Lista 3

Alicja Wiączkowska

2023-05-26

Zadanie 1 *Problem kolekcjonera kuponów*

Założenia:

- każde pudełko płatków zawiera jeden z n różnych kuponów,
- cel to zdobycie po jednym z każdego rodzaju kuponów,
- kupony są rozmieszczane w każdym pudełku niezależnie i losowo, z jednakowym prawdopodobieństwem $\frac{1}{n}$ dla każdej z możliwości.
- zmienna losowa X zlicza liczbę pudełek płatków, kóre trzeba kupić, aby zdobyć komplet kuponów.

Szukamy wartości oczekiwanej zmiennej losowej X , czyli odpowiedzi na pytanie ile średnio pudełek płatków należy kupić, aby zdobyć wszystkie rodzaje kuponów.

Rozwiązanie teoretyczne

Zdefiniujmy zmienne losowe X_i dla $i \in \{1, ..., n\}$ opisujące liczbę kupionych pudełek podczas gdy mieliśmy już dokładnie $i-1$ kuponów do uzyskania pierwszego losu nowego rodzaju.

Zauważmy że wówczas $X = \sum_{i=1}^{n} X_i$.

Przykład - ilustracja
dla n=4 i zbioru kuponów $\{A, B, C, D\}$ przy wylosowanym ciągu kuponów:

$$A, A, A, B, A, B, B, A, C, B, C, A, A, B, C, A, D$$

dzielimy go na kawałki i odpowiednie zliczeia przyporządkowujemy do zmiennych losowych X_i
$A, \mapsto X_1 = 1$
$A, A, B, \mapsto X_2 = 3$
$A, B, B, A, C, \mapsto X_3 = 5$
$B, C, A, A, B, C, A, D \mapsto X_4 = 8$

Zauważmy, że każdy z X_i ma rozkład geometrycznyny z parametrem $p_i = \frac{n-i+1}{n}$.

Rozwiązanie problemu:

- p_i to prawdopodobieństwo wylosowania innego rodzaju kuponu niż dotychczas w pojedynczym losowaniu pudełka (gd mamy już $i-1$ rodzajów kuponów),
- prawdopodobieństwo, że w k-tym kroku wypdnie nowy kolor wynosi:

$$\mathbb{P}(X_i = k) = p_i \cdot (1-p_i)^{k-1}$$

- **Lemat**

$$\sum_{k=1}^{\infty} x^{k-1} \cdot k = \frac{d}{dx}\sum_{k=1}^{\infty} x^k = \frac{d}{dx}\frac{x}{1-x} = \frac{d}{dx}\frac{-(1-x)+1}{1-x} = \frac{d}{dx}(-1+\frac{1}{1-x}) = \frac{1}{(1-x)^2}$$

- wartość oczekiwana pojedynczego X_i to:

$$\mathbb{E}(X_i) = \sum_{k=1}^{\infty}\mathbb{P}(X_i = k)\cdot k = \sum_{k=1}^{\infty} p_i \cdot (1-p_i)^{k-1}\cdot k = p_i \cdot \sum_{k=1}^{\infty}(1-p_i)^{k-1}\cdot k$$

- korzystając z **Lematu** podstawiając $x = 1 - p_i$ otrzymujemy

$$\mathbb{E}(X_i) = p_i \cdot \frac{1}{(1-(1-p_i))^2} = \frac{p_i}{{p_i}^2} = \frac{1}{p_i}$$

- następie wstawiając $p_i = \frac{n-i+1}{n}$ dostajemy

$$\mathbb{E}(X_i) = \frac{1}{\frac{n-i+1}{n}} = \frac{n}{n-i+1}$$

- korzystając z liniowości wartości oczekiwanej oraz faktu, że $X = \sum_{i=1}^{n} X_i$, otrzymujemy rozwiązanie problemu:

$$\mathbb{E}(X) = \sum_{i=1}^{n}\mathbb{E}(X_i) = \sum_{i=1}^{n}\frac{n}{n-i+1} = n\sum_{i=1}^{n}\frac{1}{n-i+1} = n\sum_{i=1}^{n}\frac{1}{i}$$

Symulacja

W celu zasymulowania wartości zmiennej X w zależności od parametru n tworzymy odpowiednie funkcje.

```
ten_sam<-function(wek1,wek2){
  if (length(wek1)!=length(wek2)) {return(F)}
  for(i in 1:length(wek1)){
    if(wek1[i]!=wek2[i]){return(F)}
  }
  return(T)
}

kolekcjoner1<-function(n){
  kroki=0
  rodzaje<-c(1:n)
  bufor<-sample(1:n,100,replace=T)
  kolekcja<-c()
  while(!ten_sam(kolekcja,rodzaje)){
    if(kroki==length(bufor)){
      bufor<-c(bufor,sample(1:n,50,replace=T))
    }
```

```
    kroki=kroki+1
    kolekcja<-sort(unique(bufor[1:kroki]))
  }
  return(kroki)
}

  kolekcjoner_seria<-function(n_wek,dokladnosc=100){
    odp<-numeric(length(n_wek))
    for (k in 1:length(n_wek)){
      n=n_wek[k]
      wyniki<-numeric(dokladnosc)
      for (i in 1:dokladnosc){
              wyniki[i]<-kolekcjoner1(n)
      }

      odp[k]<-mean(wyniki)
    }
    return(odp)
  }

teoretycznie<-function(n_wek){
  odp<-numeric(length(n_wek))
  for(k in 1:length(n_wek)){
  n=n_wek[k]
    i<-c(1:n)
    EX_i=n/(n-i+1)
    odp[k]<-sum(EX_i)
    }
  return(odp)
  }
```

```
  n=10*c(1:10)
symulacja<-kolekcjoner_seria(n,10)
teoria<-teoretycznie(n)
tabela<-data.frame(n,symulacja,teoria)

knitr::kable(tabela)
```

n	symulacja	teoria
10	29.6	29.28968
20	76.9	71.95479
30	122.2	119.84961
40	155.5	171.14172
50	250.0	224.96027
60	254.4	280.79222
70	325.3	338.29857
80	405.8	397.23834
90	456.9	457.43135
100	522.4	518.73775

Symulacja pokazuje, że wartości teoretyczne dobrze przybliżają oczekiwane wyniki.

