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

Statystyka
Laboratorium - lista III

2021-12-14

1 Zadanie pierwsze

Podaj przedział ufności dla średniej w modelu normalnym o znanej wariancji na poziomie ufności $1-\alpha$. Uzasadnij jego postać.

1.1 Rozwiązanie

σ^2 - znane, μ - szukane

$$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$$

Dla rozkładu normalnego:

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=0}^{n}(X_i-\mu) \sim N(0,1)$$

Dla dowolnego innego (z centralnego twierdzenia granicznego):

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=0}^{n}(X_i-\mu) \to N(0,1)$$

Zatem:

$$P(-z_{\alpha/2} \leq \frac{1}{\sigma\sqrt{n}}\sum_{i=0}^{n}(X_i-\mu) \leq z_{\alpha/2}) = 1-\alpha$$

Przekształcając wychodzi, że przedział ufności dla mu to:

$$(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Gdzie $z_{\alpha/2}$ to odwrotna dystybuanta rozkładu normalnego:

$$z_{\alpha/2} = \Phi^{-1}(\alpha/2)$$

2 Zadanie drugie

Wygeneruj $n = 50$ obserwacji z rozkładu

a. normalnego z parametrem przesunięcia μ i skali σ

b. logistycznego z parametrami przesunięcia μ i skali σ

c. Cauch'ego z parametrami przesunięcia μ i skali σ

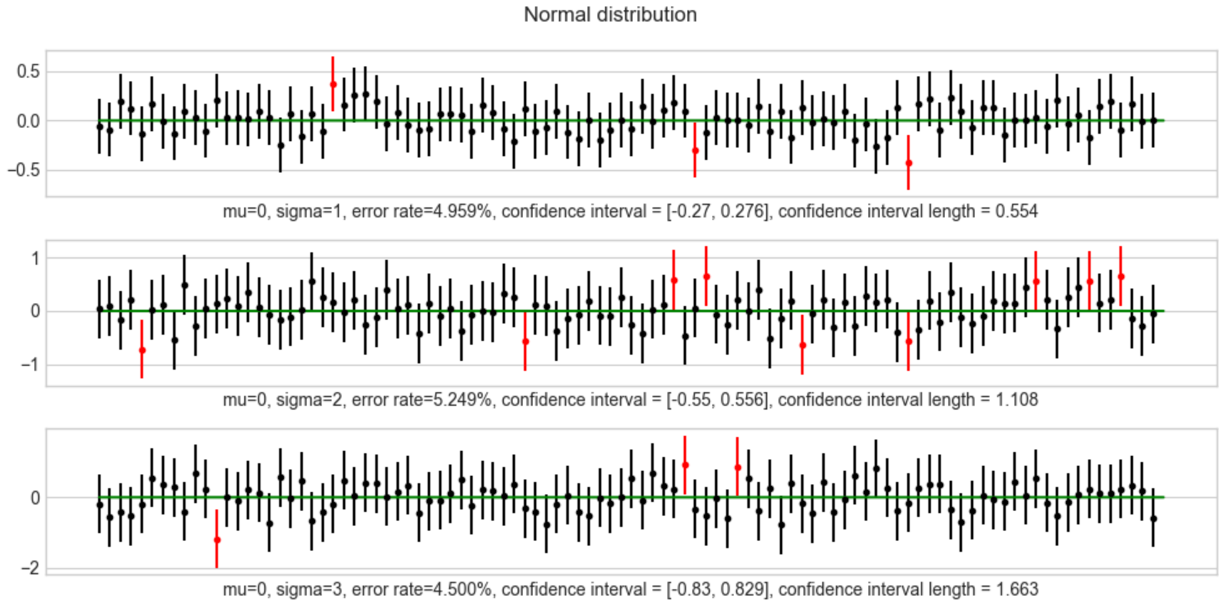
d. wykładniczego z parametrem przesunięcia λ

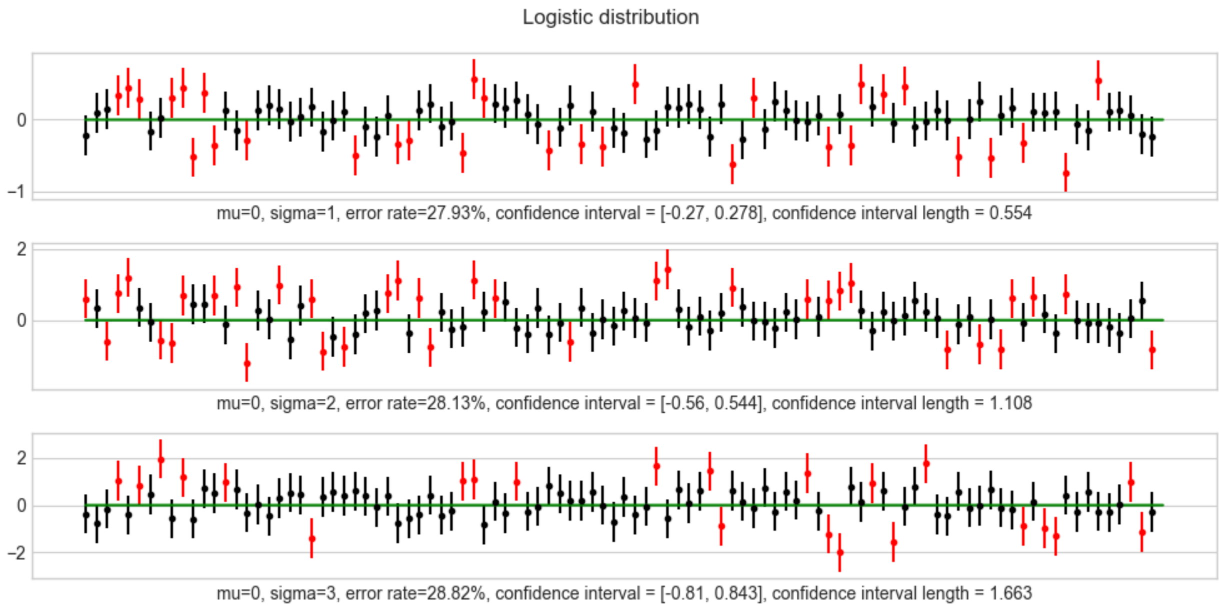
e. chi-kwadrat z ν stopniami swobody

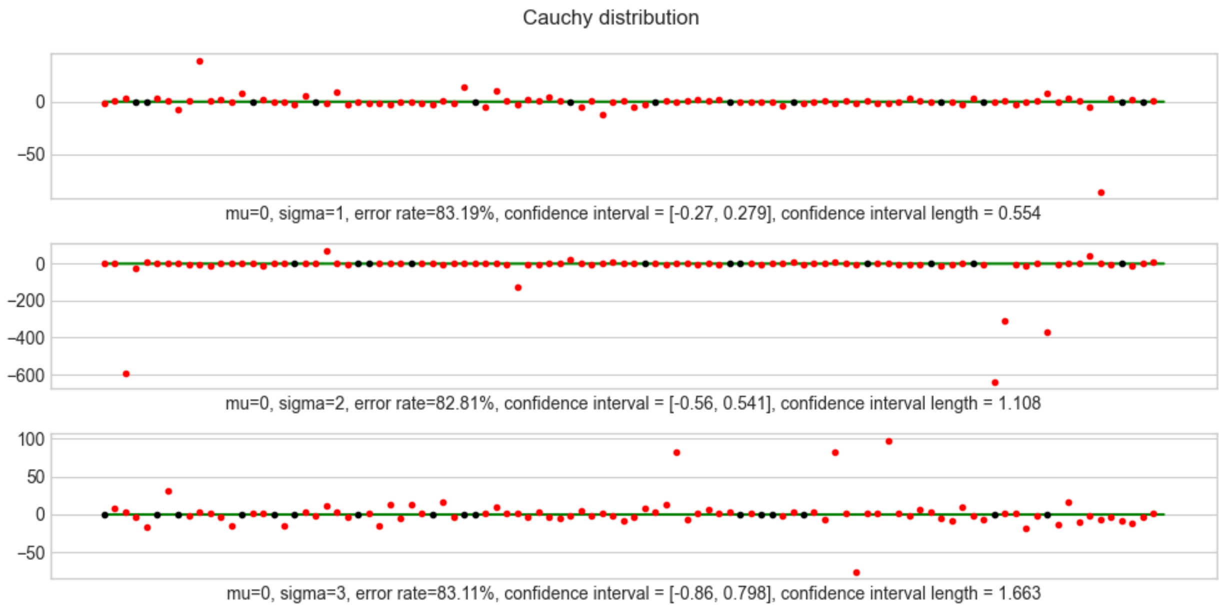
Na tej podstawie wyznacz przedziały ufności dla średniej z zadania 1 na poziomie ufności 0.95 oraz jego długość. Doświadczenie powtórz 10 000 razy. Oszacuj prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności oraz jego dzługość. Przedyskutuj uzyskane wyniki.

