Lista 2

Alicja Wiączkowska

2023-04-28

Zadanie 2 *Problem urodzin*

Załóżmy, że na sali wykładowej jest obecnych n = 25 osób z jednego roku. Jakie jest prawdopodobień-stwo, że dwoje lub więcej z nich ma tą samą datę urodzin?

Dodatkowe założenia:
- Rok ma tylko 365 dni.
- Liczba urodzeń jest równomiernie rozłożona w ciągu roku.
- Osoby w pokoju są losowo wybrane.

a) Intuicja

Około 30%

b) Obliczenia teoretyczne

Przyjmijmy oznaczenia:

- A - zdarzenie że dwoje lub więcej z nich ma tą samą datę urodzin
- A^c - zdarzenie że każda z osób ma unikalną datę urodzin
- Ω - wszystkie możliwe wariacje dat urodzin
  Skorzystamy ze wzoru:
  $$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$
  Obliczamy zatem:
  $$\mathbb{P}(A^c) = \frac{|A^c|}{|\Omega|} = \frac{\frac{365!}{(365-n)!}}{365^n} = \frac{365 \cdot 364 \cdot 363 \cdot ... \cdot (365-n+1)}{365^n} =$$
  $$= \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ ... \ \cdot \frac{(365-n+1)}{365} = \prod_{i=0}^{n-1} \frac{365-i}{365}$$

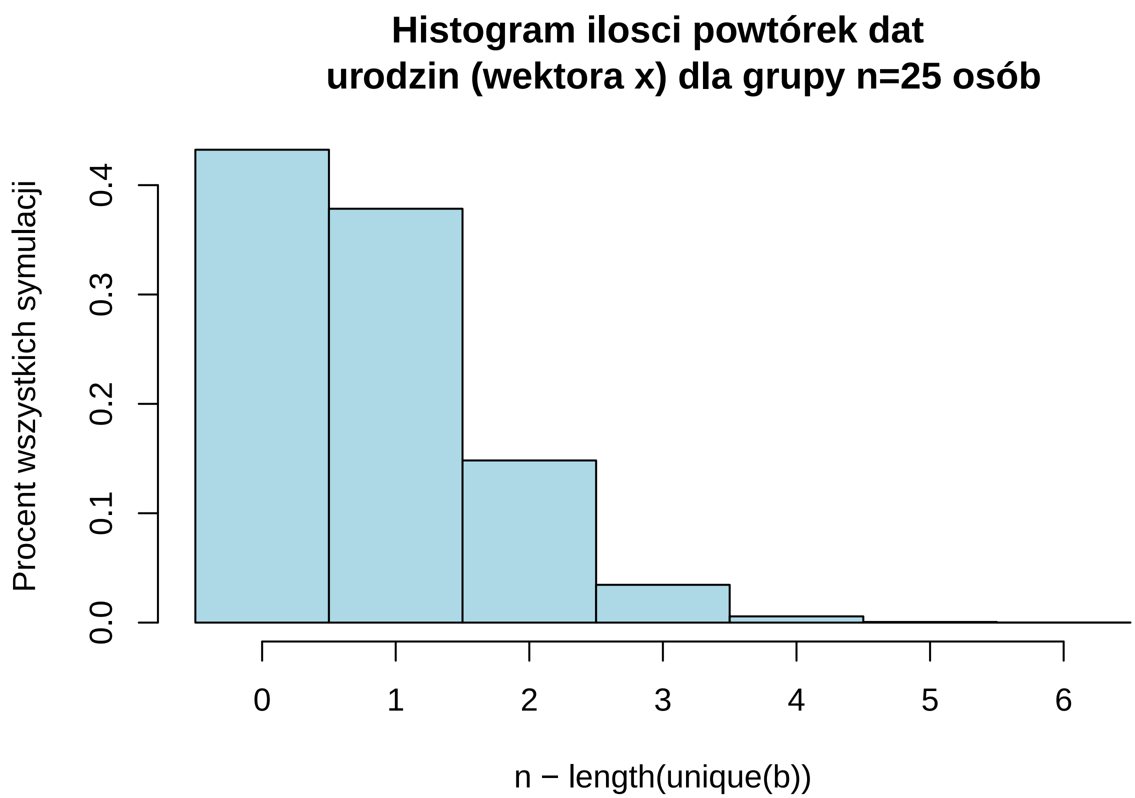
Dla n=25 otrzymujemy (z dokładnością do 2 miejsc po przecinku):
$A^c \approx 43.13\%$ $\quad$ $A \approx 56.87\%$

Zaskakującym wnioskiem może być fakt, że już przy n=25 prawdopodobieństwo znaleznienia $\geqslant 2$ osób mających urodziny tego samego dnia jest bliska 60%.

