Kaczmarek Wiktoria, Aktywność 06

3. Wyznaczyć MNW estymator parametru rozkładu Poissona.

Prawdopodobieństwo rozkładu Poissona możemy zdefiniować:

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\dots$$

Rozkład Poissona jest rozkładem jedno parametrowym.

Zdefiniujmy funkcję wiarygodności:

$$L(k_1,k_2,\dots,k_n|\lambda)=\prod_{i=1}^{n}P(X=k_i)=\prod_{i=1}^{n}\frac{\lambda^{k_i}e^{-\lambda}}{k_i!}=\frac{\lambda\sum_{i=1}^{n}k_i}{\prod_{i=1}^{n}k_i!}e^{-n\lambda}$$

Gdzie λ ∈ $(0,\infty)$ jest argumentem funkcji L, $k_1,k_2,k_3,\dots,k_n$ są realizacjami zmiennych losowych odpowiednio $X_1,X_2,\dots,X_n$.

Estymator parametru jest maksimum tej funkcji po zmiennej λ, zatem gdy przyrównamy pierwszą pochodną funkcji ln(L) do zera , otrzymamy szukany estymator.

$$\ln(L)=\ln(\lambda)\sum_{i=1}^{n}k_i-n\lambda-\sum_{i=1}^{n}\ln k_1!$$

$$\frac{\partial(\ln(L))}{\partial\lambda}=\frac{\sum_{i=1}^{n}k_i}{\lambda}-n=0$$

$$n=\frac{\sum_{i=1}^{n}k_i}{\lambda}$$

$$\lambda=\frac{\sum_{i=1}^{n}k_i}{\mathrm{n}}$$

Następnie sprawdzamy drugą pochodną.

$$\frac{\partial^2(\ln(L))}{\partial\lambda^2}=-\frac{\sum_{i=1}^{n}k_i}{\lambda^2}$$

$$\forall\lambda \quad \frac{\partial^2(\ln(L))}{\partial\lambda^2}<0$$

Zatem możemy wywnioskować, że estymatorem największej wiarygodności parametru λ jest średnia arytmetyczna populacji, wyrażona wzorem:

$$\lambda=\frac{\sum_{i=1}^{n}k_i}{\mathrm{n}}.$$

5. Wygenerować 50 elementową próbę prostą z populacji, w której cecha X ma rozkład o gęstości

$$f(x) = \frac{x}{8}\mathbf{1}_{(0;\ 4)}(x)$$

a) Sporządzić histogram.
b) Wyznaczyć wartość oczekiwaną i wariancję oraz ich oceny na podstawie wygenerowanej próby.

a)

Na początku musimy skorzystać z Twierdzenia obrócenia dystrybuanty musimy obliczyć dystrybuantę dla podanej funkcji gęstości:

$$F_X(x) = \int_{\mathbb{R}} \frac{t}{8} 1_{(0;4)}(t)dt = \int_0^x \frac{t}{8} dt = \frac{x^2}{16}$$

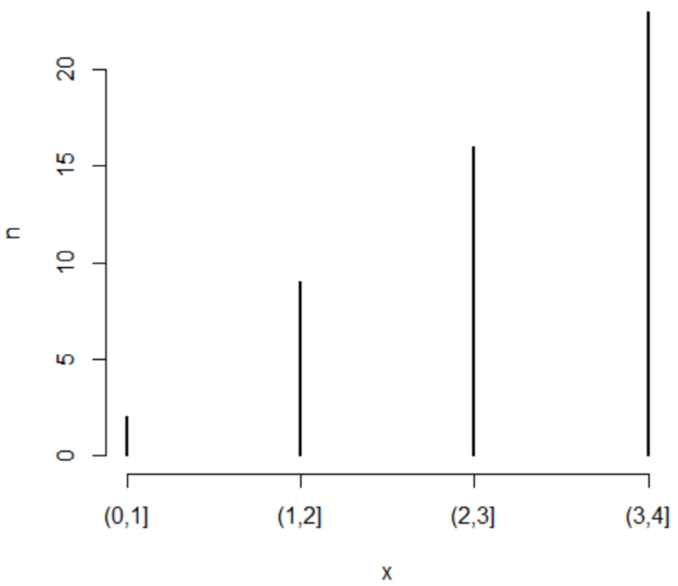
Zatem dystrybuantę możemy opisać jako:

$$F_X(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{x^2}{16}, & 0 < x < 4 \\ 1 & , x \geq 4 \end{cases}$$

Zatem dokonajmy odwrócenia dystrybuanty:

$$y = \frac{x^2}{16}$$

$$x = 4\sqrt{y}$$



Rys 5.1. Histogram dla wygenerowanej 50 elementowej próby.

b) Obliczamy wartość oczekiwaną i wariancję z podanej funkcji gęstości:

$$EX = \int_{\mathbb{R}} \frac{x^2}{8} 1_{(0;4)}(x)dx = \int_0^4 \frac{x^2}{8} dx = \frac{64}{24} \approx 2{,}67$$

$$EX^2 = \int_{\mathbb{R}} \frac{x^3}{8} 1_{(0;4)}(x)dx = \int_0^4 \frac{x^3}{8} dx = \frac{256}{32} = 8$$

$$D^2X = E(X^2) - EX^2 = 8 - \frac{4096}{576} \approx 0{,}78$$

Teraz obliczymy to samo z wygenerowanej próby.

Korzystamy z wbudowanych funkcji w R :

- Wartość oczekiwana – średnia z próby – mean;
- Wariancja – wariancja – var;

Wariancja definiujemy wzorem: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$.

```
> cat("mean ", m, " variance ", v, "\n")
mean  2.748954  variance  0.7579936
```

Możemy zauważyć, że parametry wygenerowanej próby wyszły podobne do obliczonych z parametrów. Różnią się nieznacznie, jednak próba o liczności 50 to mała próba, więc uzyskane wyniki wyszły naprawdę ładne.