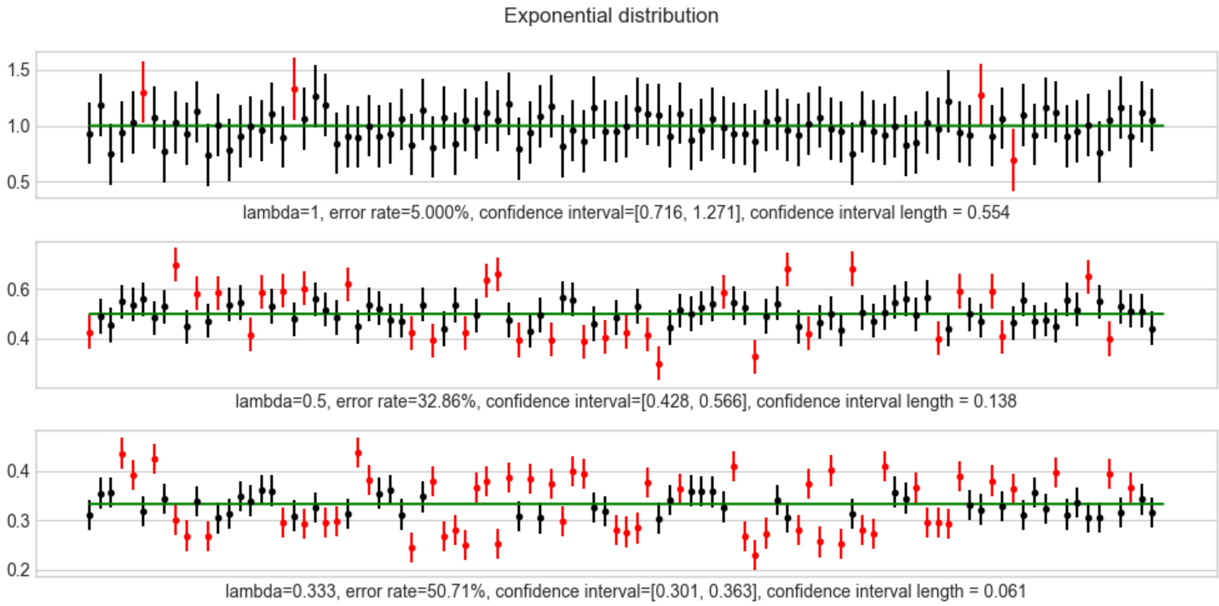
2.1 Rozwiązanie

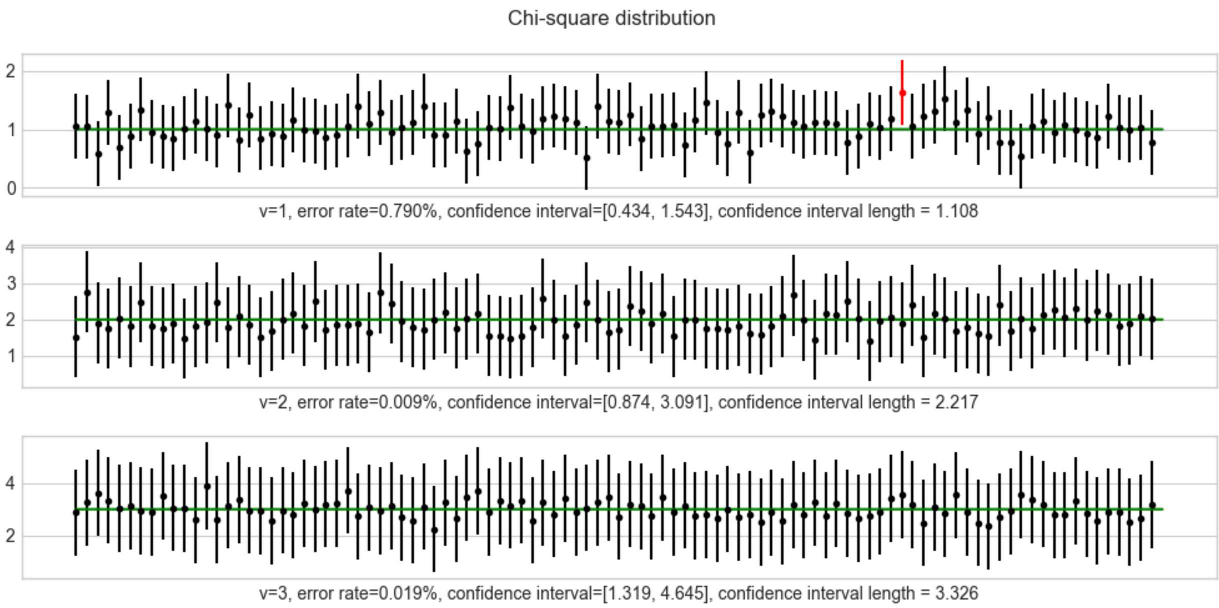
Poniższe wykresy przedstawiają 1% spośród 10000 losowań 50-elementowych próbek. Zielona linia określa prawdziwą wartość estymowanego parametru. Pionowe słupki to przedziały ufności wokół estymatu punktowego, czerwone słupki oznaczają przedziały nie zawierające prawdziwej wartości. Pod wykresem podaję 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 przedział ufności dla średniej w modelu normalnym o nieznanej wariancji na poziomie ufności $1 - \alpha$. Uzasadnij jego postać.

3.1 Rozwiązanie

Twierdzenie 1. Niech Z będzie zmienną losową o rozkładzie normalnym, U będzie zmienną losową o rozkładzie χ^2 z ν stopniami swobody, niech zmienne U i Z będą niezależne, wtedy nowa zmienna:

$$T = \frac{Z}{\sqrt{U/\nu}}$$

ma rozkład t Studenta z ν stopniami swobody.
Kontynuacja rozwiązania:
σ^2 - nie znane, μ - szukane

$$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n}) \wedge \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \chi^2_{n-1} \implies \frac{\bar{X}_n - \mu}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2/(n-1)}} \sim t_{n-1}$$

$$P(-t_{\alpha/2} \leq \frac{\bar{X}_n - \mu}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2/(n-1)}} \leq t_{\alpha/2}) = 1 - \alpha$$

Stąd przedział ufności dla mu to:

$$(\bar{X}_n - t_{\alpha/2}\frac{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}}{\sqrt{n-1}}, \bar{X}_n + t_{\alpha/2}\frac{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}}{\sqrt{n-1}})$$

Gdzie $t_{\alpha/2}$ to odwrotna dystybuanta rozkładu t Studenta:

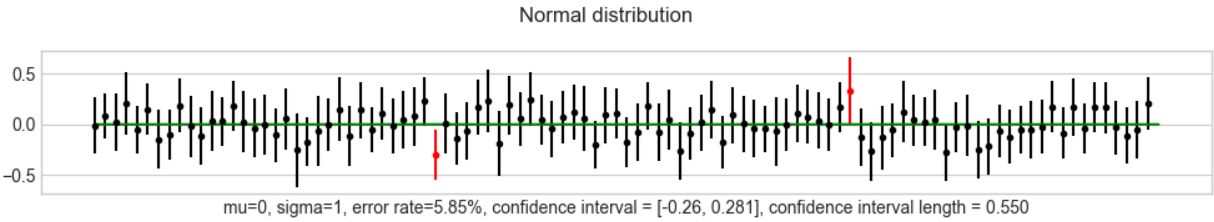
$$t_{\alpha/2} = \Phi^{-1}(\alpha/2)$$

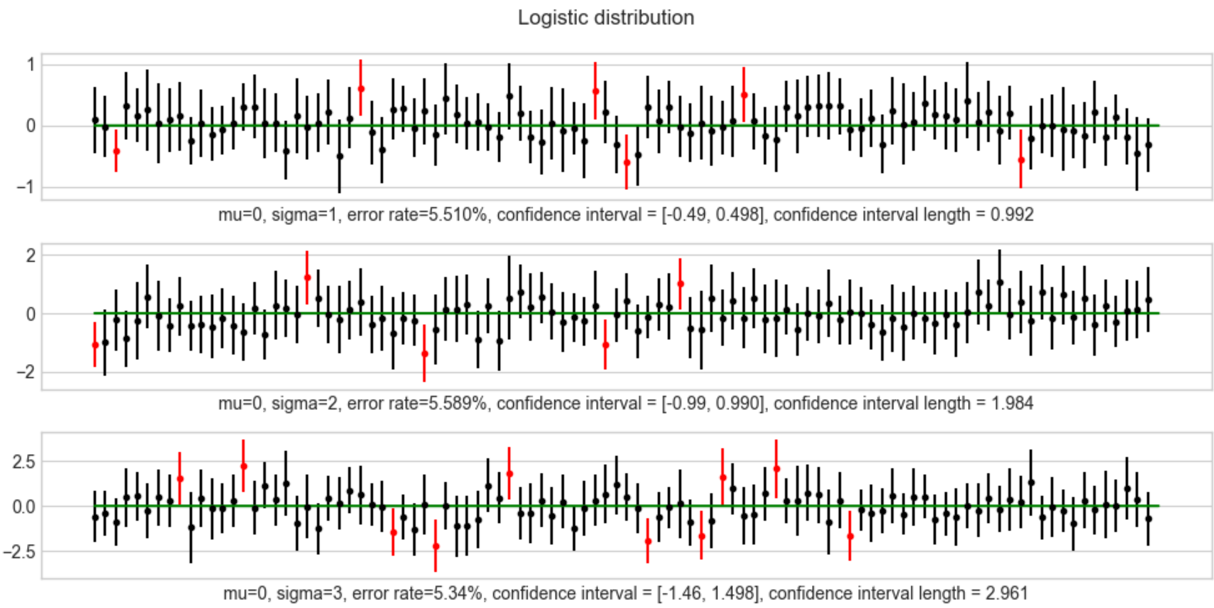
Dla dużej próby w zasadzie można by użyć metody z zadania pierwszego, ponieważ rozkład t Studenta zbiega do rozkładu normalnego wraz ze wzrostem stopnia swobody.

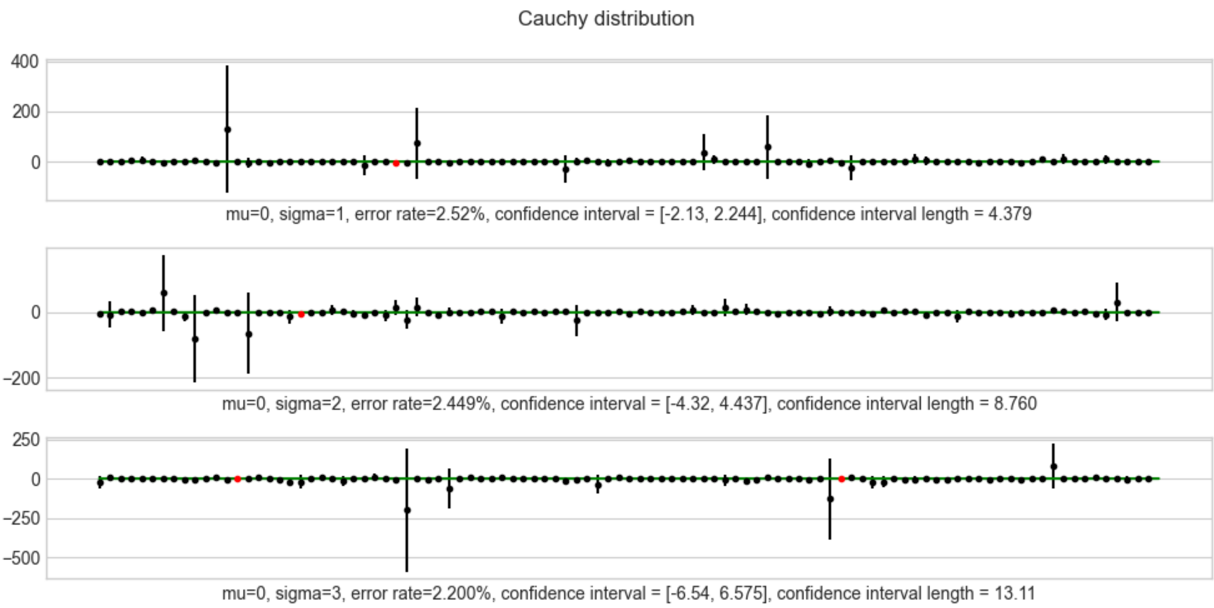
4 Zadanie czwarte

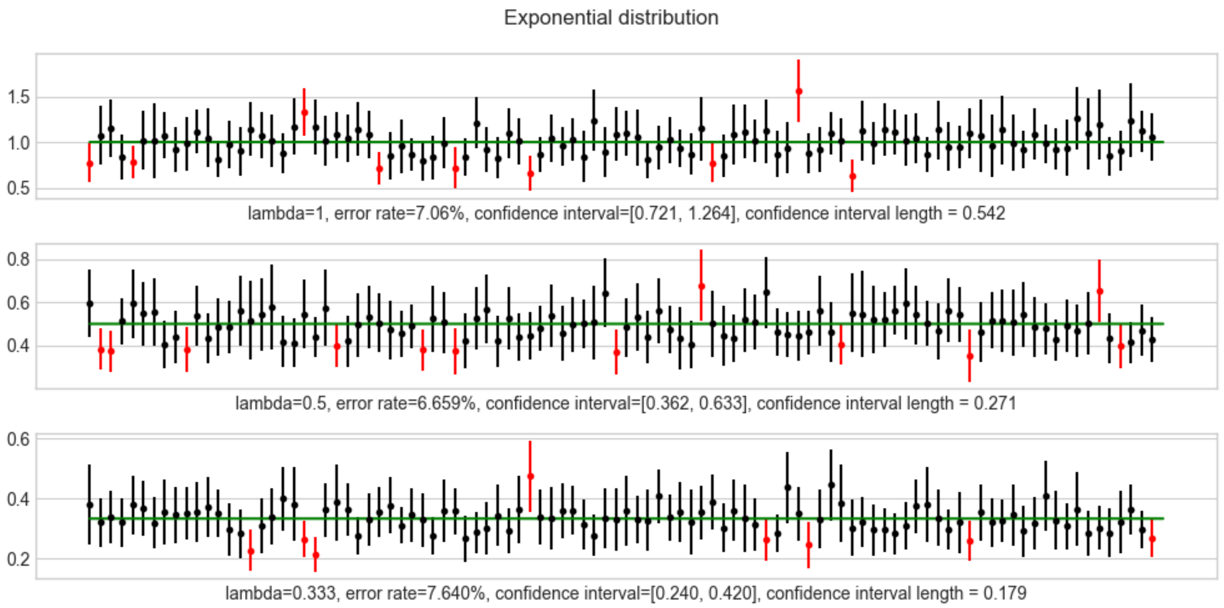
Powtórz eksperyment numeryczny z zadania 2. Na jego podstawie oszacuj prawdopodobieństwo pokrycia nieznanej średniej przez przedział ufności z zadania 3 na poziomie ufności 0.95 oraz jego długość. Przedyskutuj uzyskane rezultaty.

