

Teoria analizy dużych zbiorów - sprawozdanie 1

Stanisław Wilczyński

5 marca 2017

Zadanie1

Sprawdźmy czy zgodnie z teorią z wykładu zachodzą nierówności:

$$\frac{\phi(x) \cdot x}{x^2 + 1} \leq 1 - \Phi(x) \leq \frac{\phi(x)}{x}$$

oraz czy granice tych wyrażeń przy $x \rightarrow \infty$ są równe. Niech

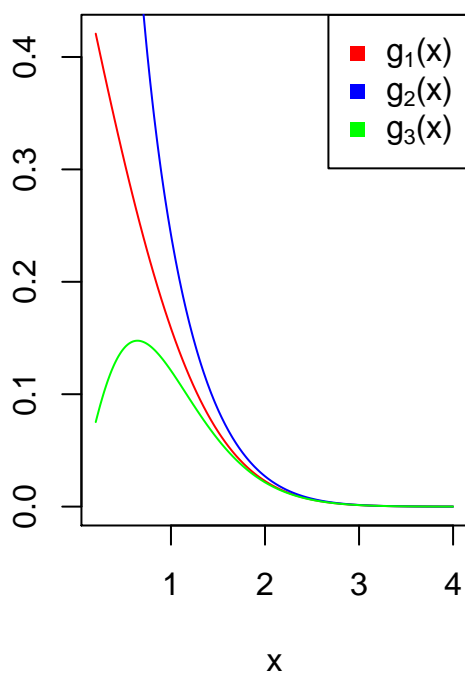
$$g_1(x) = 1 - \Phi(x)$$

$$g_2(x) = \frac{\phi(x)}{x}$$

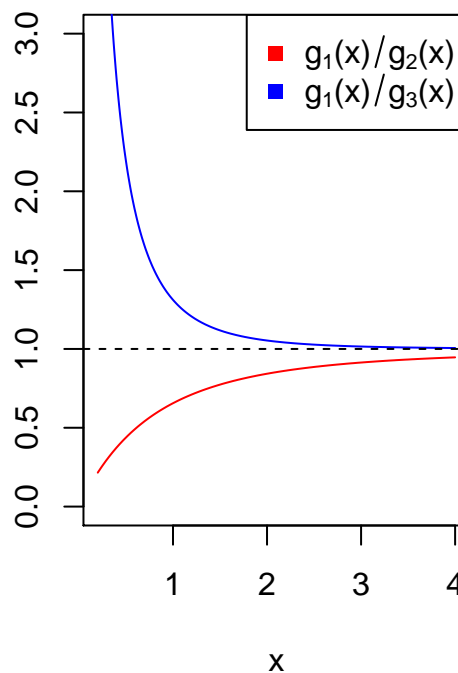
$$g_3(x) = \frac{\phi(x) \cdot x}{x^2 + 1}$$

Wtedy dostajemy

Funkcje



Ilorazy



Nasze wyniki zgadzają się z teorią z wykładu - nierówności są zachowane oraz granice są równe. Możemy jednak zauważyć, że dla $x = 2$ wyrażenia $g_1(x)$ oraz $g_3(x)$ są już prawie równe natomiast funkcje $g_1(x)$ oraz $g_2(x)$ zbliżają się do siebie wolniej.

