Drugi raport z symulacji Monte Carlo

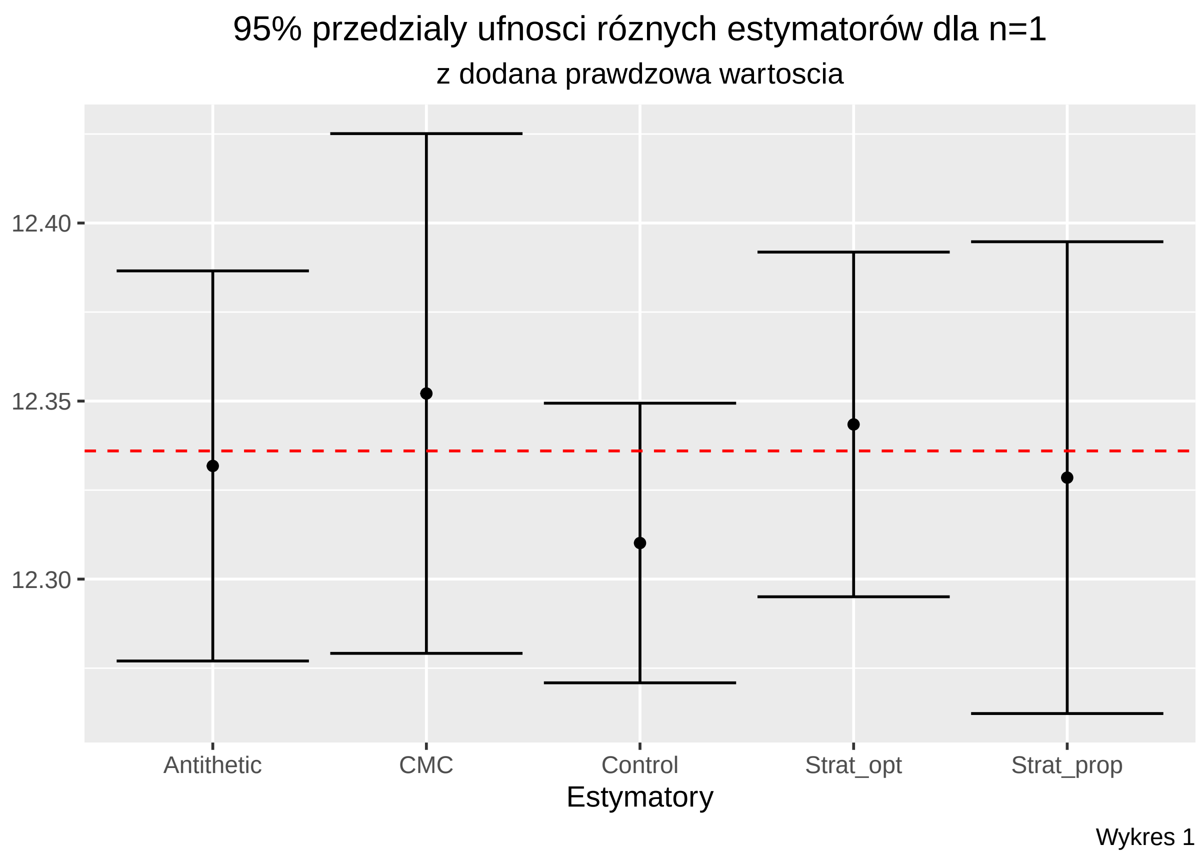
Jan Kozłowski

25-01-2025

Wprowadzenie

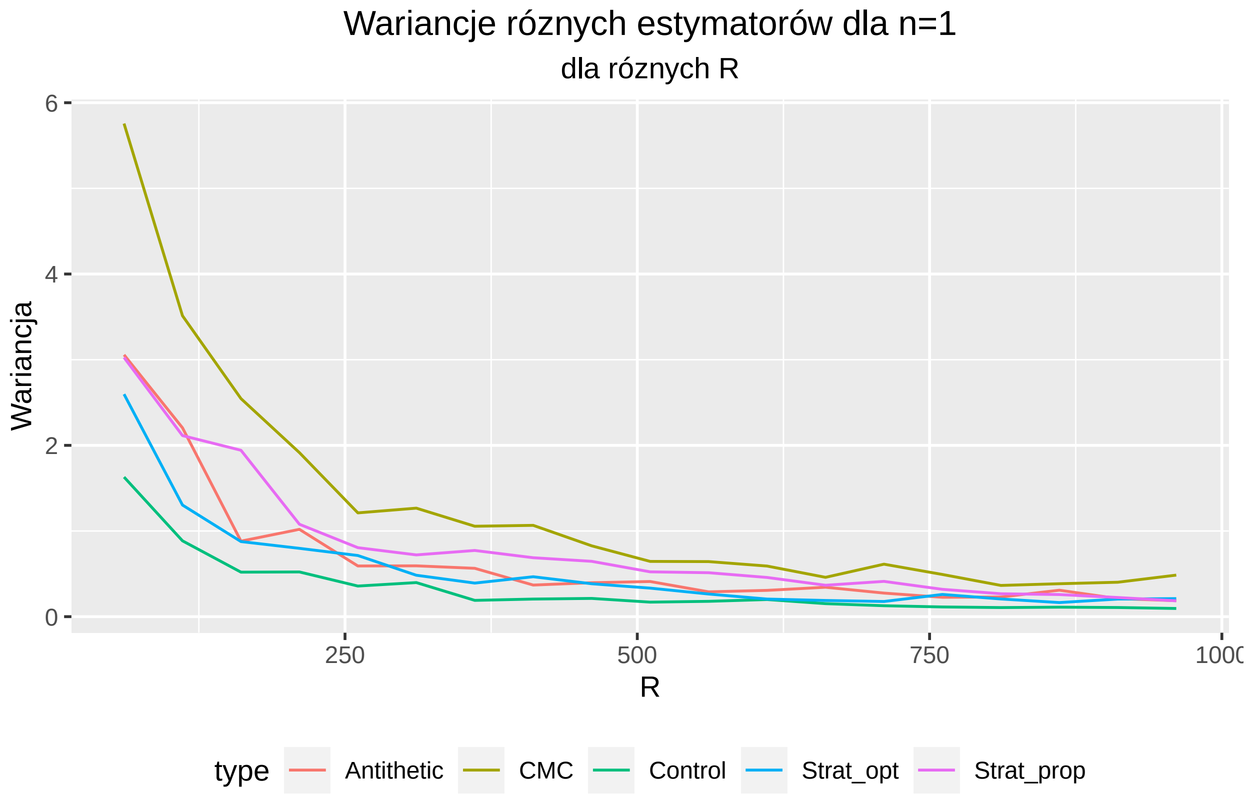
Przypadek dla n=1

Zobatrzmy jak będą sie zachowywać nasze estymatory w sytuacji, gdy $n = 1$, symulacje będziemy przeprowadzać dla $R = 500$ oraz będziemy je powtarazć $N = 500$ razy. Zacznijmy od narysowania przedziałów ufnośći dla tych estymatorów:



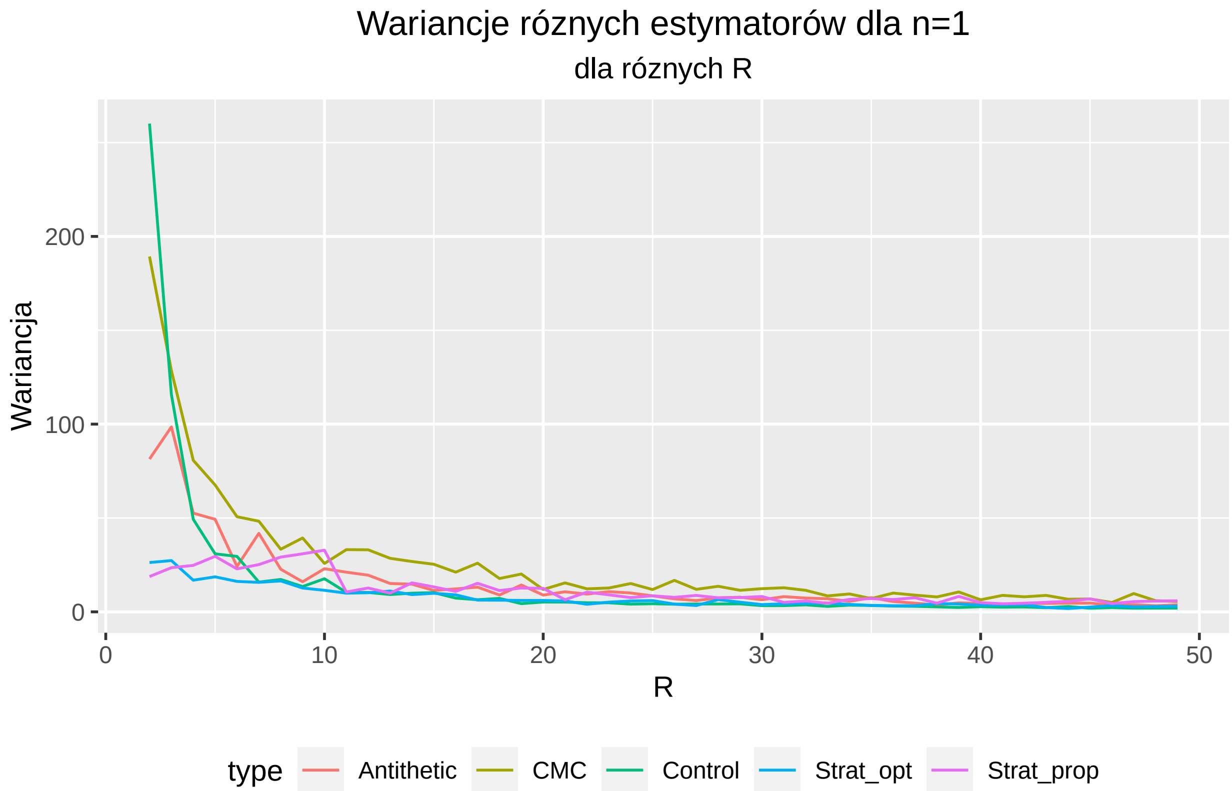
Możemy zauważyć, że dla każdego estymatora wartość średnia, czyli środek przedziału ufności jest bardzo blisko wartości prawdziwej, co zgadza się z nieobciążościami tych estymatorów. Z drugiej strony porównując wariancje, która odpowaida za "szerokość" przedziału ufności widzimy, że najniższa jest dla estymatorów Control variates oraz Stratified z opcją optymalniej alokacji, a największa wariancja jest dla podstawowego estymatora Monte Carlo.

Innym pytaniem jakie możemy sobie zadać jest zbadanie wpływu R na otrzymywane wariancje. Tutaj znów na $N = 100$ porównajmy wyniki, z powodu dużych różnic dla małych R zobaczmy najpierw wartości dla $R > 50$



Wykres 2

Możemy zauważyć, że estymator Control variates daje wciąż najlepsze wyniki na największą wariancje ma estymator CMC, choć dla każdego estymatora tempo zbieżości wydaje się być podobne. Pozostaje przyjżeć się wukresowi dla $R < 50$