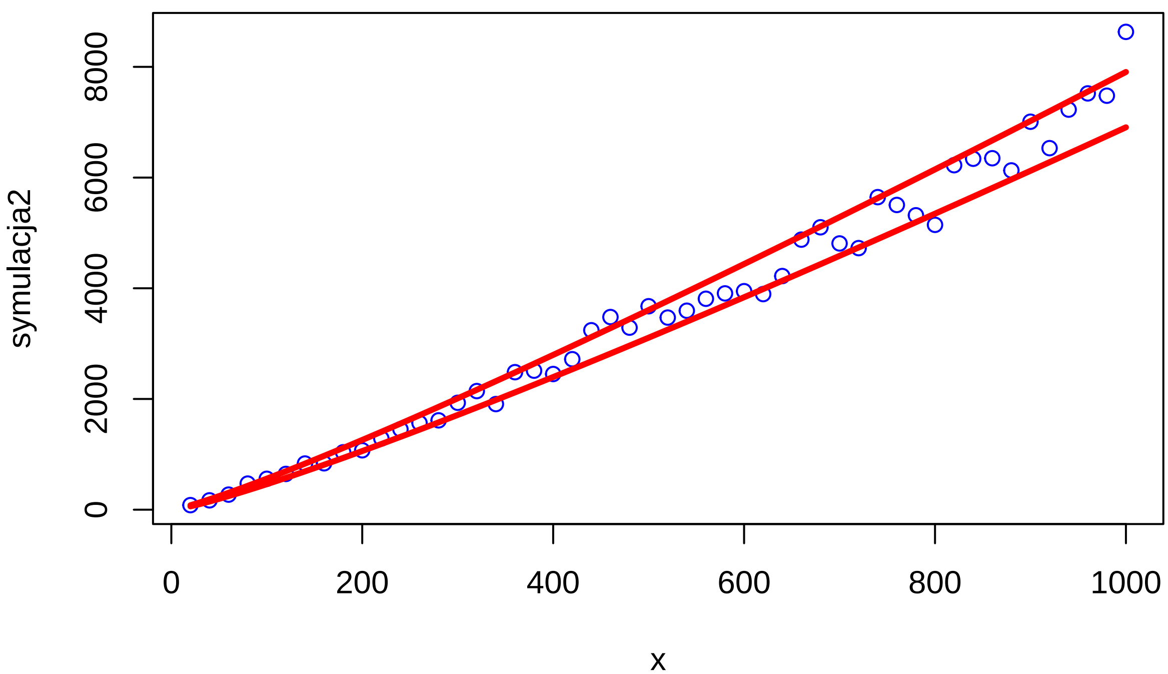
Wykres

Symulacyjnie obliczone wartości przedstawiono na wykresie, do którego dorysowano ograniczenie górne i dolne korzystając z oszacowania:

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

Zatem linie ograniczające mają wzór $n \cdot ln(n)$ i $n \cdot (ln(n) + 1)$

```
x<-20*c(1:50) #20*c(1:50)
symulacja2<-kolekcjoner_seria(x,10)
plot(symulacja2~x,col='blue')
curve(x*log(x),add = T, col='red', lwd=3)
curve((x)*(log(x)+1), add=T,col='red', lwd=3)
```



Punkty na wykresie w większości znajdują się pomiędzy liniami ograniczającymi. Istnienie obserwacji nie wpadających w ten przedział można tłumaczyć zbyt małą liczbą powtórzeń doświadczenia w symulacji - każdy punkt reprezentuje średni wynik z 10 zasymulowanych doświadczeń zliczania ilości kupionych pudełek do momentu zebrania całej kolekcji.

Zadanie 2 *Nierówność Markowa*

Twierdzenie o nierówności Markowa

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

zachodzi

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}|X|}{a}$$

Przykład z rzutem monetą

Szukamy prawdopodobieństwa uzyskania conajmniej $\frac{3n}{4}$ orłów w sekwencji n niezależnych rzutów symetryczną monetą.

Obliczenie teoretyczne

Oznaczmy przez X_i zmienną losową przyjmującą 1 przy wyrzuceniu orła w i-tym rzucie symetryczną monetą oraz 0 przy wyrzuceniu reszki. Wówczas zmienna losowa $X = \sum_{i=1}^{n} X_i$ opisuje ilość wyrzyconych orłów w n niezależnych rzutach oraz ma rozkład Bernoulliego $X \sim Bin(n, \frac{1}{2})$.