Zadanie 2

Niech

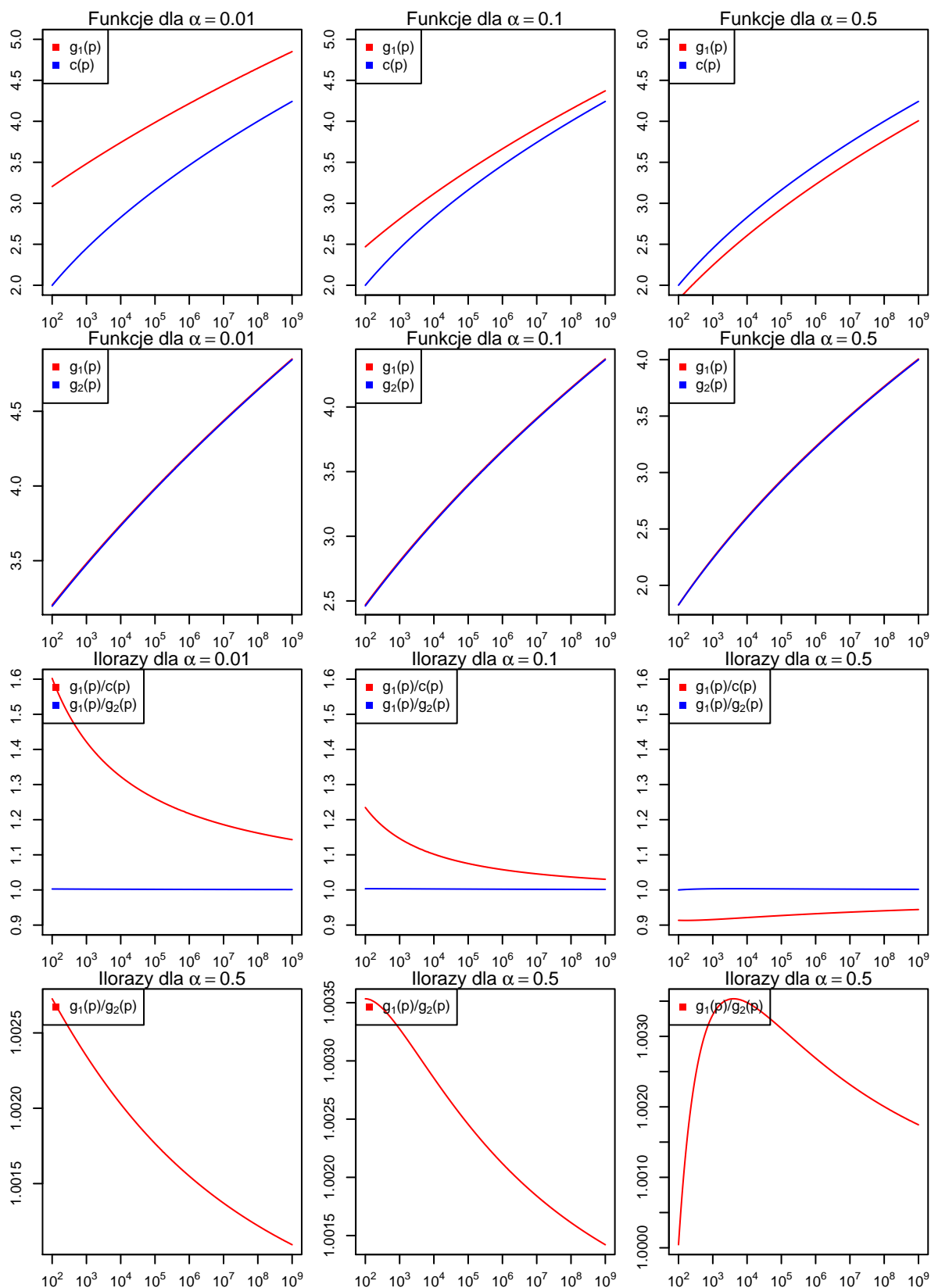
$$g_1(p, \alpha) = \Phi^{-1} \left(1 - \frac{\alpha}{2p} \right)$$

$$c(p) = \sqrt{2 \log p}$$

$$B(p, \alpha) = 2 \log \left(\frac{2p}{\alpha} \right) - \log(2\pi)$$

$$g_2(p, \alpha) = \sqrt{B \left(1 - \frac{\log B}{B} \right)}$$

Powyższe funkcje zostały przedstawione na wykresach poniżej razem z odpowiednimi ilorazami z treści zadania.



