

MWS Ćwiczenie 2

Estymacja punktowa i przedziałowa

Tomasz Korzeniowski, 265753

25 listopada 2017

1 Zadanie 1

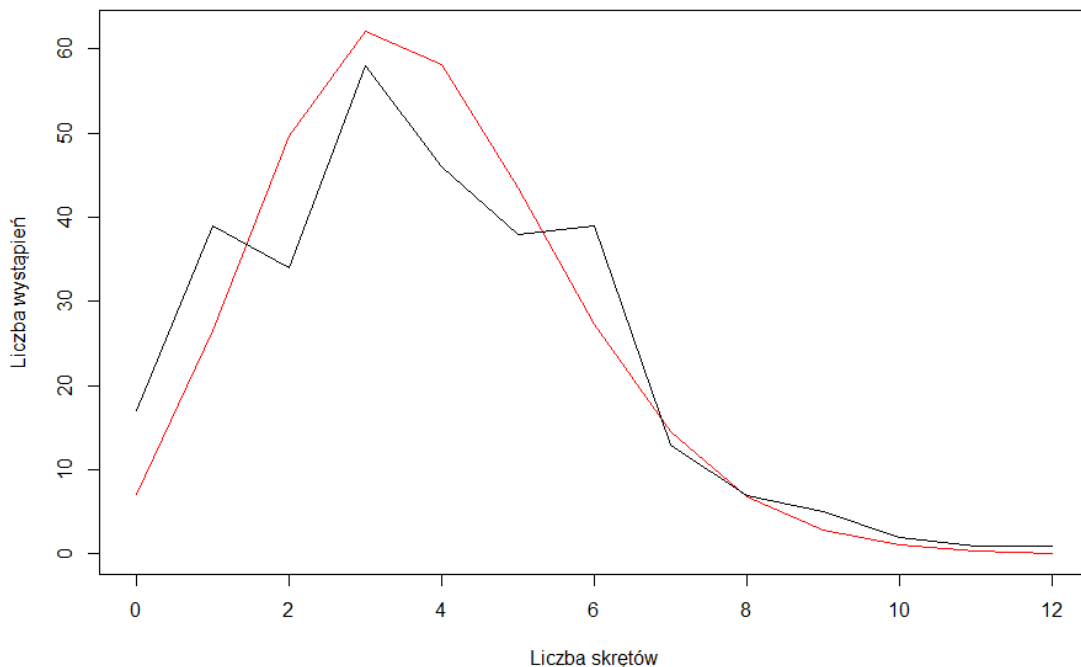
Rozkład Poissona wyznacza prawdopodobieństwo niezależnych, dyskretnych, zdarzeń następujących po sobie z daną częstotliwością w danym przedziale czasu. Jeśli oczekiwaną liczbą zdarzeń w tym przedziale jest λ , to prawdopodobieństwo, że wystąpi dokładnie k zdarzeń wyraża wzór

$$f_{\lambda}(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (1)$$

Estymatorem parametru λ rozkładu Poissona jest średnia z n -elementowej próby

$$\hat{\lambda} = \bar{X}_n \quad (2)$$

Do przedstawienia rzeczywistej liczby skrętów w prawo na wykresie, trzeba zliczyć jak często występuje określona liczba skrętów. Aby otrzymać spodziewaną liczbę skrętów w prawo należy wyznaczyć funkcję gęstości prawdopodobieństwa dla każdej liczby skrętów. Wynik porównania obu tych metod przedstawia wykres 1.