Dla ustalonego n szukna wartość to

$$\mathbb{P}(X \geq \frac{3n}{4}) = \mathbb{P}(X > \lceil\frac{3n}{4}\rceil - 1) = 1 - \mathbb{P}(X \leq \lceil\frac{3n}{4}\rceil - 1) = 1 - F_X(\lceil\frac{3n}{4}\rceil - 1)$$

gdzie $F_X(a) = \mathbb{P}(X \leq a)$ jest dystrybuantą zmiennej X o rozkładzie Bernoulliego, więc
$1 - F_X(a) = \mathbb{P}(X > a)$.

```
binom3_4<-function(n){
  q<-ceiling(3*n/4) -1
  f<-pbinom(q,n,0.5,lower.tail = F)
  return(f)
}
```

Oszacowanie nierównością Markowa

Zmienna losowa X przyjmuje tylko nieujemne wartości - można zatem użyć nierówności Markowa podstawiając $a = \frac{3n}{4}$, $\mathbb{E}(X) = \mathbb{E}|X| = \frac{n}{2}$.
Otrzymujemy

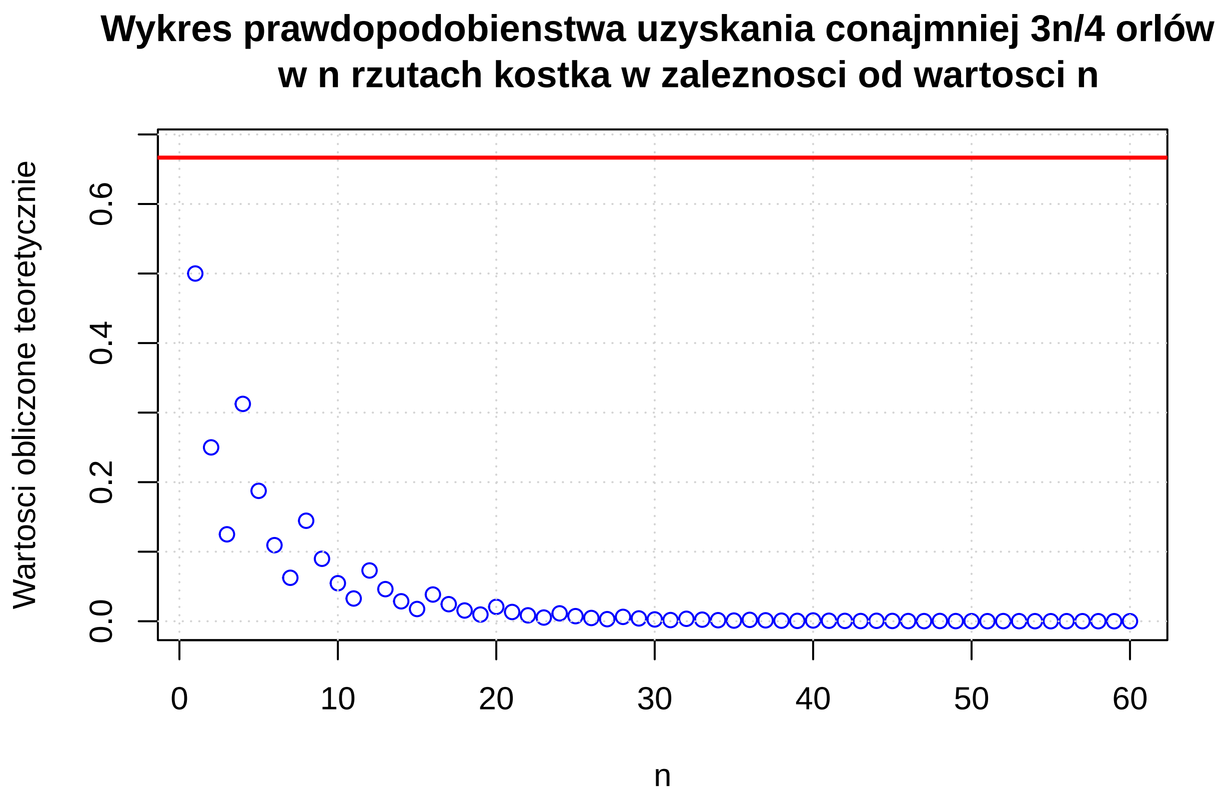
$$\mathbb{P}(X \geq \frac{3n}{4}) \leq \frac{\frac{n}{2}}{\frac{3n}{4}} = \frac{n}{2} \cdot \frac{4}{3n} = \frac{2}{3}$$

Oszacowanie Markowa w tym problemie zwraca odpowiedź $\mathbb{P}(X \geq \frac{3n}{4}) \leq \frac{2}{3}$ niezależnie od wartości n.

Porównanie wartości teoretyczntych z oszacowaniem nierównością Markowa

Poniżej przedstawiono wykres, na którym niebieskimi punktami zaznaczono wartość $\mathbb{P}(X \geq \frac{3n}{4})$ obliczoną teoretycznie.
Czerwoną linią zaznaczono oszacowanie nierównością Markowa.

```
n<- c(1:60)
y<-binom3_4(n)
plot(y~n,main="Wykres prawdopodobieństwa uzyskania conajmniej 3n/4 orłów
     w n rzutach kostką w zależności od wartości n", type='p', col="blue",
     ylim=c(0,0.68),
     ylab ="Wartości obliczone teoretycznie" )
grid()
abline(h=(2/3), col="red", lwd=2)
```



Z wykresu można odczytać, że oszacowanie nierównością Markowa jest poprawne we wszystkich przypadkach, jednak jest ono bardzo niedokładne, zwłaszcza dla dużych wartości n, dla których szukane prawdopodobieństwo jest bliskie zeru.

Przykład dla przepalających się żarówek

Zmienna losowa X przyjmuje tylko naturalne wartości i spełniony jest dla niej wzór

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

Chcielibyśmy poznać prawdopodobieństwo że w ustalonym miesiącu przepali się conajmniej dwa razw więcej żarówek niż przeciętnie.

Obliczenie teoretyczne

$$\mathbb{P}(X \geq 2\lambda) = 1 - \mathbb{P}[X \leq (2\lambda - 1)] = 1 - \sum_{k=0}^{2\lambda-1} \mathbb{P}(X = k) = 1 - \sum_{k=0}^{2\lambda-1} \frac{e^{-\lambda}\lambda^k}{k!} = 1 - F_X(2\lambda - 1)$$

gdzie $F_X(a) = \mathbb{P}(X \leq a)$ jest dystrybuantą zmiennej X o rozkładzie Poissona z parametrem λ

Oszacowanie nierównością Markowa

Zmienna losowa X przyjmuje tylko wartości nauralne - można zatem użyć nierówności Markowa podstawiając $a = 2\lambda$, $\mathbb{E}|X| = \mathbb{E}(X) = \lambda$.

