

Raport 3

Aleksandra Niedziela

2023-05-28

Problem kolekcjonera kuponów

Aby wygrać w konkursie potrzebujemy n kuponów. Interesuje nas liczba pudełek X_n , po których zakupie otrzymamy nagrodę. Liczba X_n to zmienna losowa. Przyjmijmy, że mamy już $k - 1$ kuponów, niech $X_{n,k}$ będzie liczbą pudełek, które musimy kupić, aby posiadać k kuponów.

- $X_{n,k}$ to zmienna losowa, o rozkładzie geometrycznym z parametrem $\frac{n-k+1}{n}$
- $X_n = X_{n,1} + X_{n,2} + \dots + X_{n,n} = \sum_{k=1}^n X_{n,k}$

Policzmy teraz wartość oczekiwaną zmiennej X_n

$$E[X_n] = \sum_{k=1}^n \frac{n}{n-k+1} = n \sum_{k=0}^{n-1} \frac{1}{n-k} = n \sum_{k=1}^n \frac{1}{k}$$

Symulacja

Przeprowadźmy teraz symulację, za pomocą poniższej funkcji

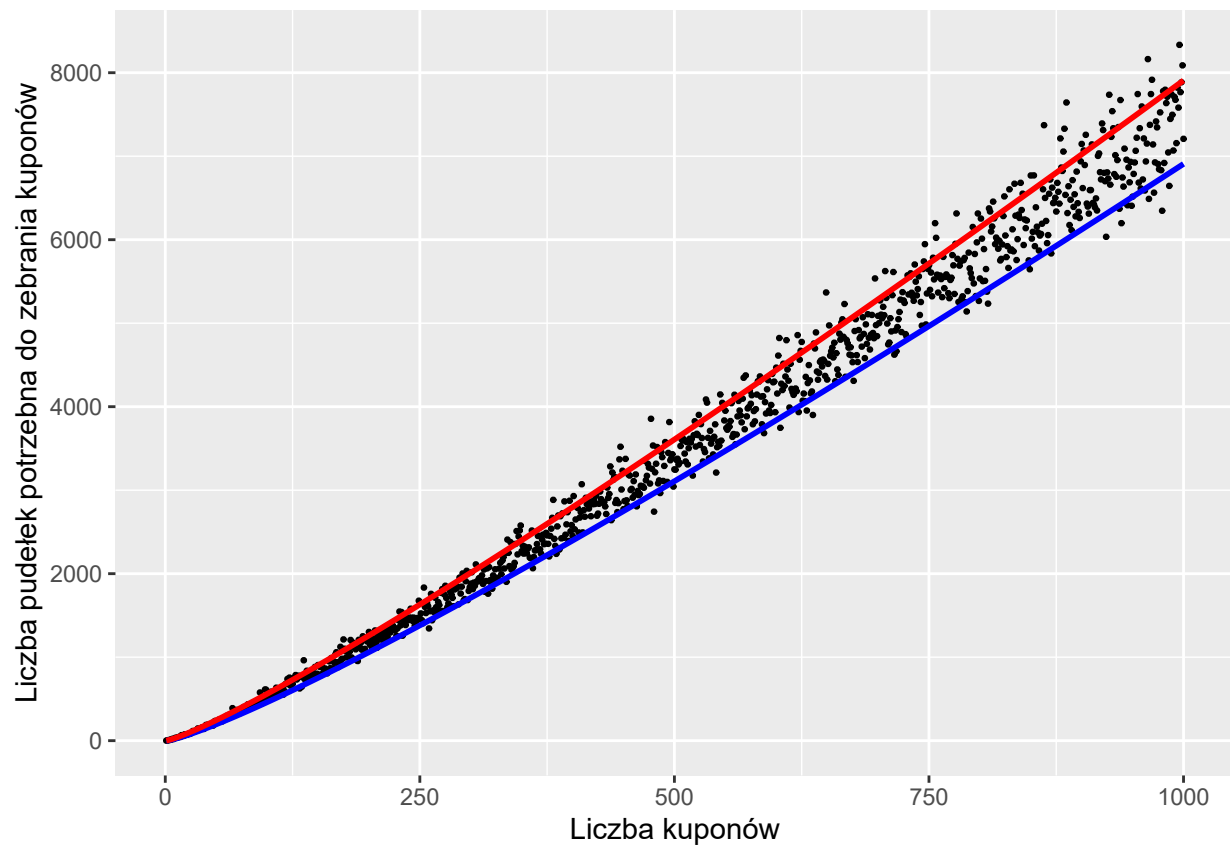
```
boxes <- function(n){  
  coupons <- 1:n  
  aquired_coupons <- numeric(n)  
  prize <- sum(1:n)  
  a <- 0  
  
  while (sum(aquired_coupons) != prize){  
    coupon <- sample(1:n, 1)  
    aquired_coupons[coupon] = coupon  
    a <- a + 1  
  }  
  return(a)  
}
```

Wyniki możemy przedstawić w tabeli:

Table 1: Porównanie wartości eksperymentalnych i teoretycznych

ilość kuponów	Wartość teoretyczna	Wartość eksperymentalna
10	29	24
25	95	115
100	519	504
500	3396	4152
1000	7485	10797

Spójrzmy teraz na wykres pokazujący średnią liczbę pudełek (10 prób), po których zakupieniu zbierzemy wszystkie kupony



Czerwona i niebieska linia oznaczają ograniczenia wynikające z faktu:

$$\ln(n) < \sum_{k=1}^n \frac{1}{k} < \ln(n) + 1$$

$$n \cdot \ln(n) < E[X_n] < n \cdot (\ln(n) + 1)$$

Nierówność Markowa

Niech X będzie zmienną losową, która przyjmuje jedynie nieujemne wartości. Wtedy dla wszystkich $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Korzystając z nierówności Markowa możemy oszacować prawdopodobieństwo uzyskania co najmniej $\frac{3n}{4}$ orłów w n rzutach monetą. Przyjmijmy za zmienną losową X_n liczbę wyrzuconych orłów w n rzutach. Widzimy, iż jest to schemat Bernoulliego, gdzie prawdopodobieństwo sukcesu (wyrzucenia orła) wynosi $\frac{1}{2}$, stąd mamy:

$$E[X_n] = \frac{n}{2}$$

Teraz z nierówności Markowa otrzymujemy:

$$P(X_n \geq \frac{3n}{4}) \leq \frac{\frac{n}{2}}{\frac{3n}{4}} = \frac{2}{3}$$

Możemy wyliczyć dokładne prawdopodobieństwa dla ustalonej ilości rzutów, a następnie zauważyć, że za każdym razem są one ograniczone przez $\frac{2}{3}$

n	prawdopodobieństwo wyrzucenia $3n/4$ orłów
4	0.31250000
10	0.17187500
50	0.00046811
100	0.00000028

Zobaczmy teraz jak nierówność Markowa sprawdza się dla rozkładu normalnego, dla różnych wartości a . Wiemy, że dla rozkładu normalnego $N(\mu, \sigma)$, $E[X] = \mu$. Weźmy $\mu = 2$ oraz $\sigma = 1$.

Porównanie oszacowania, z wartością ogona dystrybuanty

