Laboratorium 4

MAT4 - Stosowany rachunek prawdopodobieństwa

Wersja: 2023-10-25

1 Dystrybuanta empiryczna

Mając daną próbkę $x_1, \ldots, x_n$ z rozkładu o nieznanej dystrybuancie F, możemy w przybliżeniu odtworzyć F za pomocą tzw. **dystrybuanty empirycznej**, czyli funkcji F_n danej wzorem

$$F_n(t) := (\#\text{ obserwacji nie większych niż } t)/(\#\text{ wszystkich obserwacji}) = \frac{\#\{1 \leq j \leq n : x_j \leq t\}}{n}.$$

W R, zakładając, że próbka $x_1, \ldots, x_n$ zawarta jest w wektorze x, wykres jej dystrybuanty empirycznej można dostać pisząc

```
plot(ecdf(x))
```

1. Narysować wykresy dystrybuant empirycznych dla prób z rozkładu dwupunktowego (jednostajnego na punktach 1 i 2), kolejno dla $n = 5, 10, 10^2, 10^3, 10^6$.
2. Przeprowadzić $n = 10^2$ powtórzeń doświadczenia polegającego na rzucaniu symetryczną monetą do pierwszego momentu wypadnięcia orła. Stworzyć wektor v o długości n zawierający numery rzutów dających orła w poszczególnych powtórzeniach. Narysować wykres dystrybuanty empirycznej dla wektora v. Powtórzyć powyższe kroki dla $n = 10^6$.
3. Wykonać polecenia z poprzedniego punktu dla doświadczenia polegającego na wyborze punktu z odcinka $[0, 1]$ (zgodnie z rozkładem jednostajnym/prawdopodobieństwem geometrycznym).

Wstęp - rozkłady prawdopodobieństwa w R

Są zasadniczo cztery komendy do pracy z rozkładami prawdopodobieństwa w R. Ich nazwy składają się z nazwy danego rozkładu (patrz poniżej), poprzedzonej prefiksem `d`, `p`, `q` lub `r`. Znaczenie prefiksów:

- `d` - gęstość (dla rozkładów ciągłych) lub tzw. *funkcja prawdopodobieństwa* dla rozkładów dyskretnych,
- `p` - dystrybuanta rozkładu,
- `q` - kwantyl,
- `r` - wartość zmiennej losowej.

Niektóre z R-owych nazw rozkładów prawdopodobieństwa:

- `binom` - r. dwumianowy,
- `geom` - r. geometryczny,
- `pois` - r. Poissona,
- `unif` - r. jednostajny,
- `exp` - r. wykładniczy,
- `norm` - r. normalny.

Przykłady:

```
runif(6,0,1) # generuje szesc liczb pseudolosowych z rozkladu jednostajnego na (0,1)
```

```
## [1] 0.4535542 0.8667854 0.8598653 0.6109778 0.6102278 0.1965111
```

```r
pexp(3,1) # P(X<=3) dla X o rozkladzie wykladniczym z parametrem lambda=1
```

```
## [1] 0.9502129
```

```r
dbinom(10,20,0.5) # P(X=10) dla X o rozkladzie dwumianowym z parametrami n=20 i p=0.5
```

```
## [1] 0.1761971
```

```r
qnorm(0.95,0,1) #kwantyl rzędu 0.95 z rozkładu normalnego standardowego
```

```
## [1] 1.644854
```

2 Aproksymacja poissonowska

1. Wygenerować próbki z rozkładu Poissona z różnymi parametrami. Obejrzeć histogramy. Ocenić wpływ parametru λ na kształt histogramu. Ocenić wielkość ogonów. Jak duże (średnio) są elementy próbki w zależności od λ?

2. Dla ustalonych $n \in \mathbb{N}$ i $p \in (0,1)$ stworzyć macierz o trzech kolumnach i $b-a+1$ wierszach ($a, b \in \{0,1,2,\ldots\}, a < b$). W pierwszej kolumnie umieścić wartości

$$\mathbb{P}(X = k),\ k = a, \ldots, b,$$

gdzie $X \sim bin(n,p)$. W drugiej kolumnie umieścić

$$\mathbb{P}(Y = k),\ k = a, \ldots, b,$$

dla $Y \sim Poiss(np)$. Trzecia kolumna powinna być różnicą dwóch pierwszych. Stworzoną macierz wyświetlić, podając jednocześnie maksymalny moduł wartości z trzeciej kolumny.

3. Wykonać powyższe polecenia dla

- $n = 10$, $p = 0.9$, $a = 0$, $b = 10$,
- $n = 100$, $p = 0.9$, $a = 0$, $b = 100$,
- $n = 1000$, $p = 0.001$, $a = 0$, $b = 100$,

3 Czekanie na autobus

1. Wygenerować próbki z rozkładu wykładniczego z różnymi parametrami. Obejrzeć histogramy. Ocenić wpływ parametru λ na kształt histogramu. Ocenić wielkość ogonów. Jak duże (średnio) są elementy próbki w zależności od λ?

2. Do pewnego przystanku autobusy przybywają w losowych momentach czasu. Odstępy między kolejnymi autobusami są niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z parametrem $\lambda = \frac{1}{30}$. (Co to oznacza dla średniej długości odstępu między dwoma kolejnymi autobusami?) Asia przybyła na przystanek dokładnie w momencie odjazdu pierwszego autobusu (nie zdążyła na ten autobus). Basia natomiast dotarła na przystanek dokładnie 10 minut po Asi. Która z nich czekała dłużej na autobus? Przeprowadzić wielokrotnie symulację powyższego doświadczenia losowego. Porównać histogramy otrzymanych czasów oczekiwania na autobus Asi i Basi. Obliczyć średnie arytmetyczne otrzymanych czasów.