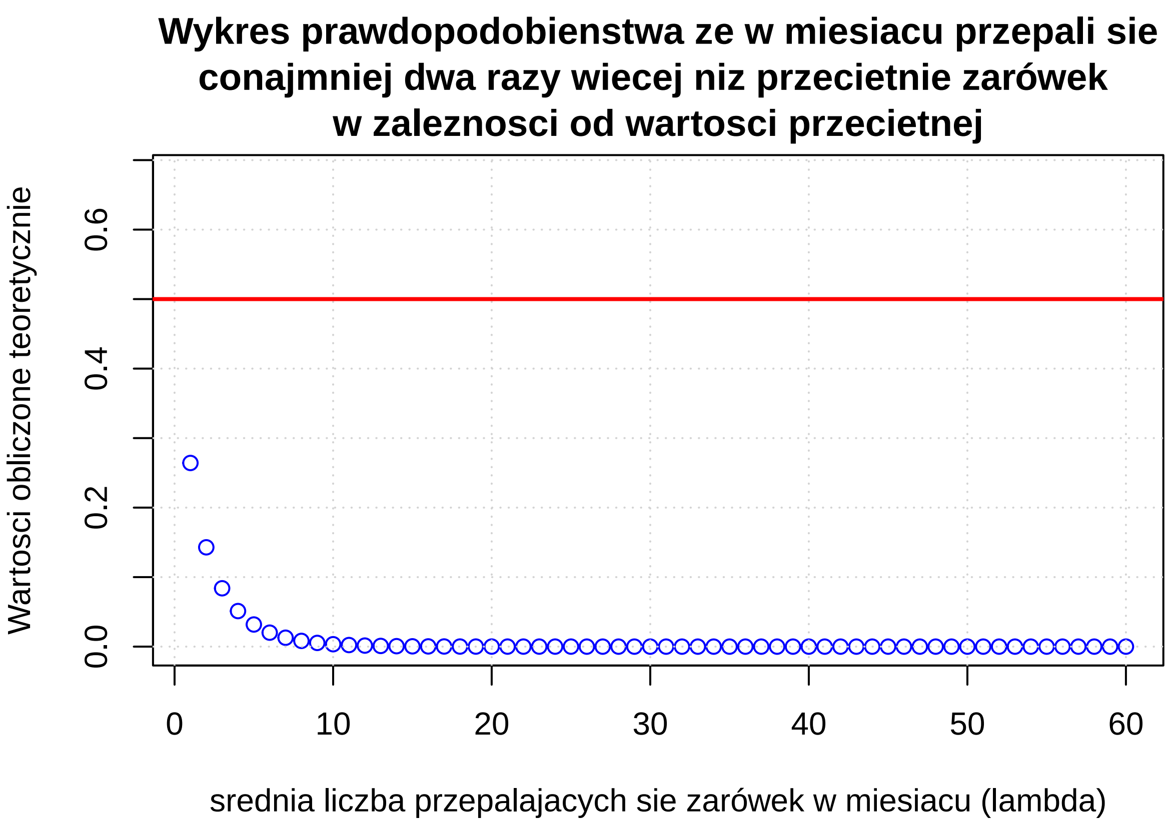
Otrzymujemy

$$\mathbb{P}(X \geq 2\lambda) = \frac{\mathbb{E}|X|}{2\lambda} = \frac{\lambda}{2\lambda} = \frac{1}{2}$$

Porównanie wartości teoretycznych z oszacowaniem nierównością Markowa

Poniżej przedstawiono wykres, na którym niebieskimi punktami zaznaczono wartość $\mathbb{P}(X \geq 2\lambda)$ obliczoną teoretycznie.
Czerwoną linią zaznaczono oszacowanie nierównością Markowa.

```
n<- c(1:60)
y<-ppois(2*n-1,n,lower.tail = F)
plot(y~n,main="Wykres prawdopodobieństwa że w miesiącu przepali się
conajmniej dwa razy więcej niż przeciętnie żarówek
w zależności od wartości przeciętnej", type='p', col="blue", ylim=c(0,0.68),
     xlab="średnia liczba przepalających się żarówek w miesiącu (lambda)",
     ylab ="Wartości obliczone teoretycznie" )
grid()
abline(h=(0.5), col="red", lwd=2)
```



Wnioski z tego przykładu są podobne do rzutów monetą. Również w tym przypadku oszacowanie nierównością Markowa jest poprawne bardzo niedokładne, zwłaszcza dla dużych wartości n, dla których szukane prawdopodobieństwo jest bliskie zeru.

Zadanie 3 *Nierówność Czebyszewa*

Twierdzenie o nierówności Czebyszewa *Dla dowolnego* $a > 0$ *zachodzi nierówność*

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{Var(X)}{a^2}$$

Nierówność Czebyszewa a Markowa

Przykład z rzutem monetą

Szukamy prawdopodobieństwa uzyskania conajmniej $\frac{3n}{4}$ orłów w sekwencji n niezależnych rzutów symetryczną monetą. Oznaczamy przez X zmienną losową zliczającą ilość orłów w sekwencji n rzutów.

Z obiczeń teoretycznych wynika że dla ustalonego n szukna wartość to

$$\mathbb{P}(X \geq \frac{3n}{4}) = 1 - F_X(\lceil \frac{3n}{4} \rceil - 1)$$

gdzie $F_X(a) = \mathbb{P}(X \leq a)$ jest dystrybuantą zmiennej X o rozkładzie Bernoulliego, więc
$1 - F_X(a) = \mathbb{P}(X > a)$.