6. Korzystając z dostępnego oprogramowania wygenerować 100 elementową próbę według rozkładu

a) bin(20; 0,8),

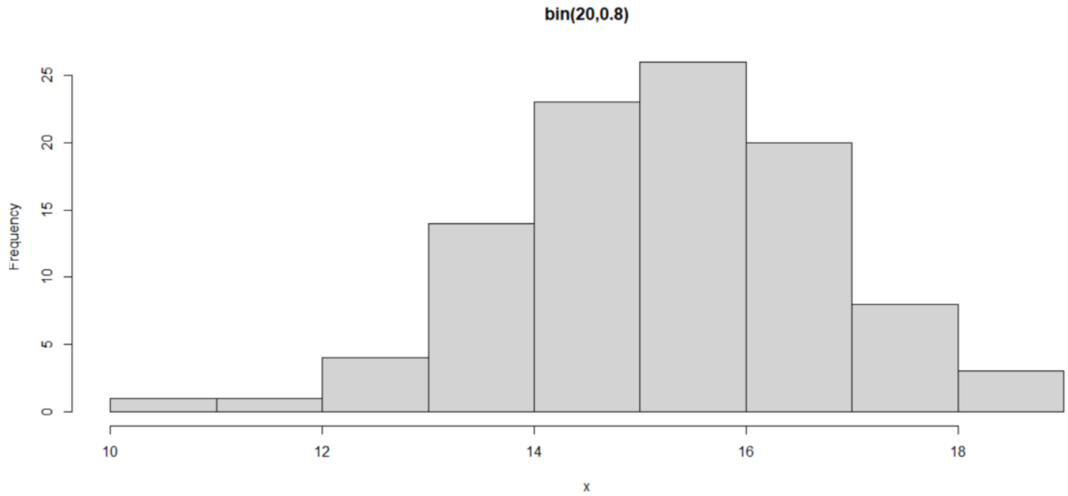
b) nbin(3; 0,1),

c) Poisson(5).

Sporządzić histogram i dokonać ocenę punktową parametrów.

a)

```
> x<-rbinom(100,20,0.8)
> hist(x,10,main="bin(20,0.8)")
```



Rys. 6.1. Wykres przedstawiający histogram dla wygenerowanej 100 elementowej próby według rozkładu bin(20;0.8).

$$EX = np = 20 \cdot 0.8 = 16$$

$$D^2X = np(1-p) = 20 \cdot 0.8 \cdot (1-0.8) = 20 \cdot 0.8 \cdot 0.2 = 3.2$$

$$DX = \sqrt{np(1-p)} = \sqrt{3.2} \approx 1{,}79$$

Znając już wartości teoretyczne obliczamy za pomocą wbudowanych funkcji w R średnia oraz wariancję

```
> mean(x)
[1] 15.72
> var(x)
[1] 2.405657
```

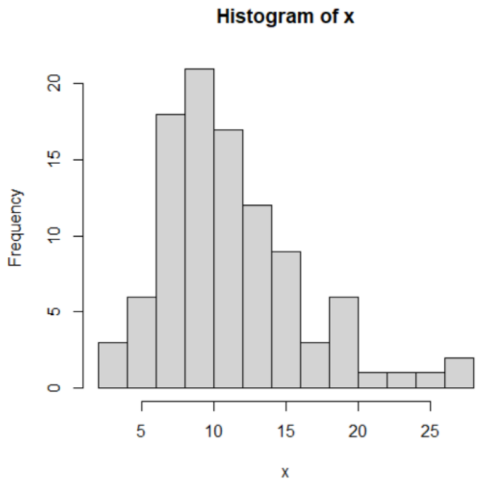
Możemy zauważyć, że wartość oczekiwana nieznacznie różni się od średniej, jednak wariancja znacząco odbiega, warto zauważyć, że 100 liczb to zbyt mała próba do globalnych porównań, jednak wystarczająca by dokonań jakiś analiz. Stąd widzimy, że histogram nie obiega znacząco od parametrów.

7. Wygenerować 100 elementową próbę według rozkładu logarytmiczno-normalnego z parametrami $\mu = 2{,}3$ i $\sigma = 0{,}5$.

a) Sporządzić histogram.

b) Dokonać estymacji parametrów, ocenić wartość oczekiwaną i wariancję oraz porównać te wartości z wartościami teoretycznymi.

a)



Rys. 7.1. Histogram do wygenerowanej próby.

```
> x<- rlnorm(100,2.3,0.5)
> hist(x,10)
```

b)Następnie obliczamy wartości teoretyczne z parametrów rozkładu

$$EX = e^{\mu} = e^{2.3} \approx 9{,}9741182$$

```
> exp(2.3)
[1] 9.974182
```

$$D^2X = \left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2} \approx 36.28$$

```
> (exp(0.5^2)-1)*exp(2*2.3+0.5^2)
[1] 36.28152
```

Teraz obliczymy to samo z wygenerowanej próby.

Korzystamy z wbudowanych funkcji w R :

- Wartość oczekiwana – średnia z próby – mean;
- Wariancja – wariancja – var;

```
> mean(x)
[1] 11.1802
> var(x)
[1] 24.19252
```

Porównując teoretyczne i obliczone parametry możemy zauważyć, że nie różnią się one znacząco. Wartość oczekiwana jest bardzo zbliżona, jednak w przypadku wariancji mamy większe rozbieżności ponieważ wygenerowana próba jest skupiona bliżej wartości oczekiwanej niż idealny rozkład z tymi parametrami.