

Praca domowa z RPIS

Tomasz Zakrzewski, tz336079

31 stycznia 2014

1 Wstęp

Praca domowa miała na celu ustalenie lepszego estymatora następującego eksperymentu:

W bibliotece jest pewna ilość książek (a). Wybieramy losowo pewne k książek i patrzymy na ich numery (wszystkie książki w bibliotece ponumerowane są od 1 do a). Ile jest książek w całej bibliotece?

Rozważane estymatory to:

- “mean”: $2 \times$ średnia wszystkich pomiarów,
- “max”: $\max \frac{n+1}{n}$, gdzie \max to maksymalny z poznanych n elementów.

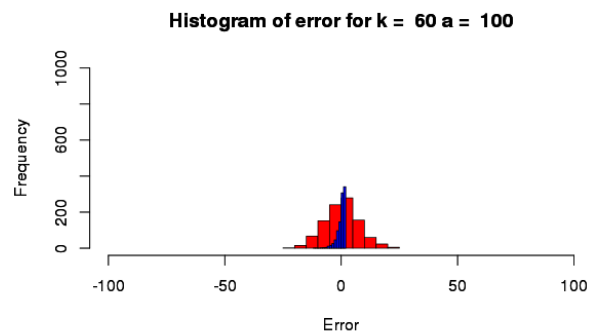
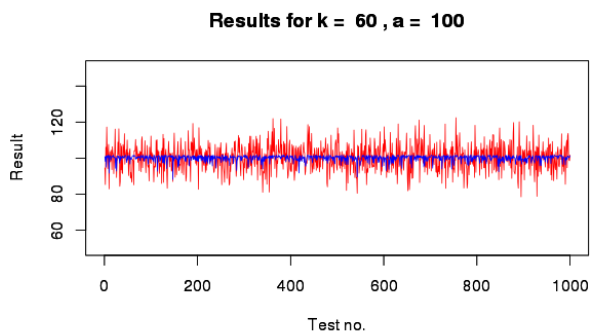
2 Eksperyment

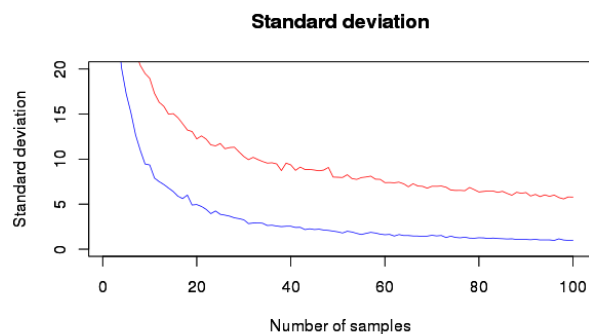
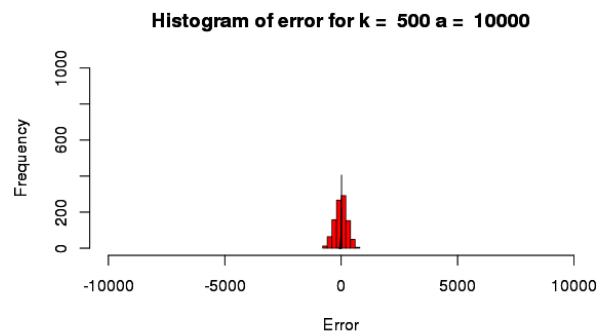
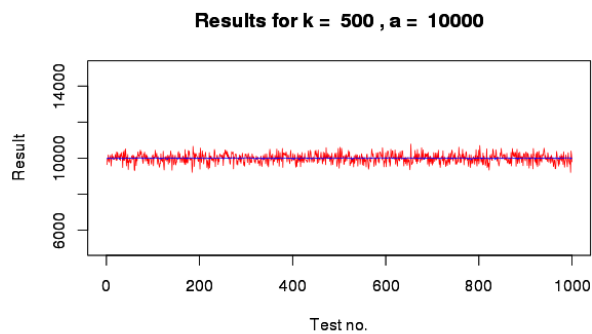
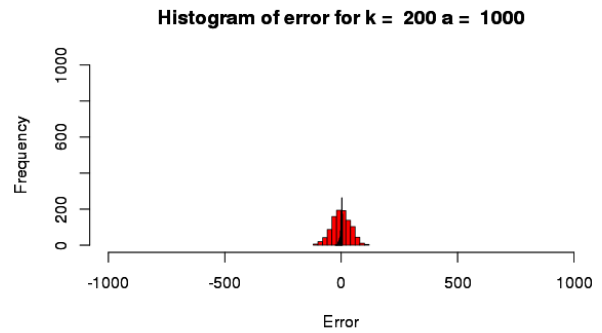
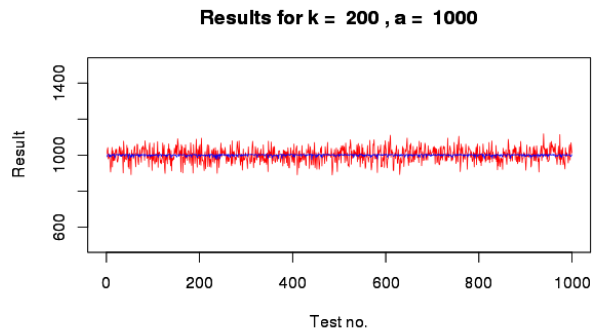
Statystyczne eksperymenty w celu wykazania wyższości jednego z dwóch proponowanych estymatorów zostały wykonane za pomocą programu R. Napisany skrypt przeprowadzał 2 pomiary:

1. W pierwszym eksperymencie, dla ustalonych k i a (wybranych arbitralnie) przeprowadzanych była pewna ilość (1000) prób. Próba polegała na wylosowaniu k próbek z rozkładu jednostajnego z przedziału $[1; a]$
2. W drugim eksperymencie pierwszy był przeprowadzany dla coraz większych wartości k i liczone było standardowe odchylenie wyniku. Wartość a wynosiła 100.

3 Wyniki

Najlepszym przybliżeniem okazało się “max”. Na poniższych wykresach próby typu 1., metoda “max” oznaczona jest kolorem niebieskim, a “mean” czerwonym. Ostatni wykres przedstawia eksperyment 2.





4 Uzasadnienie

Na wykresach pierwszego eksperymentu widać wyraźnie, że wartości uzyskiwane metodą “max” oscylują bardzo blisko faktycznego wyniku, podczas gdy “mean” ma średnio większy błąd. Ponadto, odchylenie standardowe w eksperymencie 2. maleje bardzo szybko dla “max” i już dla ilości próbek stanowiących ok. 20% całkowitej ilości badanych obiektów, jest mniejsze niż 5.

5 Uwagi

Ciekawą obserwacją jest, że ze względu na wzór metody “max”, wyniki uzyskiwane nią bardzo rzadko przekraczają szukaną wartość i można dosyć bezpiecznie przyjąć, że metoda ta osiąga prawie zawsze wartości mniejsze lub równe oczekiwanemu wynikowi, podczas gdy “mean” może wyraźnie zawyżać wynik.