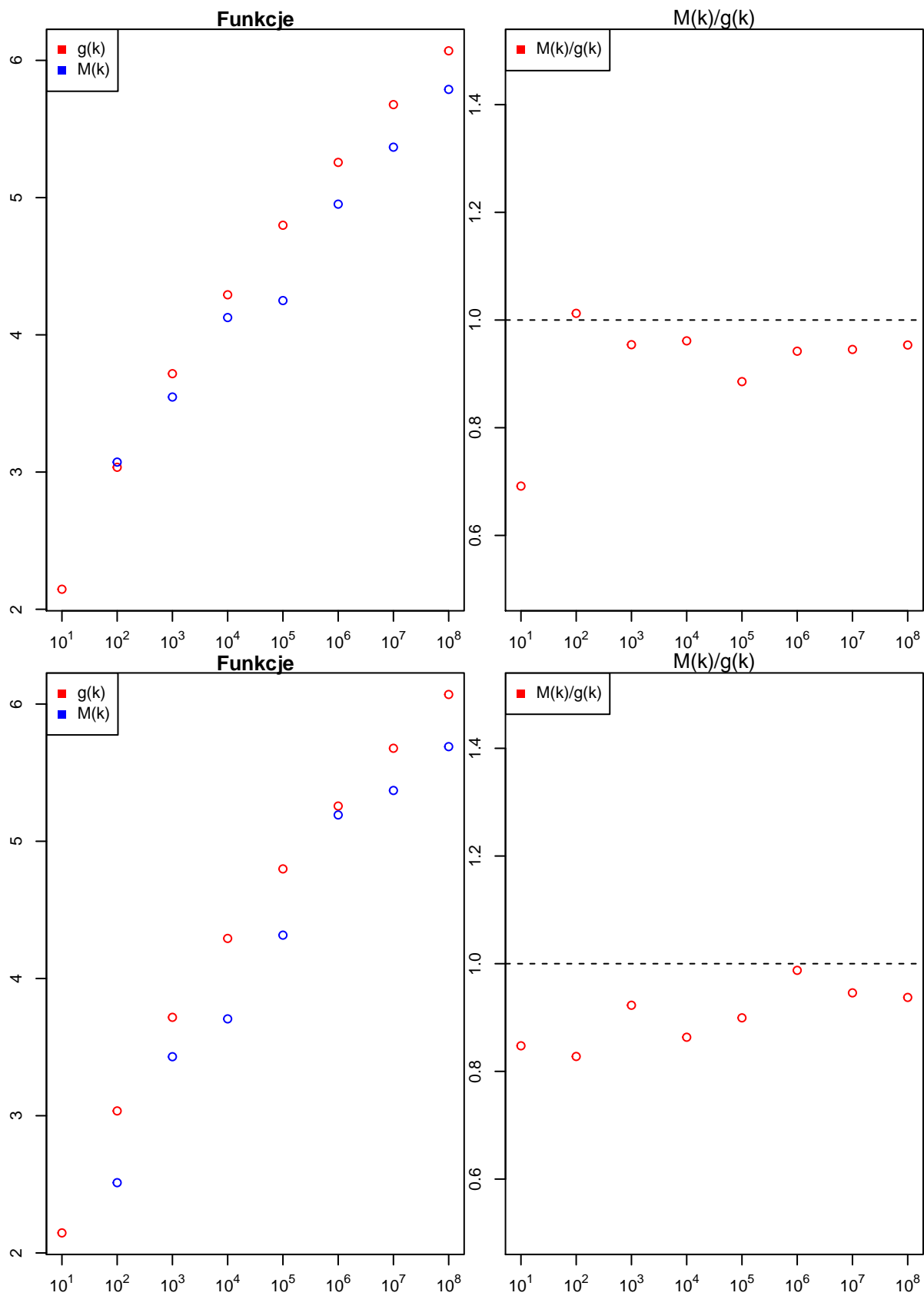
Wyniki, które możemy odczytać z wykresów potwierdzają teorię z wykładu, tzn. dla $p \rightarrow \infty$ mamy, że granice funkcji $g_1(p, \alpha)$, $g_2(p, \alpha)$ są równe, a zbieżność można zaobserwować już dla niewielkich wartości p takich jak