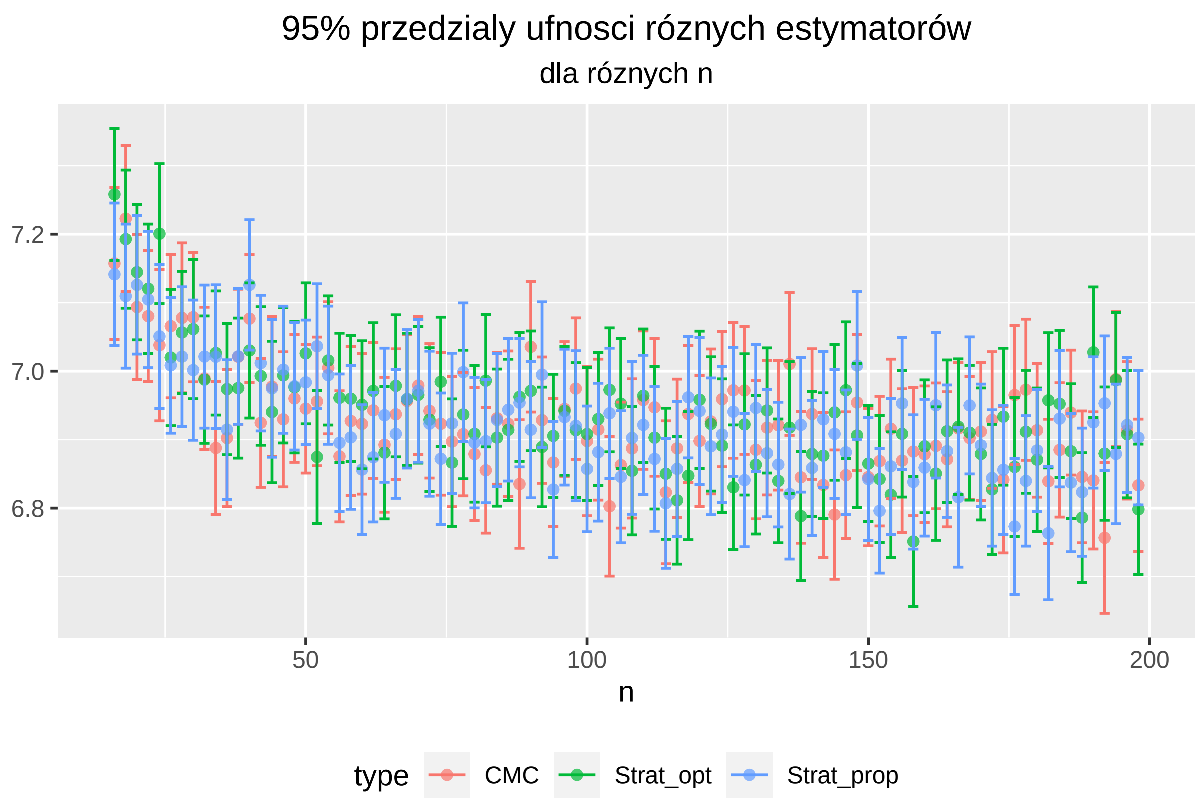


Wykres 3

Tutaj choć różnice bezwględne są duże to ogólne trend jest podobny, zastanawiająca może być duża wariancja estymatora Control variates dla $R < 0$ ale w takich losowaniach możliwe jest, że cowariancja między I a $B(1)$ będzie mała, przez co ten estymator nie będzie znacząco "poprawiać" estymatora CMC.