Zadanie 3 *Problem urodzin - symulacja*

Przeprowadzamy symulację dla n=25 osób polegającą na powtórzeniu m = 100000 razy doświadczenia wylosowania 25 dat urodzin (do wektora b) i zapisaniu w odpowiedniej komórce wektora x różnicy `n - length(unique(b))` (funkcja `unique(b)` zwraca wektor wartości z wektora b bez poworzeń). Wartość ta jest równa zeru tylko w sytuacji gdy każda z osób ma urodziny innego dnia.

```
#m = 100000; n = 25
birthday<-function(n,m = 100000,days=365,prob=F){
  x = numeric(m) #pusty wektor
for (i in 1:m){
  b = sample(1:days, n, repl=T)
  x[i] = n - length(unique(b))
}
  if(prob)return(mean(x >  0))
  return(x)
}
x=birthday(25,100000)
cutp = (0:(max(x)+1)) - .5
hist(x, breaks=cutp, prob=T, col="lightblue",
     main="Histogram ilości powtórek dat
     urodzin (wektora x) dla grupy n=25 osób ",xlab="n - length(unique(b))",
     ylab="Procent wszystkich symulacji")
```



Jak można odczytać z histogramu wektora x, prawdopodobieństwo zdarzenia A^c (dla słupka 0) wynosi nieco ponad 40% (a dokładnie 43.24%), co w przybliżeniu zgadza się z wyliczeniami teoretycznymi. Ponadto obliczone symulacyjnie $\mathbb{P}(A) = 56.59\%$.

Zauważmy jednak, że funkcja `unique` zlicza jedynie ilość wystąpienia "powtórek". Nie liczy poprawnie ilości tych samych dat urodzin w wektorze b. Nie jest też odpowiednia do zliczenia osób mających "bliźniaka".

Kontrprzykład

Weźmy n=6. Rozważmy sytuacje gdy wypadło:

1. `b = (1,1,1,2,3,4)`
   Wówczas mamy **3** osoby mające kogoś z kim dzielą dzień urodzin ("trojaczki") oraz **1** dzień który się powtarza.
   Tymczasem `unique(b)=(1,2,3,4)` oraz `6 - length(unique(b)) = 2` (powtórkami są dwie jedynki).
2. `b = (1,1,2,2,3,4)`
   Wówczas mamy **4** osoby maające kogoś z kim dzielą dzień urodzin (2 pary "bliźniaków") oraz **2** dni, które się powtarzają.
   W tym przypadku również zachodzi `unique(b)=(1,2,3,4)` oraz `6 - length(unique(b)) = 2` (powtórką jest jedna jedynka i jedna dwójka).