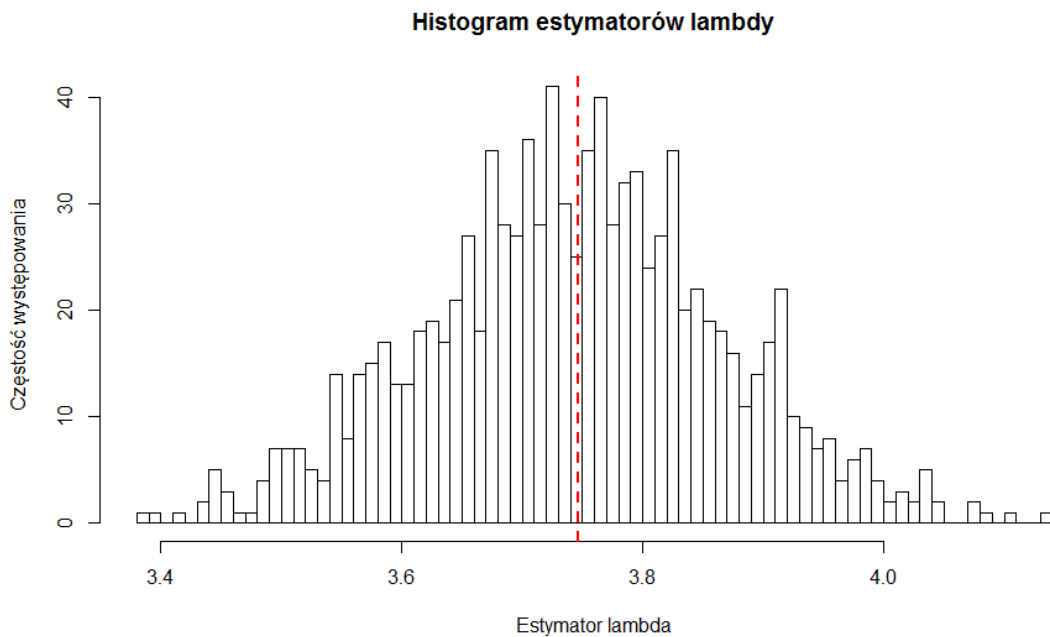
Wykres 1: Rzeczywista (czarna) i spodziewana (czerwona) liczba skrętów w prawo.

Do wyznaczenia liczby wystąpień danej liczby skrętów została napisana pętla zliczająca. Spodziewana liczba skrętów została wyznaczona z użyciem funkcji `dpois` pozwalającej na wyznaczenie gęstości rozkładu Poissona w zadanym przedziale i określonym parametrze λ .

Dla otrzymanych danych, estymowany parametr $\hat{\lambda} = 3.746667$. Na podstawie wykresu widać, że rozkład $Pois(\hat{\lambda})$ dobrze przybliża rzeczywiste liczby skrętów.

Bootstrap nieparametryczny stosujemy w sytuacji gdy nie znamy modelu parametrycznego. Polega on na wielokrotnym generowaniu prób losowych (losowanie ze zwracaniem) licznosci takiej jak oryginalna próba. Dla każdej z wygenerowanych prób wyznaczamy wartość badanej statystyki.

Do oszacowania odchylenia standardowego estymatora $\hat{\lambda}$ wykonano 1000 prób. Każda polegała na wylosowaniu ze zwracaniem n (licznosc próby) elementów z rzeczywistego zbioru skrętów w prawo. Dla każdej z prób wyznaczono estymator $\hat{\lambda} = m_1$. Wartość odchylenia standardowego estymatora $\hat{\lambda}$ to 0.1237953. Poszczególne wartości estymatorów przedstawia wykres 2.



Wykres 2: Wartości estymatorów lambdy otrzymane metodą bootstrapu nieparametrycznego.

Czerwoną przerywaną linią zaznaczono wartość $\hat{\lambda}$ wyznaczoną na podstawie dostępnych danych. Na histogramie widać, że wartości estymatorów otrzymanych metodą bootstrapu nieparametrycznego przypominają rozkład normalny. Jest to zgodne z oczekiwaniami (reguła trzech sigm), że najczęstszymi wartościami estymatora są wartości zbliżone do wartości estymatora wyznaczonego na podstawie dostępnej próby losowej. Można zatem stwierdzić, że metoda ta pozwala na dość dokładne oszacowanie $\hat{\lambda}$ (małe odchylenie standardowe).

2 Zadanie 2

Do wyznaczenia estymatorów parametrów α i β rozkładu Gamma

$$f_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad (3)$$

można wykorzystać metodę momentów. Momentem k -tego rzędu z próby X_1, \dots, X_n nazywamy

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k = \overline{(X^k)}_n \quad (4)$$

Korzystając z tej metody wyznaczamy estymatory parametrów α i β

$$\hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}, \quad \hat{\beta} = \frac{m_2 - m_1^2}{m_1} \quad (5)$$

Do wyznaczenia estymatorów metodą momentów wystarczy policzyć m_1 i m_2 korzystając z funkcji *mean* i powyższych wzorów.