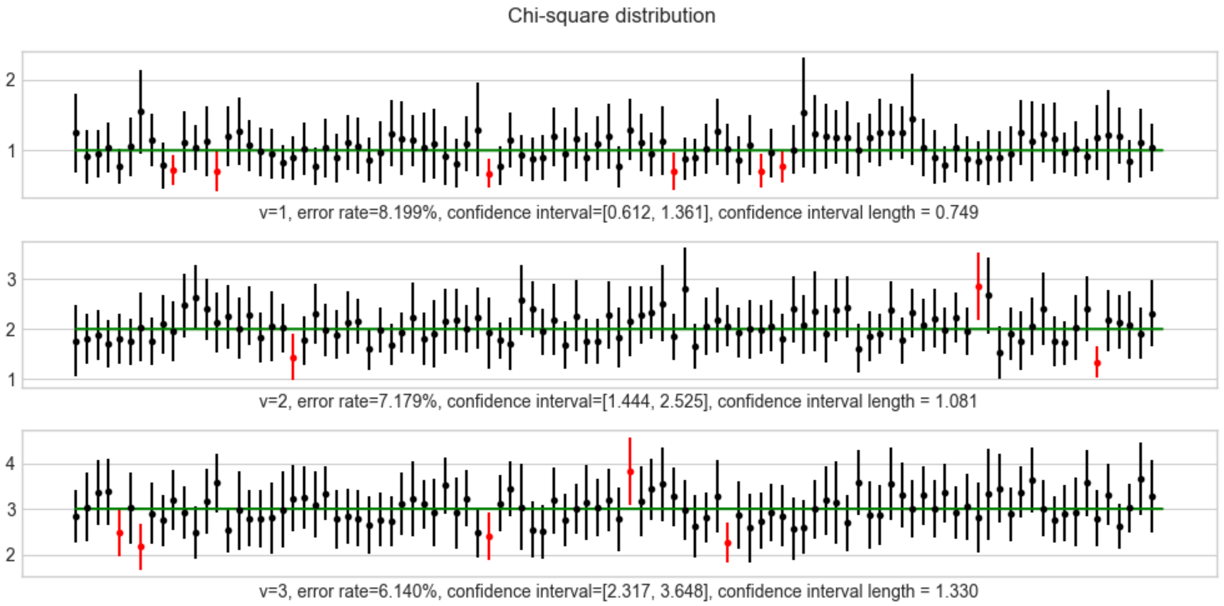
4.1 Rozwiązanie











W przypadku rozkładu normalnego otrzymałem podobne rezultaty do tych z zadania drugiego. Jednak w przypadku pozostałych rozkładów znalezione przedziały ufności były znacznie dłuższe. Mniejsza niepewność w zadaniu drugim prawdopodobnie wynika z dodatkowej wiedzy jaką jest wartość dokładna wariancji.

5 Zadanie piąte

Podaj przedział ufności dla wariancji w modelu normalnym o znanej średniej na poziomie ufności $1-\alpha$. Uzasadnij jego poprawność.

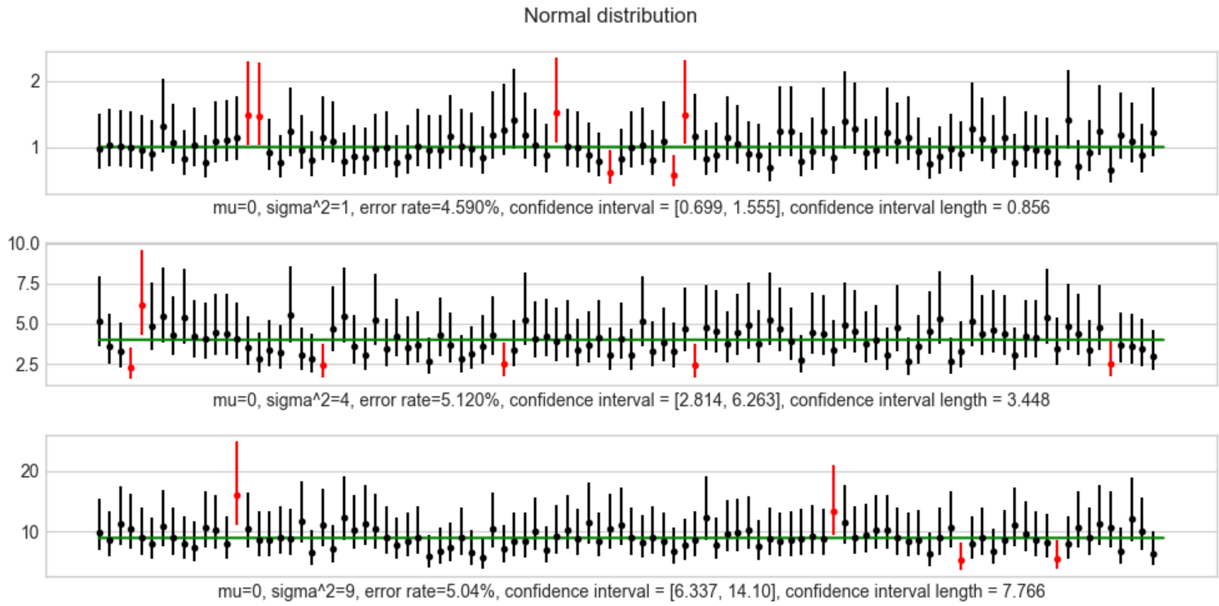
5.1 Rozwiązanie

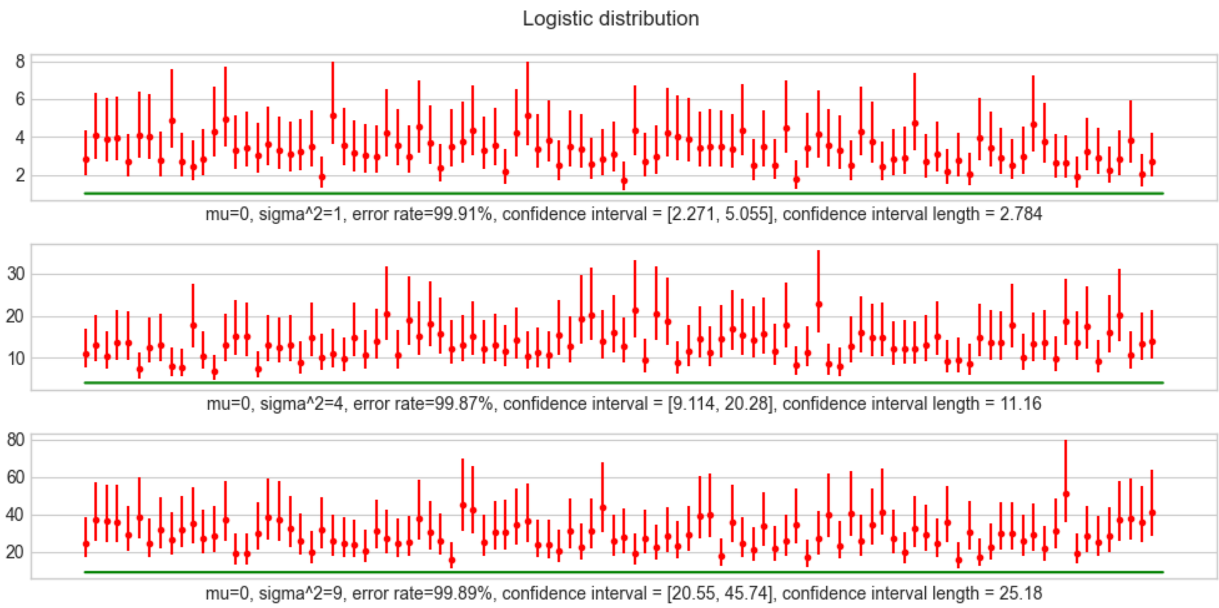
σ^2 - szukane, μ - znane
Można rozwiązać to zadanie analogicznie do kolejnego, gdzie średnia jest nieznana. Jednak skoro ją znam, to intuicyjnie wydaje się, że można uzyskać lepsze wyniki używając jej do wyznaczenia wariancji próbkowej (zamiast średniej próbkowej).