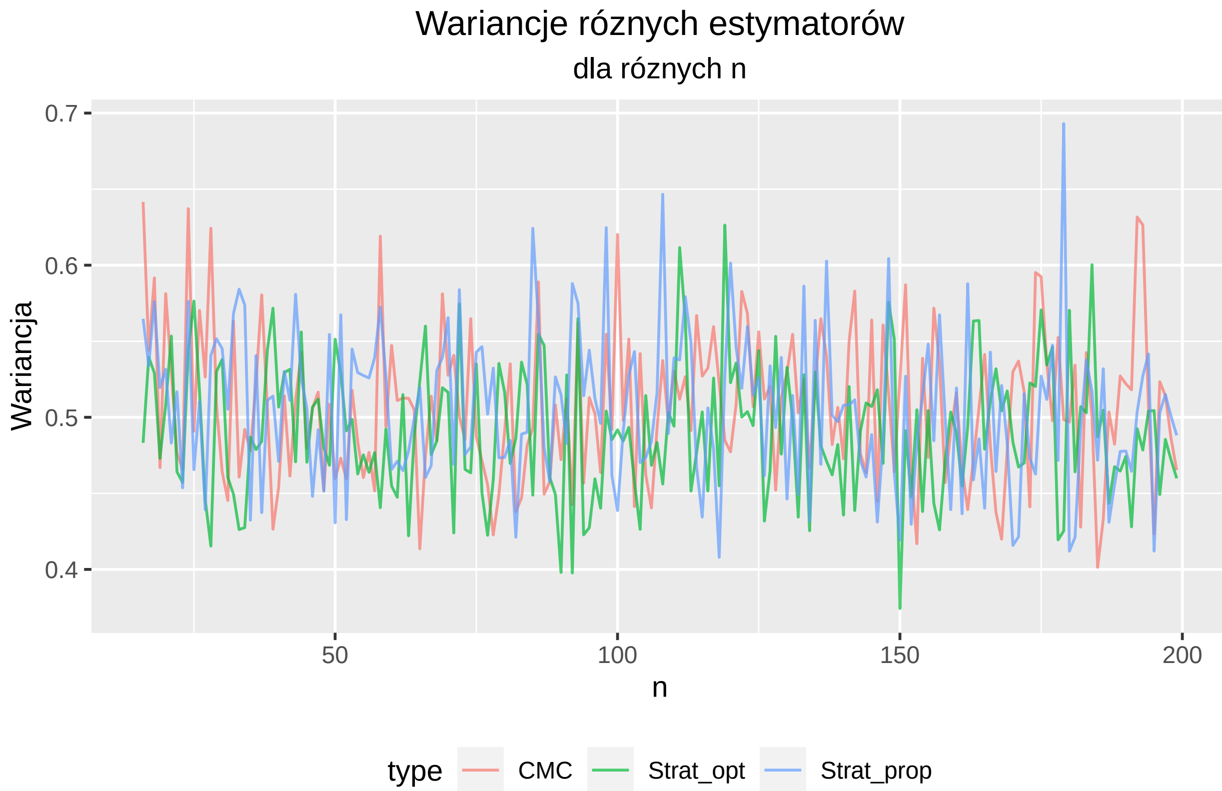
Przypadek dla n>1

Teraz skupmy się na sytuacji gdy $n > 1$ oraz na estymatorach CMC oraz Stratified. Znów zacznijmy od narywosania przedziałów ufności oraz ograniczmy się do $n > 15 (R = 200$ i $N = 200)$:

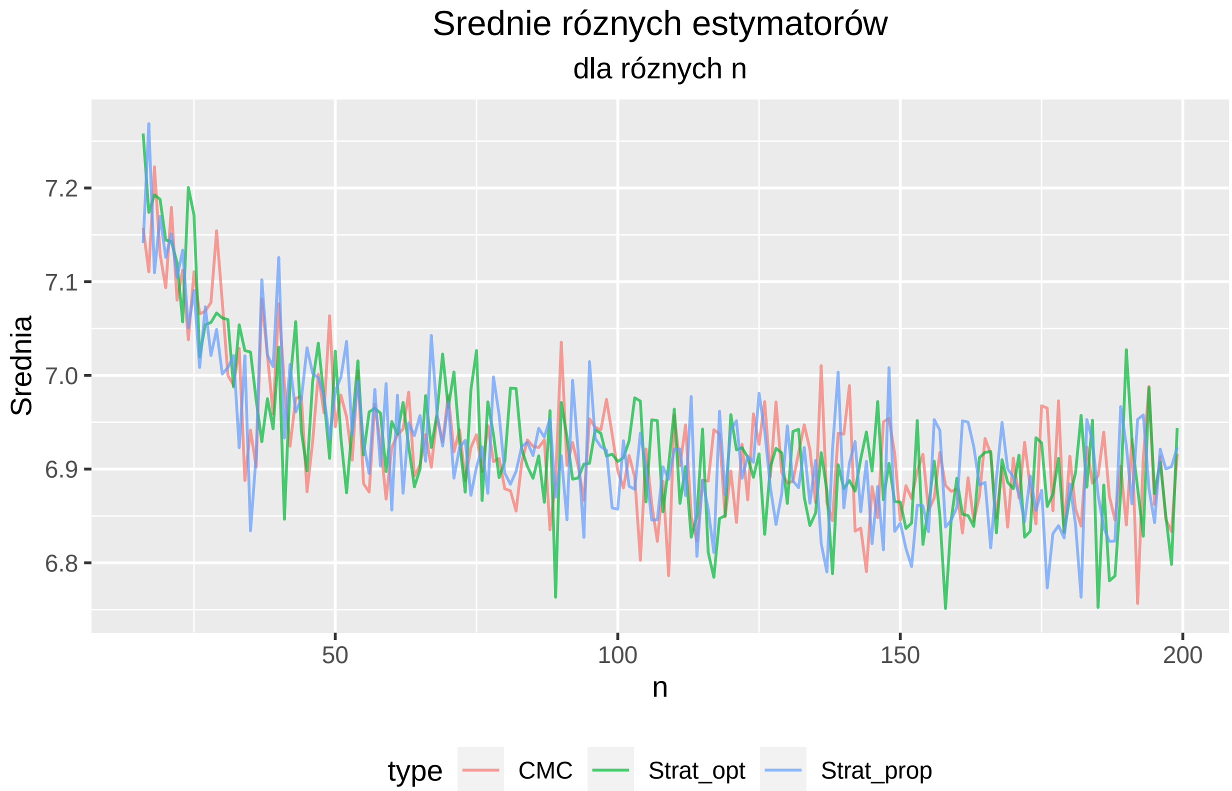


Wykres 4

W tym przypadku nie widzimy już znacząco lepszych wyników dla którego kolwiek z estymatorów, więc powinniśmy popatrzeć na wykresy średnich i wariancji.