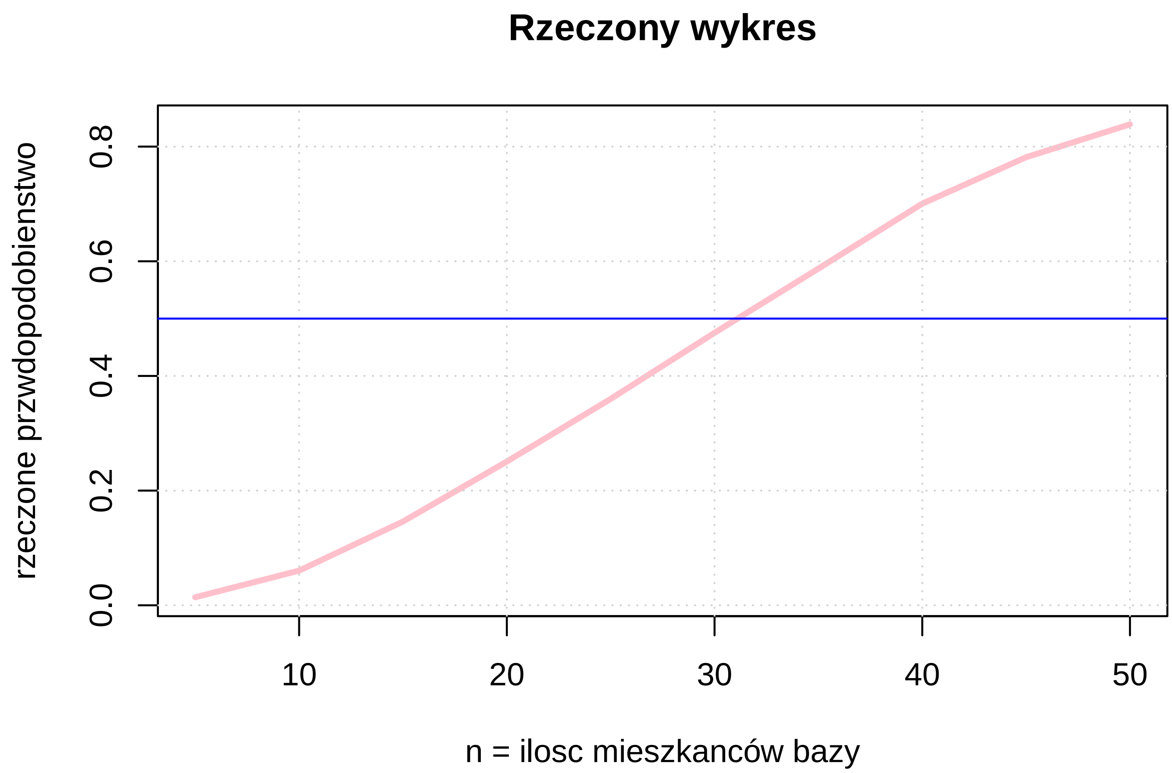
Zadanie 4 *Urodziny na Marsie*

Założenia: - na Marsie rok trwa 669 dni, czyli 12 miesięcy, - populacja Marsa składa się wyłącznie z ludzi urodzonych na Marsie, - urodziny w populacji urodzonej na Marsie są równomiernie rozłożone przez cały rok.

a) Wykres obrazujący zależność prawdopodobieństwa zdarzenia, że dwoje lub więcej osób mieszkających wspólnie w bazie marsjańskiej obchodzi urodziny w tym samym dniu w zależnościod liczby n osób mieszkających wspólnie.

Do rozwiązania tego zadania można posłużyć się funkcją `birthday` napisaną w poprzednim zadaniu.

```
m=10
y<-numeric(m)
for(i in 1:m){
  y[i]<-birthday(n=5*i,days=669,prob=T, m=10000)
}
x=5*c(1:m)
plot(y~x, main="Rzeczony wykres", xlab='n = ilość mieszkańców bazy',
     ylab="rzeczone przwdopodobieństwo",type='l', lwd=3,col="pink")
grid()
abline(h=0.5, col="blue")
```



b) Prawdopodobieństwo $> 50\%$

Z powyższego wykresu wynika, że prawdopodobieństwo zdarzenia, że dwoje lub więcej osób mieszkających wspólnie w bazie marsjańskiej obchodzi urodziny w tym samym dniu przekroczy 50% na pewno dla n $\geq$ 35 (ale powinno wystarczyć dla liczby nieco większej niż 30 - nie można dokładniej stwierdzić z wykresu, gdyż wartości na osi odciętych występują co 5 jednostek), bowiem dla n=30 symulacyjne prawdpodobieństwo wyniosło 48%.

Zadanie 5 *Niejednostajny rozkład dat urodzin*

Załóżmy, że wskaźnik urodzeńprzez 65 dni w roku, na przykład w środku lata, jest trzykrotnością wskaźnika urodzeń dla pozostałychdni roku.

a) Prawdopodobieństwo urodzenia się w środku lata

B - zdarzenie urodzenia się w jeden z 65 dni w środku lata

```
p=c(rep(3,65),rep(1,300),0.25)
p=p/sum(p) # żeby prawdopodobieństwa sumowały się do 1
```

Przybliżone prawdopodobieństwo zdarzenia B wynosi 39.37%.