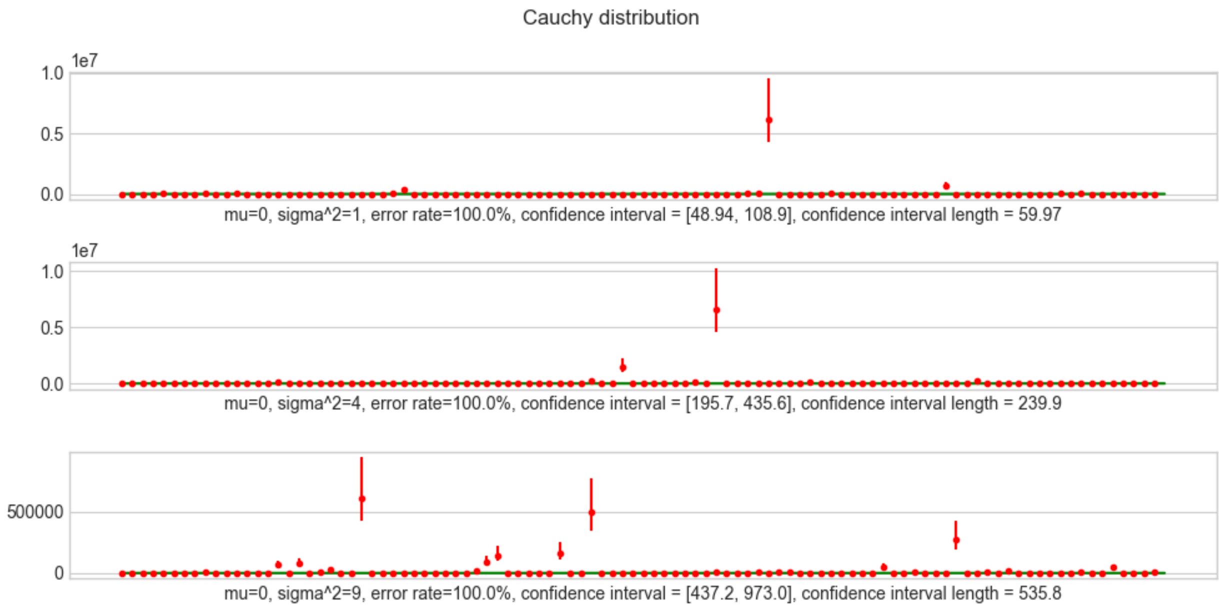
6 Zadanie szóste

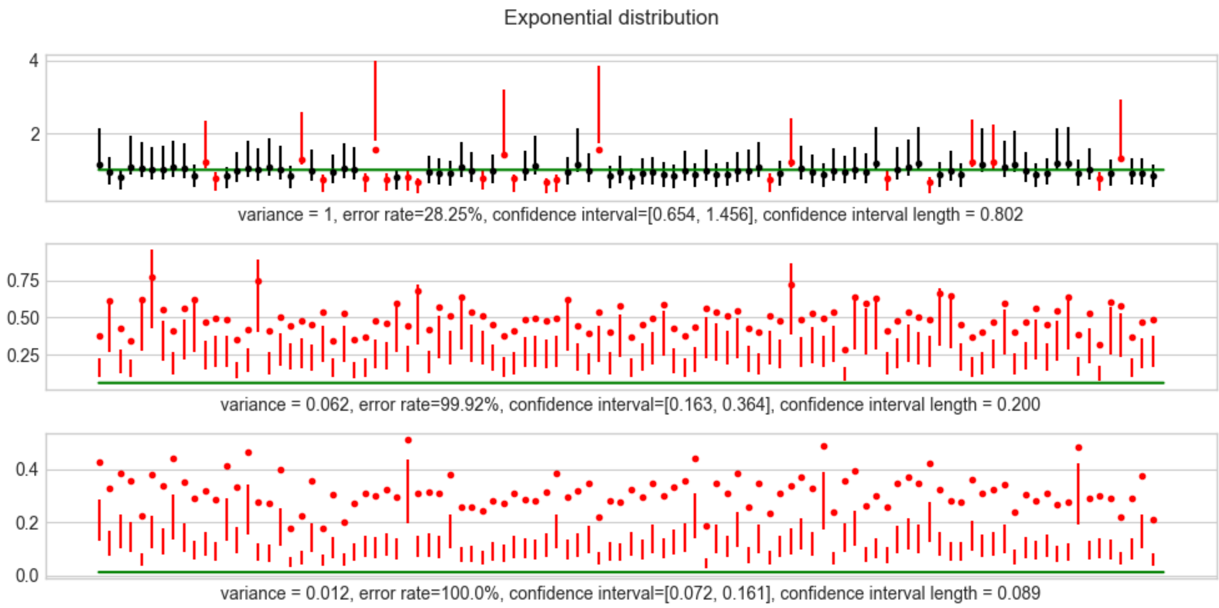
Powtórz eksperyment numeryczny z zadania 2. Na jego podstawie oszacuj prawdopodobieństwo pokrycia nieznanej wariancji przez przedział ufności z zadania 5 na poziomie ufności 0.95 oraz jego długość. Przedyskutuj uzyskane rezultaty.

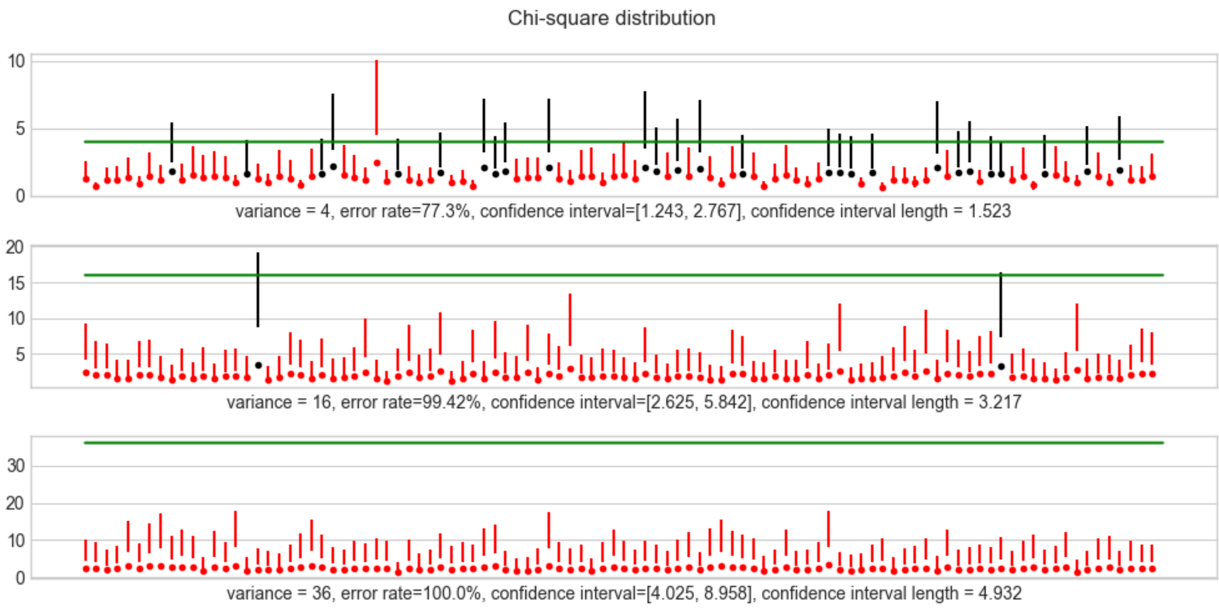
6.1 Rozwiązanie











Teoria z poprzedniego zadania dobrze sprawdza się w przypadku rozważanego rozkładu normalnego, jednak nie ma żadnej pewności, że będzie ona działać dla dowolnego innego rozkładu, co widać na powyższych wykresach.