Z powodu symetryczności monety prawdopodobieństwo wyrzucenia orła to $p = \frac{1}{2}$. Wartość oczekiwania dla rozkładu Bernoulliego to $\mathbb{E}(X) = np = \frac{n}{2}$, a wariancja $Var(X) = np(1-p) = \frac{n}{4}$

Oszacowanie nierównością Czebyszewa W celu oszacowania wartości $\mathbb{P}(X \geq \frac{3n}{4})$ możemy skorzystać z nierówności

$$\mathbb{P}(X \geq \frac{3n}{4}) = \mathbb{P}(X \geq \frac{n}{4} + \frac{n}{2}) = P(X - \frac{n}{2} \geq \frac{n}{4}) = \mathbb{P}(X - \mathbb{E}X \geq \frac{n}{4}) \leq \mathbb{P}(|X - \mathbb{E}X| \geq \frac{n}{4})$$

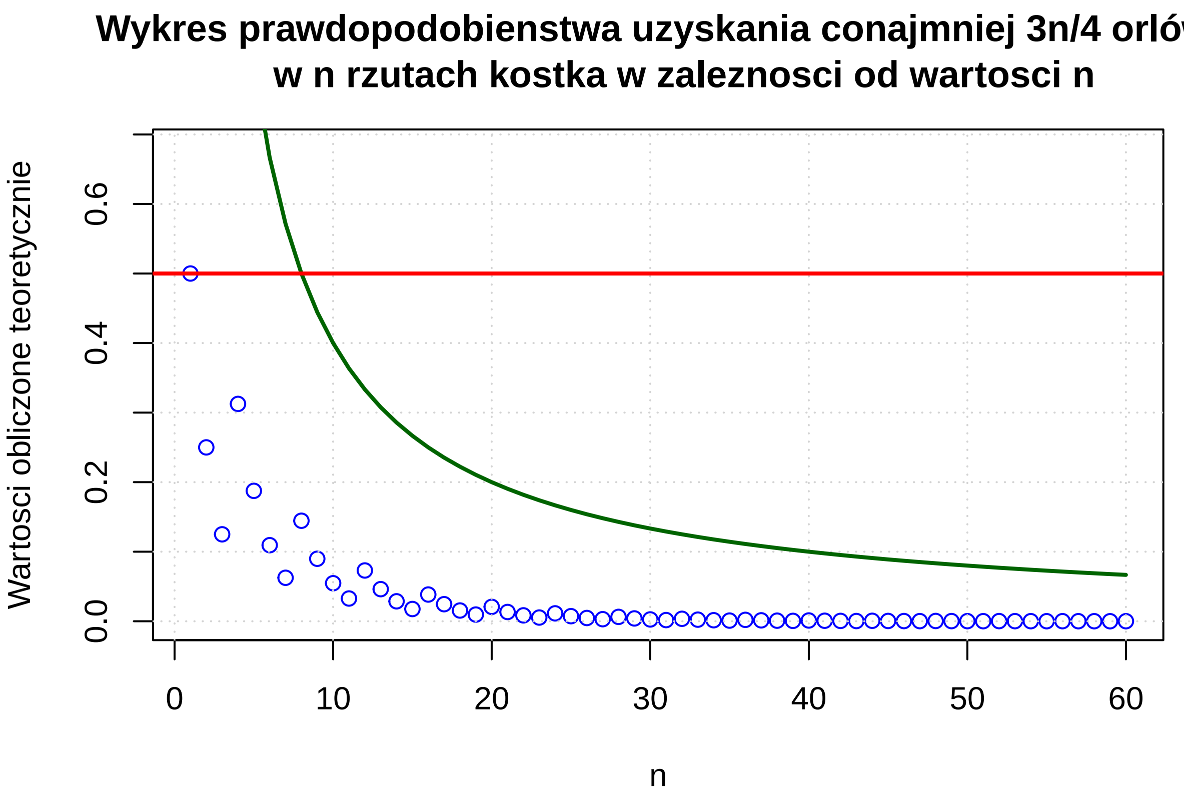
Dla nierówności $\mathbb{P}(|X - \mathbb{E}X| \geq \frac{n}{4})$ możemy użyć nierówności Czebyszewa podstawiając $a = \frac{n}{4}$, $Var(X) = \frac{n}{4}$.
Otrzymujemy

$$\mathbb{P}(X \geq \frac{3n}{4}) \leq \mathbb{P}(|X - \mathbb{E}X| \geq \frac{n}{4}) \leq \frac{\frac{n}{4}}{(\frac{n}{4})^2} = \frac{4}{n}$$

Oszacowanie Markowa w tym problemie zwraca odpowiedź $\mathbb{P}(X \geq \frac{3n}{4}) \leq \frac{2}{3}$ niezależnie od wartości n.

Porównanie wartości teoretycznych z oszacowaniem nierównością Czebyszewa i Markowa
Poniżej przedstawiono wykres, na którym niebieskimi punktami zaznaczono wartość $\mathbb{P}(X \geq \frac{3n}{4})$ obliczoną teoretycznie.
Czerwoną linią zaznaczono oszacowanie nierównością Markowa, a zieloną oszacowanie Czebyszewa.

```r
n<- c(1:60)
y<-binom3_4(n)
plot(y~n,main="Wykres prawdopodobieństwa uzyskania conajmniej 3n/4 orłów
     w n rzutach kostką w zależności od wartości n", type='p', col="blue",
     ylim=c(0,0.68),
     ylab ="Wartości obliczone teoretycznie" )
grid()
lines(n,4/n , col="darkgreen",type='l', add=T, lwd=2)
abline(h=(0.5), col="red", lwd=2)
```



Oszacowanie Czebyszewa również poprawnie ogranicza wartości prawdopodobieństwa, ale znacznie dokładniejsze niż Markowskie dla n większych niż 8. W tym przypadku zauważyć można również, że nierówność Czebyszewa zwraca oszacowanie zależne od parametru n, w przeciwieństwie do oszacowania MArkowa.