Metoda największej wiarygodności polega na maksymalizacji logarytmicznej funkcji wiarygodności. Funkcja ta reprezentuje łączną funkcję gęstości prawdopodobieństwa. Na jej podstawie, dla rozkładu Gamma, estymatory parametrów α i β są następujące

$$\hat{\beta} = \frac{\bar{X}_n}{\hat{\alpha}}, \quad \overline{(\ln X)}_n - \ln \bar{X}_n = \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} - \ln \hat{\alpha} \quad (6)$$

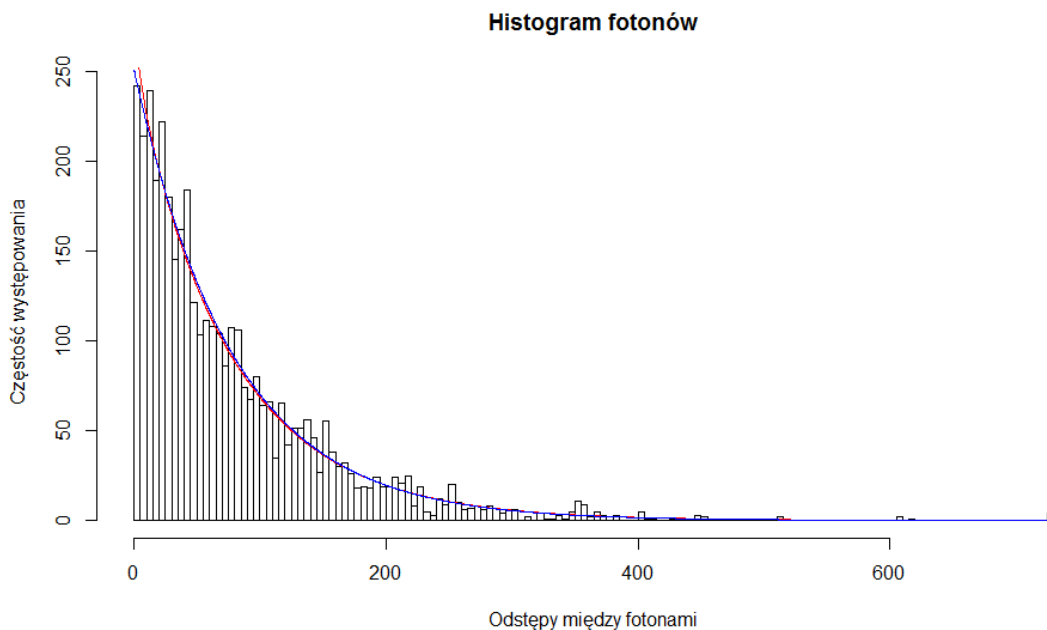
Wynika z tego, że estymatorów tych należy szukać numerycznie. Rozwiązanie równania można otrzymać korzystając z metody *uniroot*. Wyznaczane są wtedy pierwiastki zadanej funkcji. W naszym przypadku funkcją podawaną do metody będzie przekształcone równanie (6) tak by wyznaczyć $\ln \hat{\alpha}$. Drugim parametrem metody jest przedział w jakim szukamy pierwiastków. Przyjmijmy przedział $[0, 2]$.

Do wyznaczenia i narysowania gęstości rozkładu Gamma została wykorzystana funkcja *dgamma*, która przyjmuje przedział w jakim wyznaczyć rozkład oraz parametry *shape* i *scale* odpowiadające estymatorom (w ogólności dokładnym wartościom) $\hat{\alpha}$ i $\hat{\beta}$ szukanego rozkładu.

Otrzymane wartości estymatorów prezentuje tabela 1. Histogram odstępów między fotonami wraz z funkcjami rozkładu gęstości rozkładu Gamma o parametrach wyznaczonych dwiema metodami przedstawia wykres 3.