7 Zadanie siódme

Podaj przedział ufności dla wariancji w modelu normalnym o nieznanej średniej na poziomie ufności $1-\alpha$. Uzasadnij jego poprawność.

7.1 Rozwiązanie

σ^2 - szukane, μ - nie znane
Niech $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ oznacza wariancję próbki. Wtedy $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2$ z $n-1$ stopniami swobody. Stąd:

$$P(\chi^2_{\alpha/2} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}) = 1-\alpha$$

gdzie $\chi^2_{1-\alpha/2}$ jest kwantylem rzędu $1-\alpha/2$ rozkładu χ^2 o $n-1$ stopniach swobody, a $\chi^2_{\alpha/2}$ jest odpowiednio kwantylem rzędu $\alpha/2$. Ponieważ funkcja gęstości rozkładu χ^2 nie jest symetryczna również przedział ufności nie będzie symetryczny. Przekształcając otrzymamy szukany przedział:

$$P(\frac{\chi^2_{\alpha/2}}{(n-1)s^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{1-\alpha/2}}{(n-1)s^2}) = 1-\alpha$$

$$P(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \geq \sigma^2 \geq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}) = 1-\alpha$$

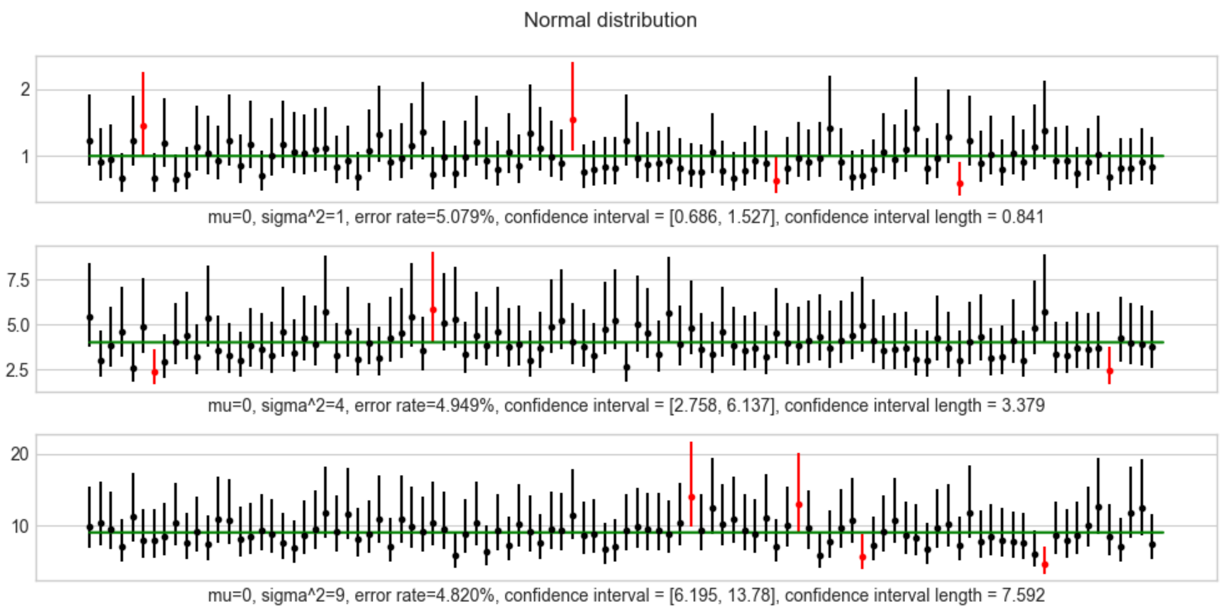
$$P(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\alpha/2}}) = 1-\alpha$$

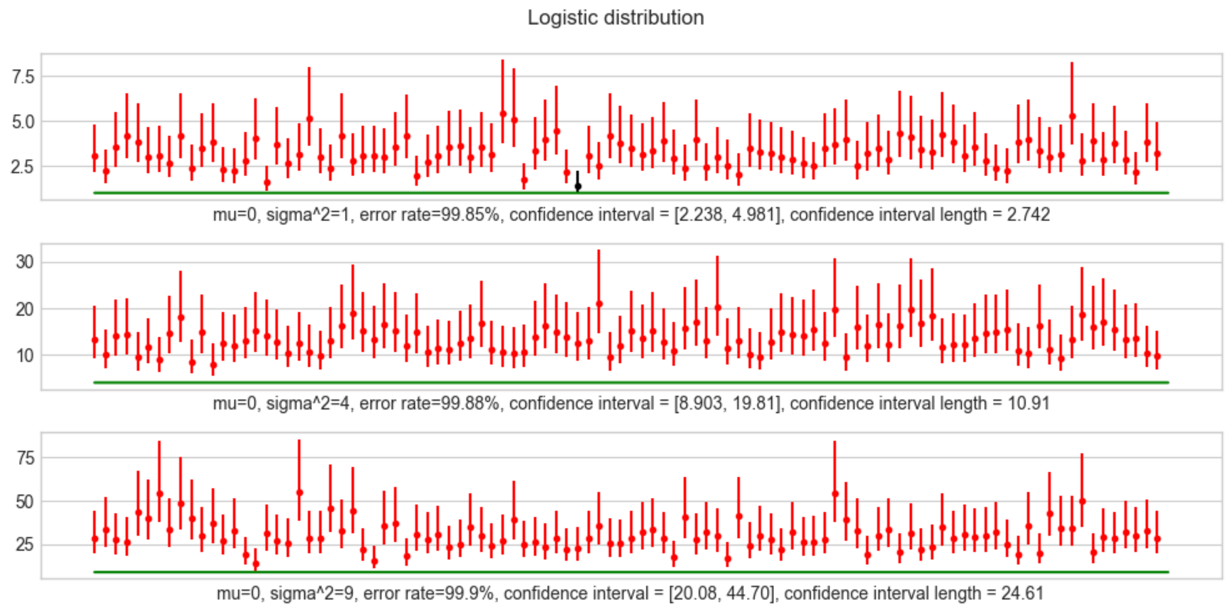
Stąd przedział ufności dla σ^2 to:

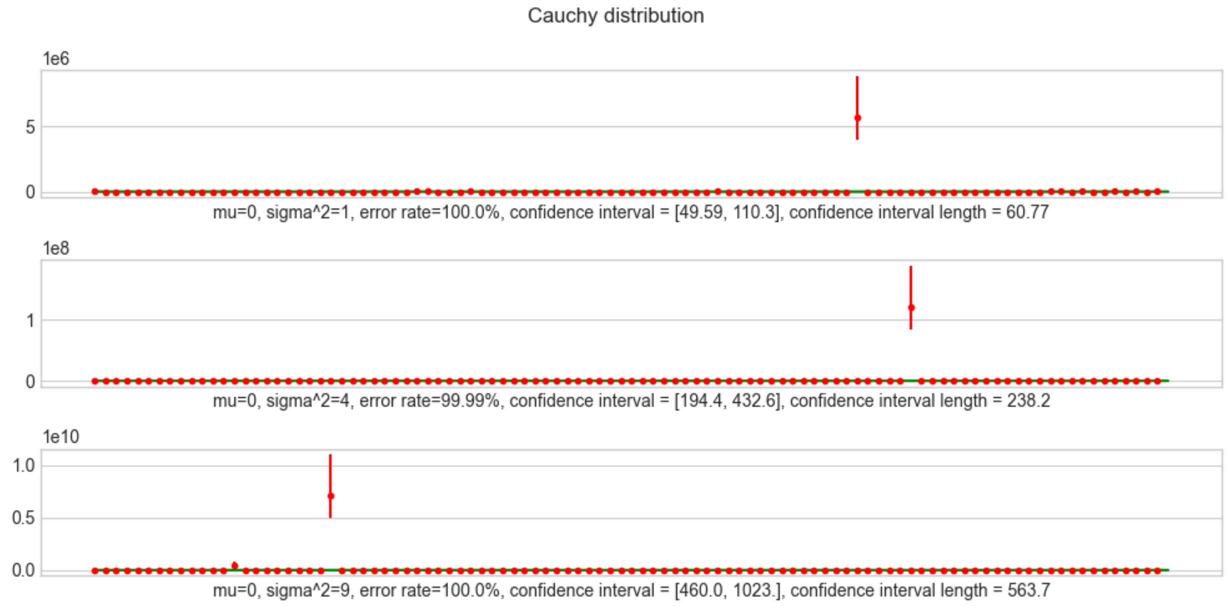
$$(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}})$$

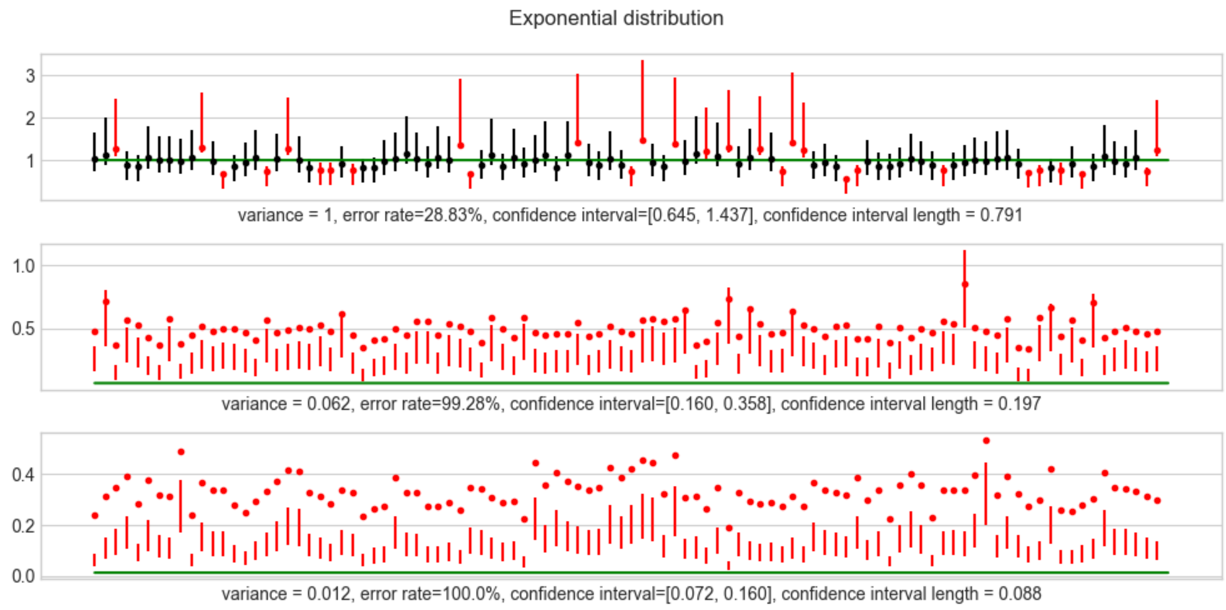
8 Zadanie ósme

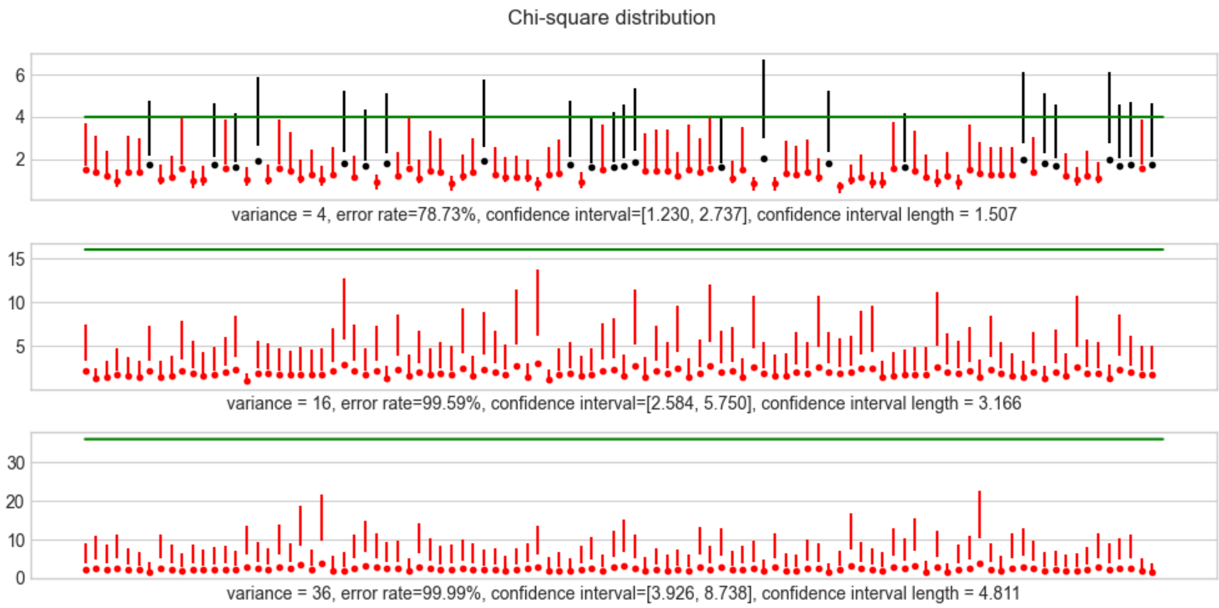
8.1 Rozwiązanie




Logistic distribution
mu=0, sigma^2=1, error rate=99.85%, confidence interval = [2.238, 4.981], confidence interval length = 2.742
mu=0, sigma^2=4, error rate=99.88%, confidence interval = [8.903, 19.81], confidence interval length = 10.91
mu=0, sigma^2=9, error rate=99.9%, confidence interval = [20.08, 44.70], confidence interval length = 24.61


Cauchy distribution
1e6
mu=0, sigma^2=1, error rate=100.0%, confidence interval = [49.59, 110.3], confidence interval length = 60.77
1e8
mu=0, sigma^2=4, error rate=99.99%, confidence interval = [194.4, 432.6], confidence interval length = 238.2
1e10
mu=0, sigma^2=9, error rate=100.0%, confidence interval = [460.0, 1023.], confidence interval length = 563.7


Exponential distribution
variance = 1, error rate=28.83%, confidence interval=[0.645, 1.437], confidence interval length = 0.791
variance = 0.062, error rate=99.28%, confidence interval=[0.160, 0.358], confidence interval length = 0.197
variance = 0.012, error rate=100.0%, confidence interval=[0.072, 0.160], confidence interval length = 0.088



Brak informacji o średniej nie wpłyną znacząco na rezultaty.

9 Zadanie dziewiąte

Podaj asymptotyczny przedział ufności dla proporcji na poziomie ufności $1-\alpha$. Uzasadnij jego postać.