Przykład dla przepalających się żarówek

Chcielibyśmy poznać prawdopodobieństwo że w ustalonym miesiącu przepali się conajmniej dwa razw więcej żarówek niż przeciętnie.
Przez X oznaczamy zmienną losową opisującą liczbę żarówek, które się przepaliły w danym miesiącu.
Z obiczeń teoretycznych wynika że dla ustalonego n szukna wartość to

$$\mathbb{P}(X \geq 2\lambda) = 1 - F_X(2\lambda - 1)$$

gdzie $F_X(a) = \mathbb{P}(X \leq a)$ jest dystrybuantą zmiennej X o rozkładzie Poissona z parametrem λ

Wartość oczekiwana dla zmiennej X, która w tym przypadku ma rozkład Poissona z parametrem λ wynosi $\mathbb{E}X = \lambda$ oraz wariancja $Var(X) = \lambda$.

Oszacowanie nierównością Czebyszewa W celu oszacowania wartości $\mathbb{P}(X \geq \frac{3n}{4})$ możemy skorzystać z nierówności

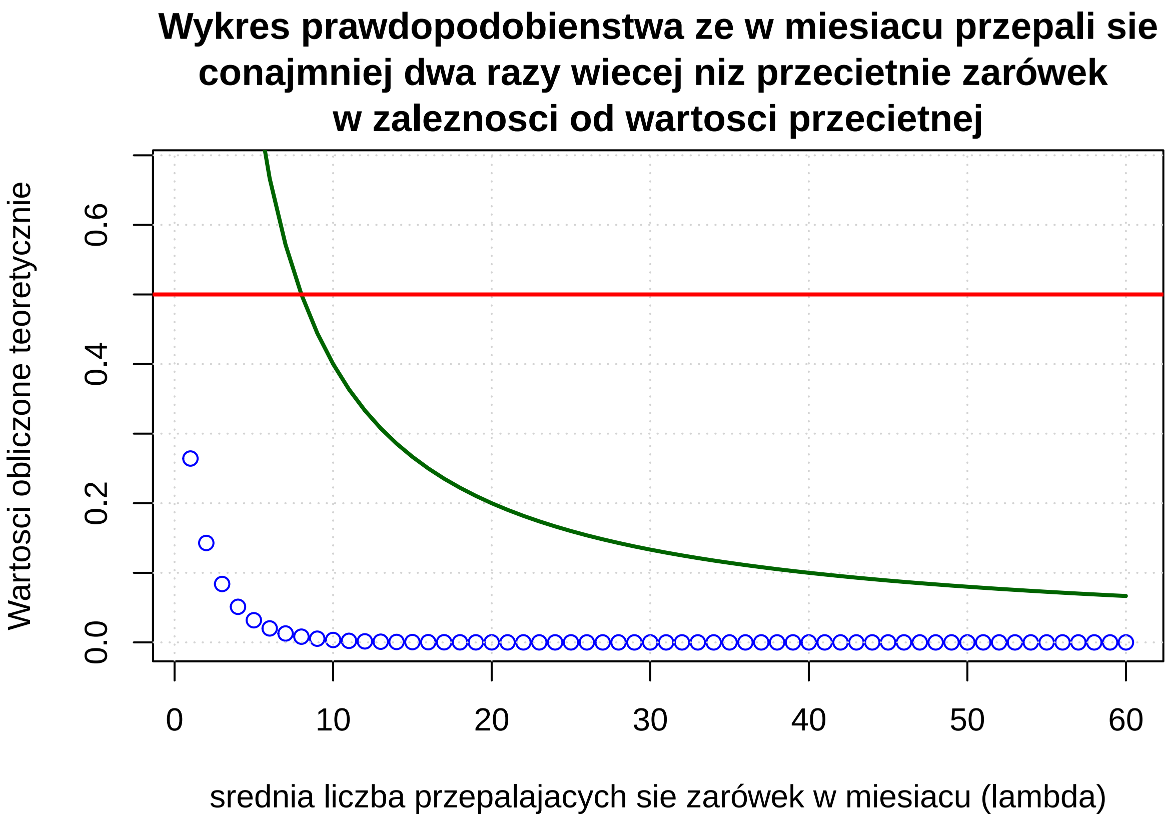
$$\mathbb{P}(X \geq 2\lambda) = \mathbb{P}(X - \lambda \geq \lambda) \leq \mathbb{P}(|X - \lambda| \geq \lambda)$$

Następnie $\mathbb{P}(|X - \lambda| \geq \lambda)$ możemy oszacować nierównością Czebyszewa. Podstawiając $a = \lambda$, $Var(X) = \lambda$ otrzymujemy

$$\mathbb{P}(X \geq 2\lambda) \leq \mathbb{P}(|X - \lambda| \geq \lambda) \leq \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

Porównanie wartości teoretycztnych z oszacowaniem nierównością Czebyszewa i Markowa Poniżej przedstawiono wykres, na którym niebieskimi punktami zaznaczono wartość $\mathbb{P}(X \geq 2\lambda)$ obliczoną teoretycznie.
Czerwoną linią zaznaczono oszacowanie nierównością Markowa, a zieloną oszacowanie Czebyszewa.

```
n<- c(1:60)
y<-ppois(2*n-1,n,lower.tail = F)
plot(y~n,main="Wykres prawdopodobieństwa że w miesiącu przepali się
conajmniej dwa razy więcej niż przeciętnie żarówek
w zależności od wartości przeciętnej", type='p', col="blue", ylim=c(0,0.68),
     xlab="średnia liczba przepalających się żarówek w miesiącu (lambda)",
     ylab ="Wartości obliczone teoretycznie" )
grid()
lines(n,4/n , col="darkgreen",type='l', add=T, lwd=2)
abline(h=(0.5), col="red", lwd=2)
```



W przypadku rozkładu Poissona nierówność Czebyszewa również zwraca lesze ograniczenie prawdopodobieństwa niż nierówność Markowa dla $\lambda \leq 8$. Również w tym przypadku Nierówność Markowa jest stała dla wszystkich parametrów λ, a Czebyszewa od tego parametru zależy, co pozwala uzyskać lepszą dokładność.