b) Ilość koalescencji

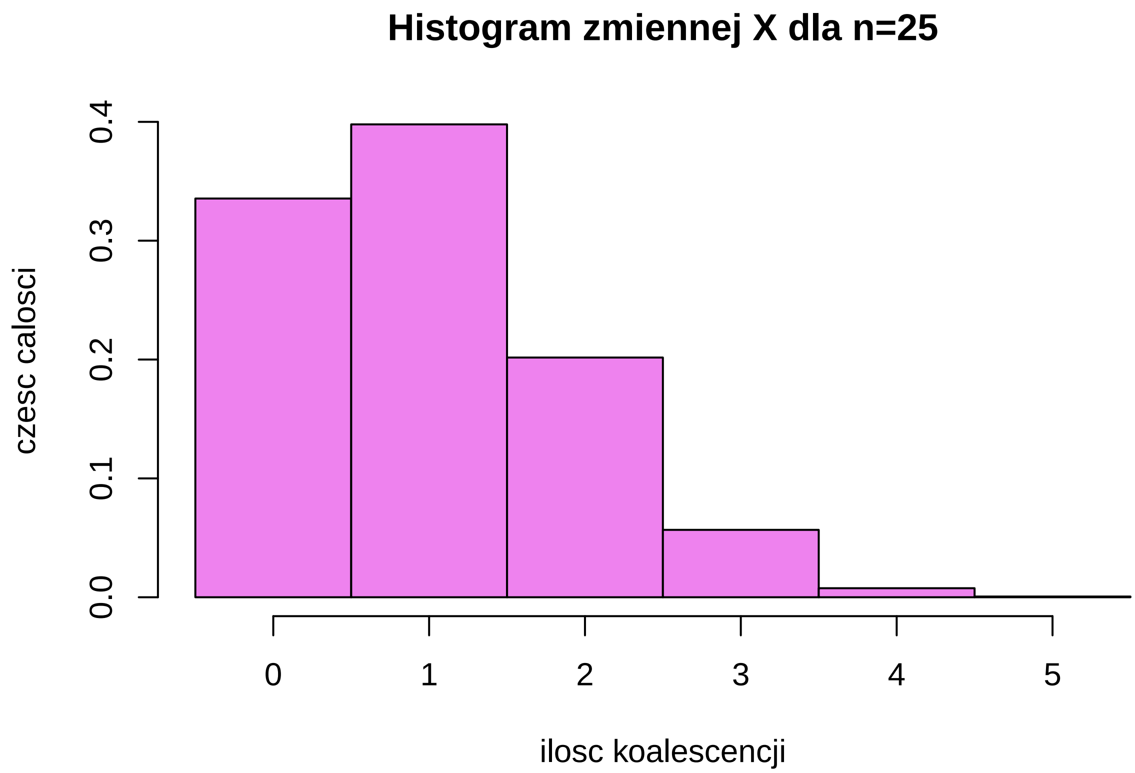
X - zmienna losowa opisaująca liczbę koalescencji (czyli dat w roku, w których ≥ 2 osoby obchodzą urodziny) wśród n=25 osób w pokoju.

Szukamy $\mathbb{P}(\{X=1\})$ oraz $\mathbb{E}(X)$. Do przeprowadzenia symulacji posłużymy się funkcją `birthday2`.

```
birthday2<- function(bday,days=366){
  #funkcja zwraca ilość dni, w których urodziło się >= 2
  x=numeric(days)
  for(i in 1:length(bday)){
    # do wektora x zapisujemy ilość osób mających urodziny w dniu x[i]
    x[bday[i]]= x[bday[i]] +1
  }
  return(length(x[x>=2]))
}

# ta funkcja działa dla **kontrprzykładu**
# birthday2(c(1,1,1,2,3),6)
# birthday2(c(1,1,2,2,3),6)

p=c(rep(3,65),rep(1,300),0.25)
p=p/sum(p)

m=10000
koalescencje<-function(days=366, probs=c(rep(3,65),rep(1,300),0.25) ){
  m=10000;n=25
  koal<-numeric(m)
  year<-c(1:days)
  for(i in 1:m){
    x<-sample(year,n,replace=T, prob=probs)
    koal[i]=birthday2(x)
  }
  return(koal)
}
```

Zmienna losowa X ma rozkład przybliżony poniższym histogramem zdudowanym na podstawie symulacji.

```
x<-koalescencje()
cutp = (0:(max(x)+1)) - .5
hist(x, breaks=cutp, prob=T, main="Histogram zmiennej X dla n=25",
     xlab="ilość koalescencji", ylab="część całości", col="violet")
```