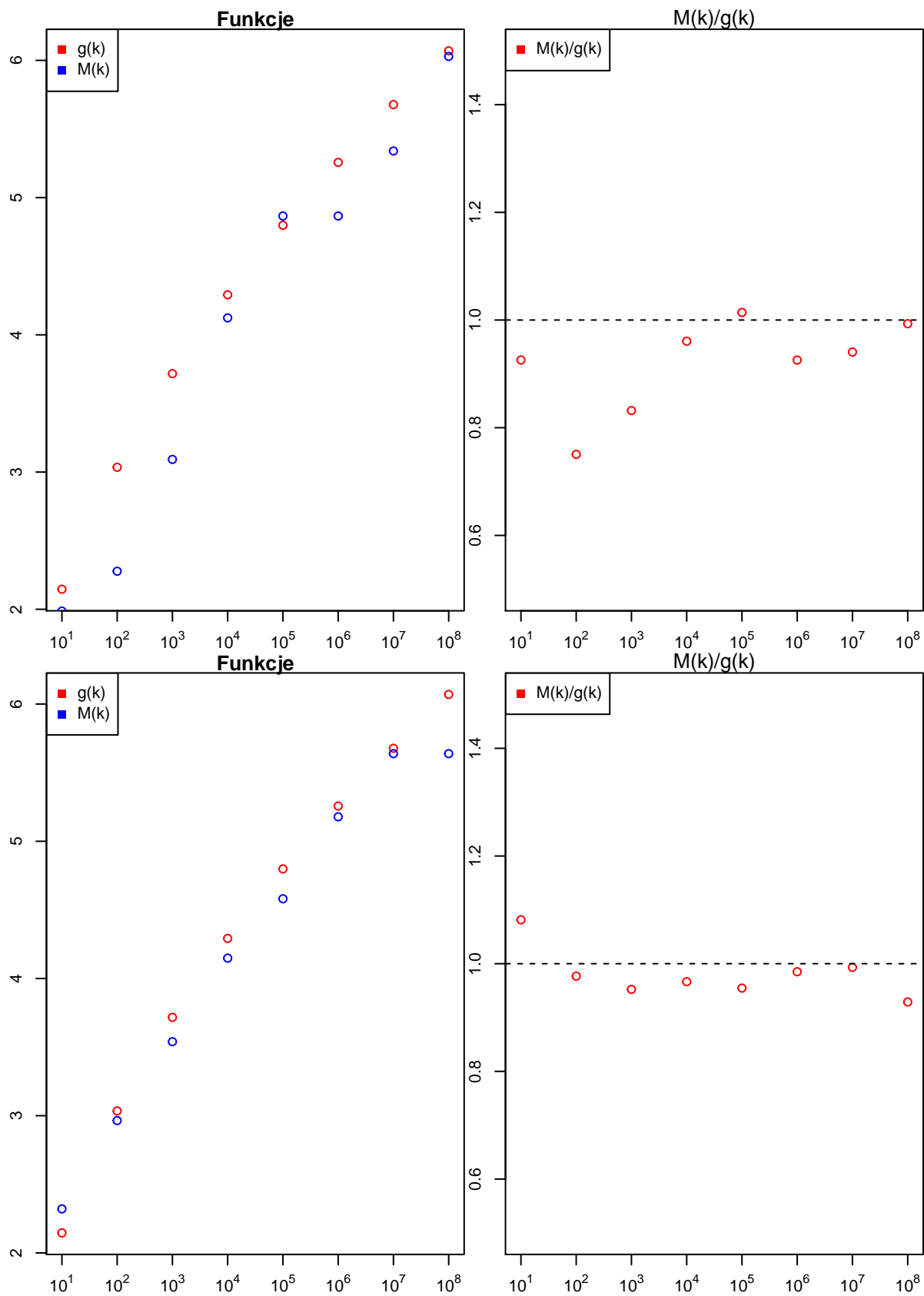
$10^2, 10^3$ i to niezależnie od parametru α . Jeśli chodzi o funkcję $c(p)$ jest ona pewnym przybliżeniem $g_1(p)$, jednak dla pewnych wartości parametru α , np. 0.01 jest ona dość słabym przybliżeniem.