Szacowanie zmiennej losowej z problemu kolekcjonera

Chcielibyśmy poznać wartość $\mathbb{P}(X \geq 2nH_n)$, gdzie H_n jest n-tą liczbą harmoniczną $H_n = \sum_{i=1}^{n} \frac{1}{i}$. Biorąc pod uwagę, że $\mathbb{E}(X) = nH_n$, szukamy prawdopodobieństwa, że zmienna losowa X przyjmie wartość równą

$2 \cdot \mathbb{E}(X)$ - czyli poszukujemy prawdopodobieństwa, że chcąc zebrać wszystkie n rodzajów kuponów kupimy 2 razy więcej opakowań płatków niż przeciętnie powinniśmy.

Szacowanie nierównością Markowa

Zmienna losowa X przyjmuje tylko wrtości dodatnie, więc użycie nierówności Markowa jest uzasadnione. OTrzymujemy wówczas dla $a = 2nH_n$,$E|X| = nH_n$

$$\mathbb{P}(X \geq 2nH_n) \leq \frac{nH_n}{2nH_n} = \frac{1}{2}$$

Szacowanie nierównością Czebyszewa

Do oszacowania szukanego prawopodobieństwa możemy również wykorzystać nierówność Czebyszewa, która powinna zwrócić dokładniejszy wynik niż nierówność Markowa. Jak w zadaniu pierwszym zmienną losową X możemy zapisać jako sumę $X = \sum_{i=1}^{n} X_i$ gdzie X_i opisuje liczbę kupionych pudełek podczas gdy mieliśmy już dokładnie $i-1$ kuponów do uzyskania pierwszego losu nowego rodzaju. Zauważmy, że składowe zmienne są od siebie niezależne. Możemy więc zapisać, że

$$Var(X) = \sum_{i=1}^{n} Var(X_i)$$

Zauważmy, że skoro X_i ma rozkład gepoetryczny z parametrem $p_i = \frac{n-i+1}{n}$, to zachodzi

$$Var(X_i) = \frac{1-p_i}{{p_i}^2} = \frac{1-\frac{n-i+1}{n}}{(\frac{n-i+1}{n})^2} = \frac{\frac{i-1}{n}}{(\frac{n-i+1}{n})^2} = \frac{i-1}{n} \cdot \frac{n^2}{(n-i+1)^2} = \frac{n(i-1)}{(n-i+1)^2}$$

biorąc pod uwagę, że i przebiega $1, 2, ..., n$, można oszacować $i-1 \leq n$. Mamy zatem

$$Var(X_i) = \frac{n(i-1)}{(n-i+1)^2} \leq \frac{n^2}{(n-i+1)^2} = (\ \frac{n}{n-i+1}\ )^2$$

W takim razie wariancję zmiennej losowej X możemy oszacować następująco

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = Var(X) \leq \sum_{i=1}^{n} (\ \frac{n}{n-i+1}\ )^2 = n^2 \sum_{i=1}^{n} \frac{1}{i^2} = n^2 \cdot \frac{\pi^2}{6}$$

Oszacowanie szukanego prawdopodobieństwa nierównością Czebyszewa prezentuje się następująco

$$\mathbb{P}(X \geq 2nH_n) = \mathbb{P}(X - nH_n \geq nH_n) \leq \mathbb{P}(|X - nH_n| \geq nH_n) = \mathbb{P}(|X - \mathbb{E}X| \geq nH_n) \leq \frac{Var(X)}{(nH_n)^2} \leq \frac{\frac{\pi^2 n^2}{6}}{n^2 {H_n}^2} = \frac{\pi^2}{6 \cdot {H_n}^2}$$

Porównanie oszacowań

Na poniższym wykresie przedstawiono wartości oszacowań prawdopodobieństwa $\mathbb{P}(X \geq 2nH_n)$ nierównością Markowa (zaznaczone kolorem czerwonym) oraz nierównością Czebyszewa (oznaczone kolorem zielonym).