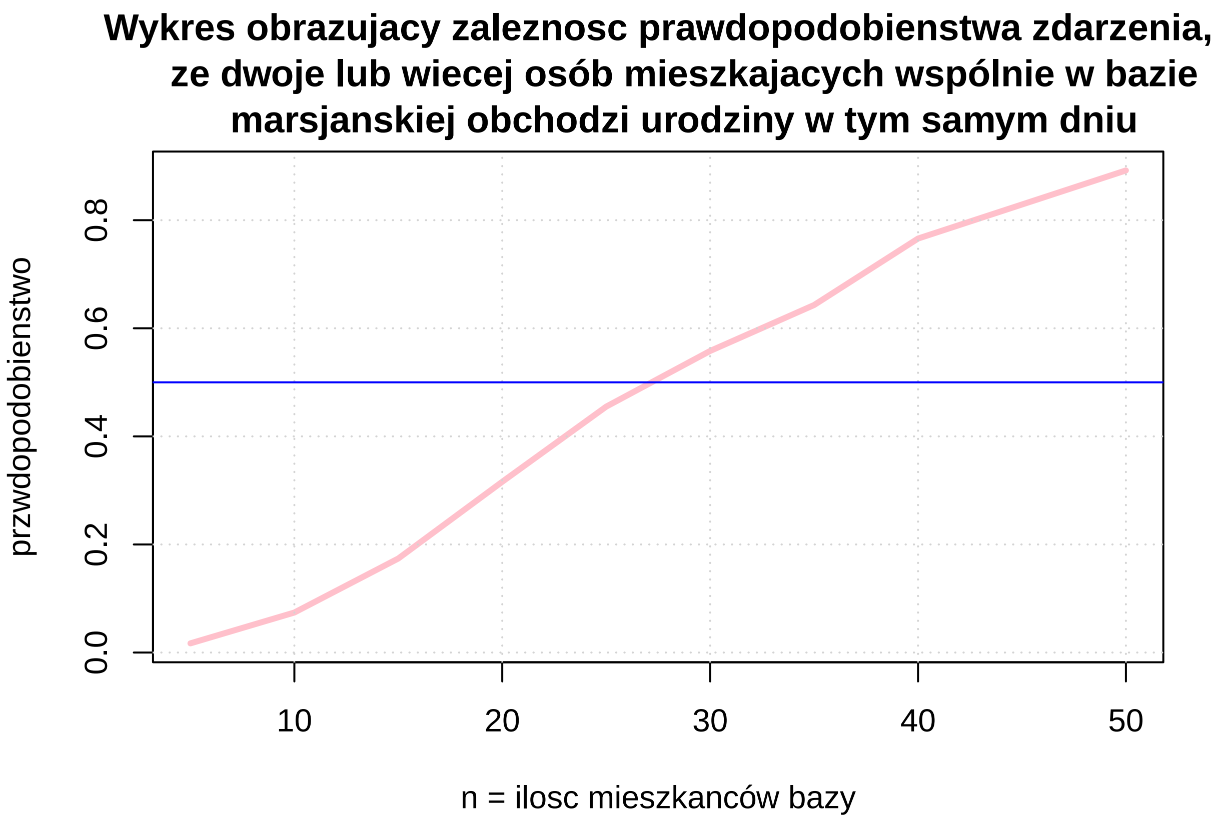
Jak można odczytać z histogramu, prawdopobieństwo $\mathbb{P}(X \geq 1)$ wynosi 66%. Symulacyjnie obliczona wartość oczekiwana zmiennej X wynosi 1.005.

c) Niejednostajny rozkład urodzin na Marsie

Zakładamy, że prawdopodobieństwo urodzenia się w pierwsze 100 dni w roku jest 3 razy wyższe niż w przeciętny dzień marsjański; ponadto kolejne 200 dni roku marsjańskiego jest 2 razy bardziej prawdopodobne pozostałe przeciętne dni.

```
p=c(rep(3,100),rep(2,200), rep(1,369))

birthday3<-function(n,m = 1000,days=669,
                    probs=c(rep(3,100),rep(2,200), rep(1,369))){
  x = numeric(m) #pusty wektor
for (i in 1:m){
  b = sample(1:days, n, repl=T, prob=probs)
  x[i] = n - length(unique(b))
}
  return(mean(x >  0))
}


m=10
y<-numeric(m)
for(i in 1:m){
  y[i]<-birthday3(n=5*i)
```

```
}
x=5*c(1:m)
plot(y~x, main="Wykres obrazujący zależność prawdopodobieństwa zdarzenia,
     że dwoje lub więcej osób mieszkających wspólnie w bazie
     marsjańskiej obchodzi urodziny w tym samym dniu",
     xlab='n = ilość mieszkańców bazy',
     ylab="przwdopodobieństwo",type='l', lwd=3,col="pink")
grid()
abline(h=0.5, col="blue")
```



Ze sporządzonego wykresu (na którym ilość mieszkańców wzrasta co 5) można wywnioskować, że w przypadku takiego rozkładu urodzeń prawdopodobieństwo wystąpienia koalescencji na pewno przeroczy 50% dla n=30 (być może próg ten da się obniżyć, ale potrzebny byłby do tego dokładniejszy wykres).