9.1 Rozwiązanie

Niech X będzie zmienną losową o rozkładzie $B(n,p)$, chcemy na podstawie realizacji tej zmiennej oszacować wartość p. Niech $X_1, ..., X_n$ będzie próbą losową z określonego rozkładu, wtedy za estymat punktowy p, weźmiemy $\hat{p} = \bar{X}$. Z centralnego twierdzenia granicznego wiadomo, że $\hat{p}$ asymptotycznie ma rozkład $N(p, p(1-p)/n)$. Ponieważ $\hat{p}$ asymptotycznie zbiega do p, a $\hat{p}(1-\hat{p})$ zbiega do $p(1-p)$ to:

$$P(-z_{\alpha/2} \leq \frac{p-\hat{p}}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq z_{\alpha/2}) = 1-\alpha$$

$$P(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}) = 1-\alpha$$

Stąd asymptotyczny przedział ufności dla p to:

$$(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n})$$

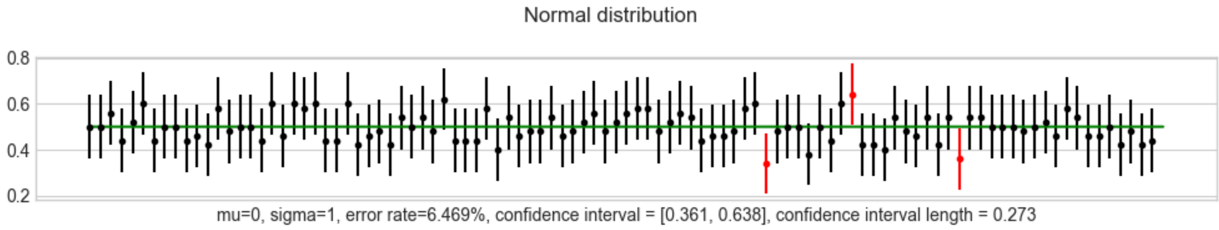
Gdzie $z_{\alpha/2}$ to odwrotna dystybuanta rozkładu normalnego:

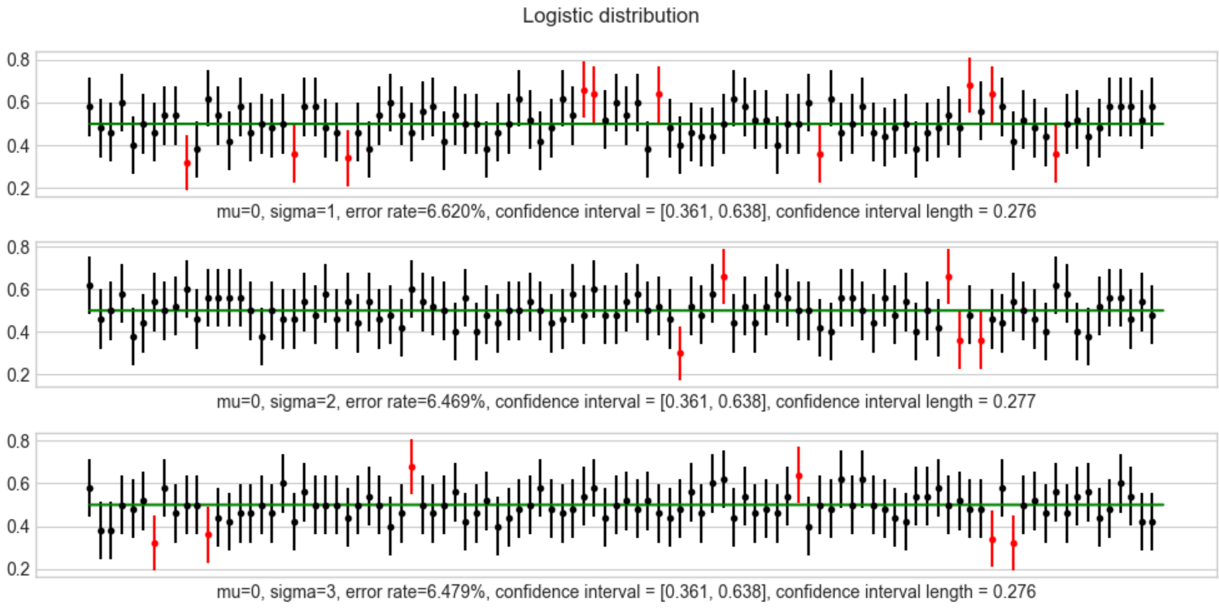
$$z_{\alpha/2} = \Phi^{-1}(\alpha/2)$$

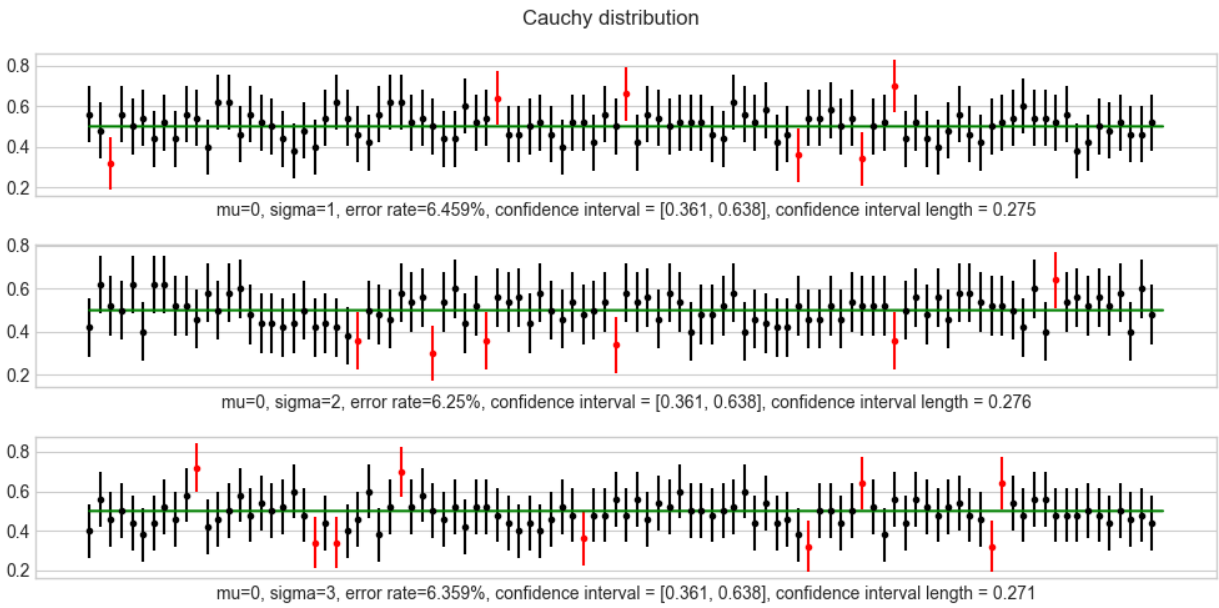
10 Zadanie dziesiąte

Powtórz eksperyment numeryczny z zadania 2, podpunkty: a, b, c. Na jego podstawie oszacuj prawdopodobieństwo pokrycia nieznanej proporcji dodatnich obserwacji przez przedział ufności z zadania 9 na poziomie ufności 0.95 oraz jego długość. Przedyskutuj uzyskane rezultaty.

10.1 Rozwiązanie







Powyższe wyniki potwierdzają prawdziwość teori dyskutowanej w poprzednim zadaniu. Przedziały ufności zdają się pokrywać z zadaną pewnością wartość dokładną.

11 Zadanie jedenaste

Powtórz eksperyment numeryczny z zadań 2, 4, 6, 8, 10, dla $n = 20$ i $n = 100$. 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.
- Błędne przedziały funości w zadaniu szóstym i ósmym nie są wynikiem zbyt małej próby
- Wartości error rate z zadania 10 faktycznie zbiegają do 5%