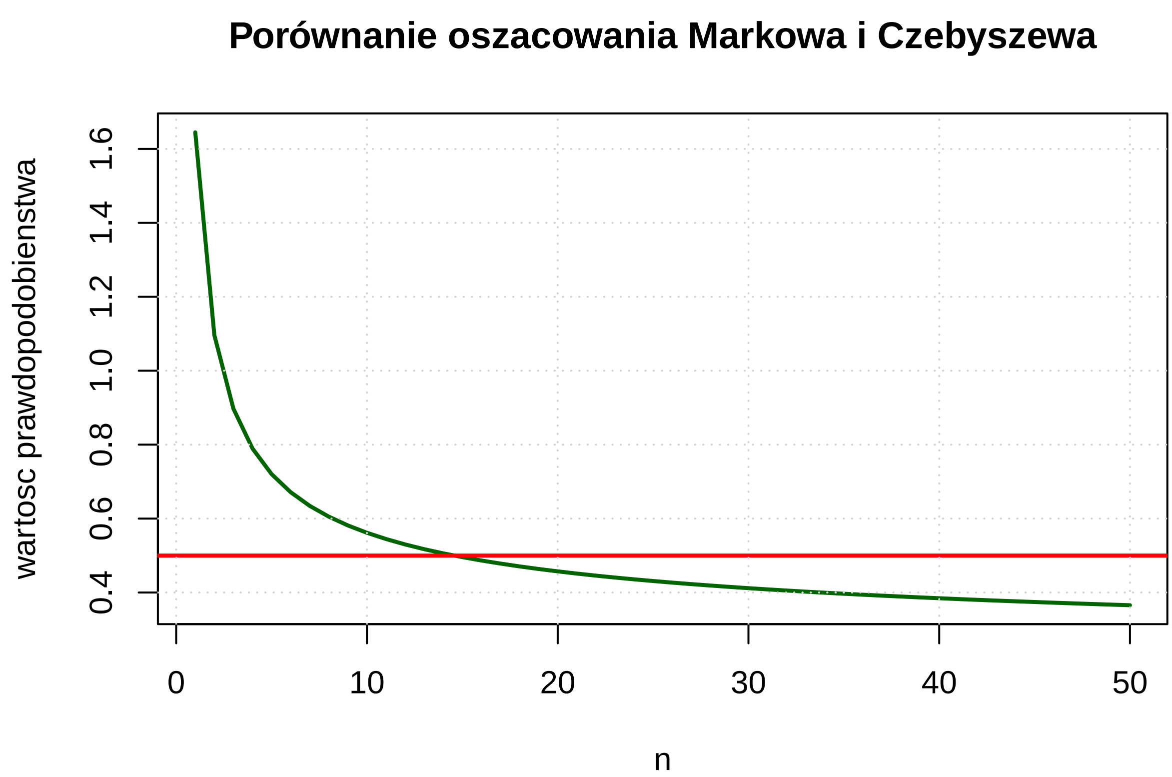
```
H<-function(n_wek){
  odp<-numeric(length(n_wek))
  for(i in 1:length(n_wek)){
    n=n_wek[i]
    v<-c(1:n)
```

```
    odp[i]<-sum(1/v)
  }
  return(odp)
}

czebysz<-function(n){
  return(pi**2 /(6*H(n)) )
}

n<- c(1:50)
czeb<-czebysz(n)

plot(n,czeb,main='Porównanie oszacowania Markowa i Czebyszewa',
     col='darkgreen',lwd=2,type='l',ylab="wartość prawdopodobieństwa")
abline(h=0.5,col='red',lwd=2)
grid()
```




```
n<- c(10:20)
czeb<-czebysz(n)
```

Obie nierówności stanowią ograniczenie górne szukanego prawdopodobieństwa. Ograniczenie jest tym lepsze im mniejszą wartość przyjmuje.
Oszacowanie prawdopodobieństwa przy pomocy nierówności Czebyszewa jest dokładniejsze dla $n \geq 15$.

Zadanie 4 *Średnia, dyspersja i mediana*

Średnia to wartość oczekiwana zmiennej losowej. **Dyspersja** to odchylenie standardowe, czyli pierwiastek z wariancji. **Mediana** to taka wartość m, dla której ilości obserwacji powyżej m i poniżej m są sobie równe.
Twierdzenie o odległośc mediany od średniej
Jeśli X jest zmienną losową o średniej μ odchyleniu standardowym σ i medianie m, wtedy

$$|\mu - m| \leq \sigma$$

Zbadajmy prawdziwość tego twierdzenia gdy średnia, dyspersja i mediana zostaną zastąpione estymatorami. W tym celu wygenerujemy próby 5-,10- i 50-elementowe z rozkładu jednostajnego (na zbiorze [0,1]), standardowego normalnego, Poissona ($\lambda = 10$), wykładniczego ($\lambda = 1$) i geometrycznego($p = 0.1$), a następnie sprawdzimy prawdziwość tezy.

```
tabelka<-function(n){
  unif<-runif(n)
  norm<-rnorm(n)
  pois<-rpois(n,10)
  exp<-rexp(n)
  geom<-rgeom(n,0.1)

  head<-c('unif','norm','pois','exp','geom')

  mi<-c(mean(unif),mean(norm),mean(pois),mean(exp),mean(geom) )
  m<-c(median(unif),median(norm),median(pois),median(exp),median(geom) )
  sd<-c(sd(unif),sd(norm),sd(pois),sd(exp),sd(geom) )

  teza<-( abs(mi-m)<=sd )

  d<-data.frame(mi,m,sd,teza)
  rownames(d)<-head
  return(d)
}


t1<-tabelka(5)
t2<-tabelka(10)
t3<-tabelka(50)


knitr::kable(t1,caption='n=5')
```

Table 2: n=5

	mi	m	sd	teza
unif	0.5042755	0.3120381	0.3878577	TRUE
norm	0.2119374	0.1254073	0.8602179	TRUE
pois	10.0000000	9.0000000	3.8078866	TRUE
exp	0.9449745	1.0117914	0.3244337	TRUE
geom	3.6000000	1.0000000	5.5045436	TRUE

```
knitr::kable(t2,caption='n=10')
```

Table 3: n=10

	mi	m	sd	teza
unif	0.4974566	0.3833067	0.3365658	TRUE
norm	0.0437118	0.0853568	0.8487416	TRUE
pois	11.5000000	12.0000000	3.9228674	TRUE
exp	0.9592560	0.8401580	0.6409132	TRUE
geom	6.8000000	3.0000000	9.5777520	TRUE

```
knitr::kable(t3,caption='n=50')
```

Table 4: n=50

	mi	m	sd	teza
unif	0.5390475	0.5143217	0.3157357	TRUE
norm	0.0996180	0.0964333	1.1809218	TRUE
pois	10.0600000	10.0000000	3.5016614	TRUE
exp	0.8226869	0.6530844	0.7369980	TRUE
geom	10.0600000	6.0000000	10.2248397	TRUE

W zaprezentowanych symulajach teza twierdzenia jest spełniona również dla estymatorów.