Wariancje róznych estymatorów
dla róznych n
Wariancja
0.7
0.6
0.5
0.4
50
100
150
200
n
type
CMC
Strat_opt
Strat_prop

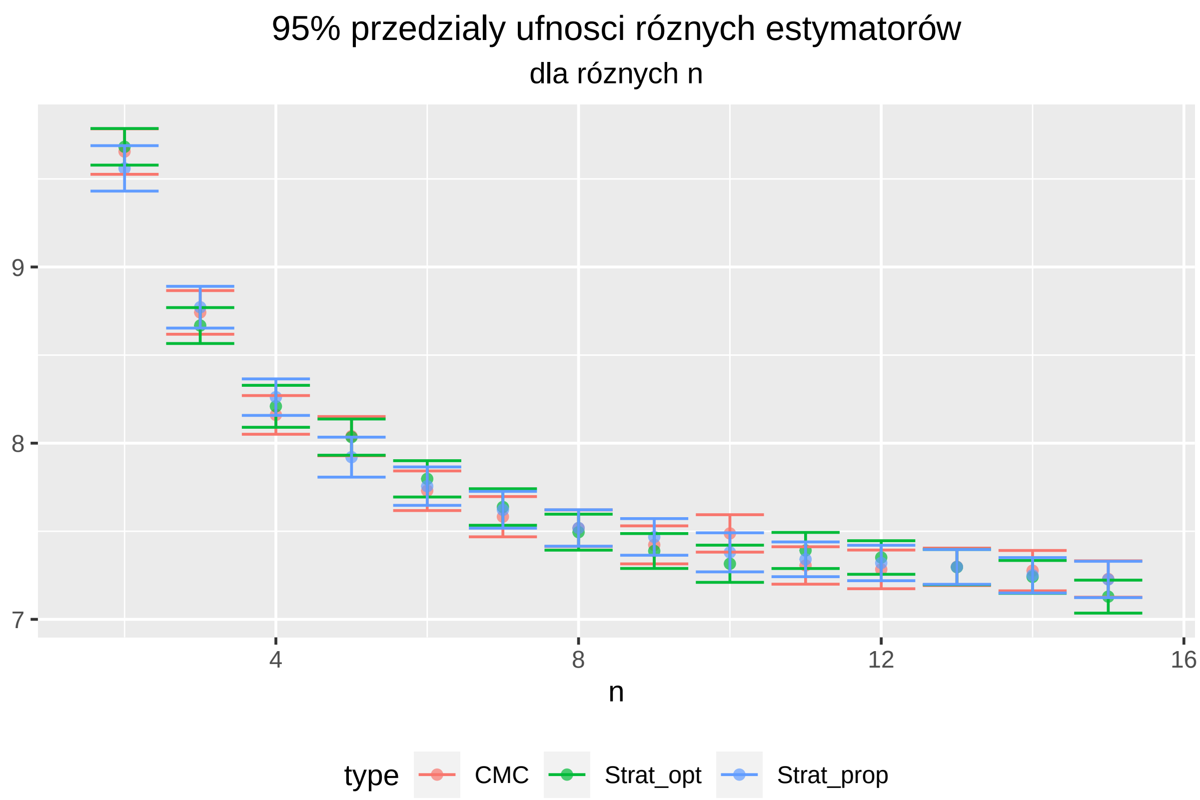
Wykres 5



Wykres 6

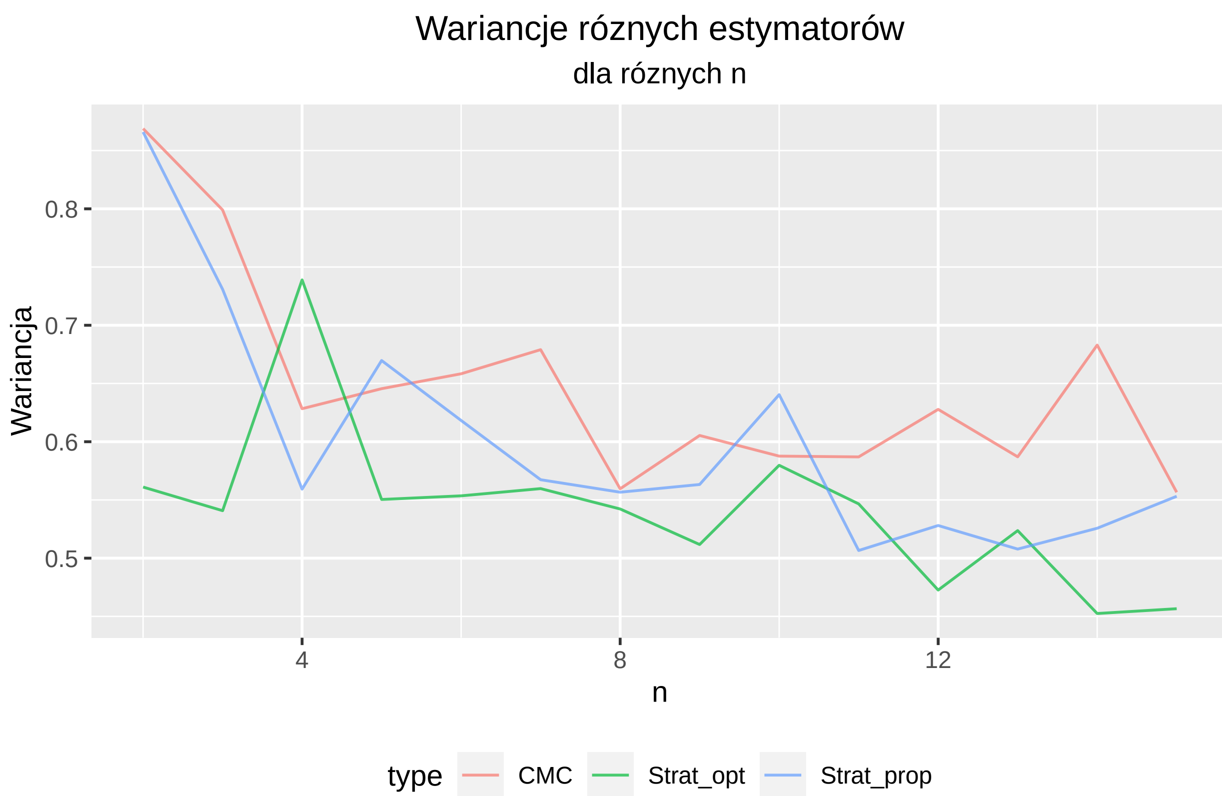
Tutaj również nie mamy znaczących różnic pomiędzy estymatorami ale ogólnie możemy zauważyć, że dla dużych n wariancja zdaje się nie zmieniać, a średnia dążdyć do ok 6,9.

Zobatrzmy jescze jak wyglądają te wykresy gdy weźmieny mnijesze n, tzn. $1 < n < 16$:

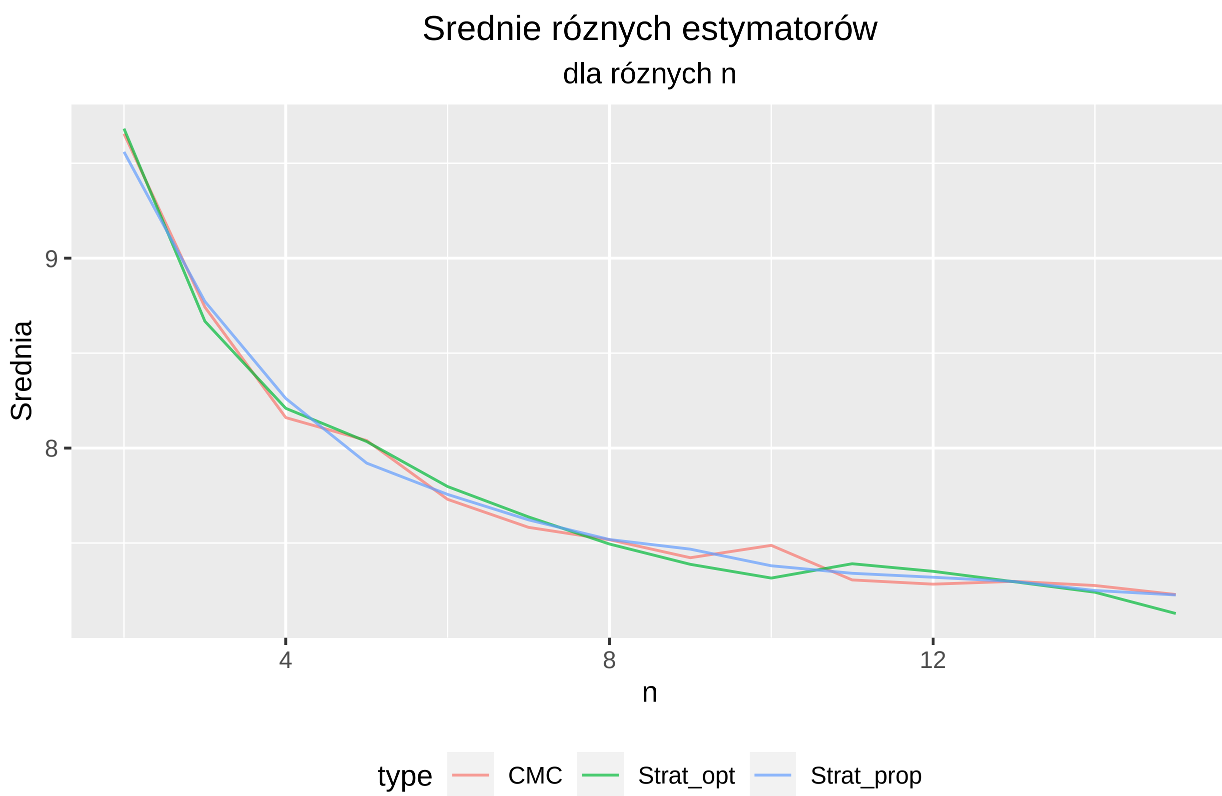


95% przedzialy ufnosci róznych estymatorów
dla róznych n
9
8
7
4
8
12
16
n
type
CMC
Strat_opt
Strat_prop

Wykres 7


Wariancje róznych estymatorów
dla róznych n
Wariancja
0.8
0.7
0.6
0.5
4
8
12
n
type
CMC
Strat_opt
Strat_prop

Wykres 8



Wykres 9

Tutaj wariancje estymatora Stratified z optymalną alokacją są mniejsze od pozostałych, jednak wartości przedziałów ufności są już bardzo podobne.

Tabela wywoływania funkcji

W tym projekcjie postanowiłem najpierw gererować liczby pseudolosowe i testować je "pierwszo poziomowo" w pythonie, później zapisywać wyniki w postaci .csv, a wykresy i second level testong przeprowadziać w R. Poniżej jest przedstwaiona tabela z nazwami funkcji, które zapisują dane to wykesu o danej nazwie.

	nazwa pliku	nazwa funkcji	parametry
Wykres 1	data_n_1.csv	zapisz_wyk1	N=500,R=500
Wykres 2	data_n_1_different_R.csv	zapisz_wyk2	Rs = np.arange(11,1000,50),N = 100
Wykres 3	data_n_1_different_R2.csv	zapisz_wyk2	Rs = np.arange(2,50,1),N = 100
Wykres 4	data_n_different.csv	zapisz_wyk3	Ns=np.arange(1,200,1),N=200,R=200
Wykres 5	data_n_different.csv	zapisz_wyk3	Ns=np.arange(1,200,1),N=200,R=200
Wykres 6	data_n_different.csv	zapisz_wyk3	Ns=np.arange(1,200,1),N=200,R=200
Wykres 7	data_n_different.csv	zapisz_wyk3	Ns=np.arange(1,200,1),N=200,R=200
Wykres 8	data_n_different.csv	zapisz_wyk3	Ns=np.arange(1,200,1),N=200,R=200
Wykres 9	data_n_different.csv	zapisz_wyk3	Ns=np.arange(1,200,1),N=200,R=200