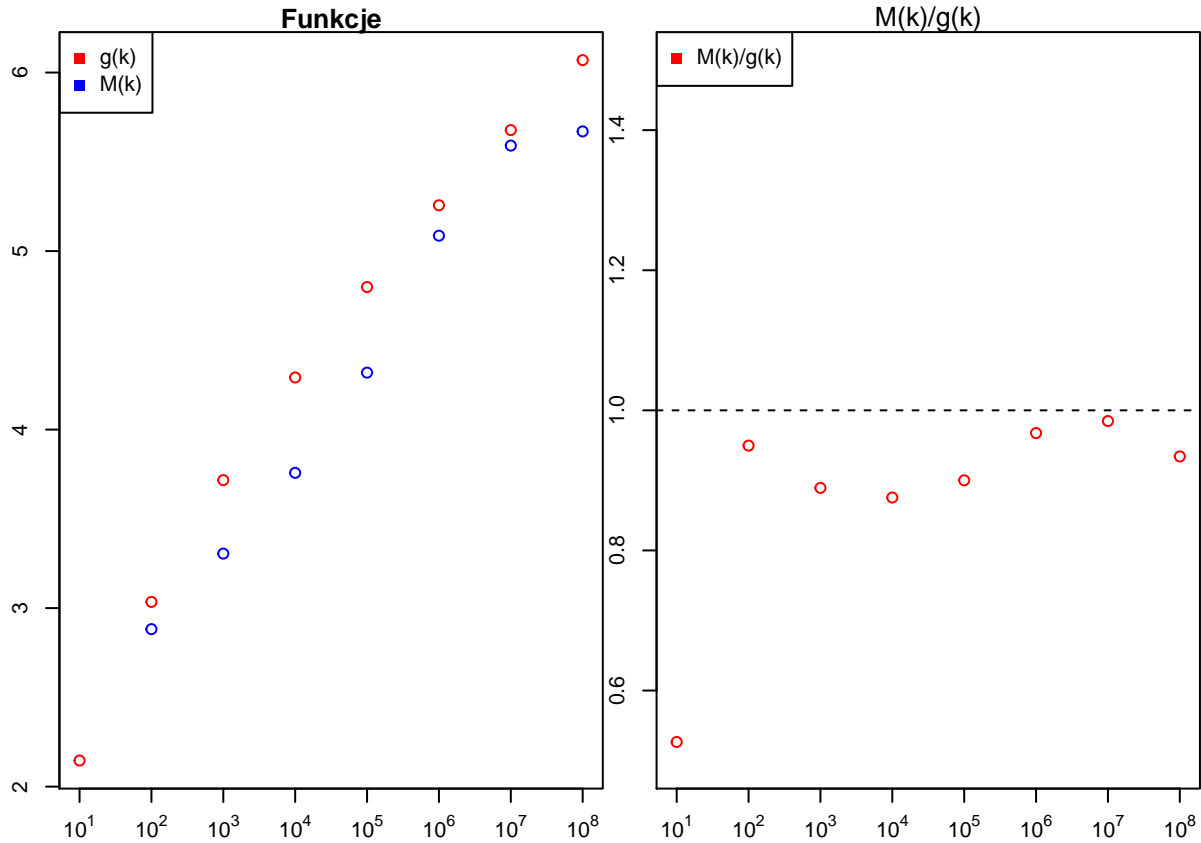
Zadanie 3

Niech Y_1, \dots, Y_p dla $p = 10^8$ będą i.i.d $\sim N(0, 1)$. Niech $M(k) = \max_{j \in \{1, \dots, k\}} |Y_j|$ oraz $g(k) = \sqrt{2 \log k}$ dla $k \in \{10^1, 10^2, \dots, 10^8\}$. Będziemy porównywać funkcje M i g . Przetestujemy w ten sposób teorię mówiącą, że dla Z_1, \dots, Z_n i.i.d zmiennych losowych z rozkładu $N(0, 1)$ mamy

$$\frac{\max_{i \in \{1, \dots, n\}} |Z_i|}{\sqrt{2 \log n}} \xrightarrow{P} 1$$







Możemy zauważyć, że rzeczywiście dla dużych wartości k (od około 10^5) funkcje $M(k), g(k)$ zbliżają się do siebie, jednak nie jest to zbyt wyraźna i szybka zbieżność tak jak w przykładach w poprzednich zadaniach.

Zadanie 4

W tym zadaniu będziemy estymować moc testów Bonferroniego i Fishera. Niech $p = 5000$. Zakładamy, że nasze zmienne pochodzą z rozkładu normalnego o wariancji 1. Dla $H_0 : \mu_1, \dots, \mu_p = 0$ będziemy testować alternatywy:

$$H_1 : \mu_1 = 1.2\sqrt{2\log p}, \mu_2, \dots, \mu_p = 0$$

$$H_2 : \mu_1, \dots, \mu_{1000} = 0.15\sqrt{2\log p}, \mu_{1001}, \dots, \mu_p = 0$$

Aby oszacować moc, wygenerujemy zmienne z rozkładów przy hipotezie alternatywnej i sprawdzimy w jak wielu przypadkach hipoteza zerowa została odrzucona.

Niech $\hat{\gamma}_B, \hat{\gamma}_F, \hat{\gamma}_C$ oznaczają wyestymowane przez nas moce testów odpowiednio Bonferroniego, Fishera i chi-kwadrat. Przy testowaniu przeciwko H_1 otrzymujemy $\hat{\gamma}_B = 0.699$, $\hat{\gamma}_F = 0.071$, $\hat{\gamma}_C = 0.086$.

Przy testowaniu przeciwko H_2 otrzymujemy natomiast $\hat{\gamma}_B = 0.11$, $\hat{\gamma}_F = 0.98$ oraz $\hat{\gamma}_C = 0.982$. Te wyniki potwierdzają przedstawione na wykładzie cechy tych testów. Otoż test Bonferroniego ma w pierwszym przypadku dużą moc, gdyż H_0 odrzucamy, gdy $\min(p_{value}) \leq \frac{\alpha}{p}$, służy, więc do wykrywania dużego odstępstwa od normy na jednej zmiennej. Z kolei test Fishera odrzuca H_0 dla dużych wartości statystyki $-2 \sum \log p_{value}$, dlatego działa lepiej w drugim przypadku - kiedy wiele zmiennych odbiega od normy, ale niekoniecznie w znacznym stopniu którakolwiek z nich. Podobnie test chi-kwadrat - odrzuca on H_0 dla dużych wartości statystyki $\sum Y_i^2$, gdzie Y_i były naszą próbą losową, dlatego działa lepiej w drugim przypadku.