Wartość estymatora	$\hat{\alpha}$	$\hat{\beta}$
Metoda momentów	0.9608926	81.45088
Metoda największej wiarygodności	1.000356	78.23772

Tabela 1: Wartości estymatorów $\hat{\alpha}$ i $\hat{\beta}$.



Wykres 3: Histogram odstępów fotonów wraz z funkcjami gęstości dla parametrów estymowanych metodą momentów (czerwony) oraz metodą największego zaufania (niebieski).

Wartości estymatorów $\hat{\alpha}$ i $\hat{\beta}$ otrzymane obiema metodami są zbliżone do siebie. Potwierdzają to także obie funkcje gęstości prawdopodobieństwa, które niemal nakładają się na siebie. Na podstawie wykresu 3 widać także, że funkcje gęstości dla znalezionych estymatorów dość dokładnie odwzorowują histogram. Dowodzi to faktu, że jest to dobre przybliżenie szukanych parametrów rozkładu.

Bootstrap parametryczny różni się od nieparametrycznego znajomością modelu parametrycznego. Dla znanego modelu wielokrotnie generujemy próby losowe licznosci takiej jak oryginalna próba. Generujemy je na podstawie przyjętego rozkładu prawdopodobieństwa. Możemy do tego wykorzystać wyznaczone estymatory. Dla każdej z tak wygenerowanych prób wyznaczamy wartość badanej statystyki.

Przedziałem ufności na poziomie ufności γ jest przedział $[A(X_1, \dots, X_n), B(X_1, \dots, X_n)]$ dla którego zachodzi nierówność

$$\mathbb{P}(\theta \in [A(X_1, \dots, X_n), B(X_1, \dots, X_n)]) \geq \gamma \quad (7)$$

Istotą przedziału ufności jest wskazanie możliwych skrajnych wartości, dla których, z ufnością na poziomie γ , dokładna wartość parametru θ znajdzie się między nimi. Czasami, zamiast poziomu ufności γ podaje się współczynnik ufności $\alpha = 1 - \gamma$.

Obliczanie odchyłeń standardowych dla estymatorów $\hat{\alpha}$ i $\hat{\beta}$ będzie przebiegało następująco. W kolejnych iteracjach (1000) obliczamy wartości estymatorów tak jak opisano wcześniej (dla obu metod: MM i NW). Zamiast funkcji *dgamma* użyjemy funkcji *rgamma* o takich samych parametrach, lecz generującej losowe wartości z zadanego rozkładu zamiast gęstość rozkładu. Każdy z wyników zostanie zapisany na liście. Po zakończeniu iteracji, wyznaczmy odchylenia standardowe. Wyznaczanie przedziałów ufności polega na posortowaniu otrzymanych iteracyjnie wyników, a następnie odrzuceniu skrajnych $\alpha/2$ z nich. Największa i najmniejsza wartość listy będzie wyznaczała przedział ufności. Wyniki takiego działania prezentują tabele 2 i 3.

Odchylenie standardowe	$\hat{\alpha}$	$\hat{\beta}$
Metoda momentów	0.0307076	2.96457
Metoda największej wiarygodności	0.02043947	2.066343

Tabela 2: Odchylenie standardowe estymatorów $\hat{\alpha}$ i $\hat{\beta}$.

Przedział ufności	$\hat{\alpha}$	$\hat{\beta}$
Metoda momentów	[0.9034094, 1.0265143]	[75.65117, 87.18680]
Metoda największej wiarygodności	[0.9620835, 1.0459473]	[74.23167, 82.23115]

Tabela 3: Przedziały ufności dla $\hat{\alpha}$ i $\hat{\beta}$.

Można zaobserwować, że wyznaczone wcześniej wartości estymatorów $\hat{\alpha}$ i $\hat{\beta}$ znajdują się niemal pośrodku przedziałów ufności. Interpretacja jest analogiczna jak w przypadku wykresu 2.

Przedziały ufności metody największej wiarygodności są krótsze niż w przypadku metody momentów. Odchylenie standardowe również jest mniejsze dla metody największej wiarygodności. Wskazuje to na większą precyzję w modelowaniu analizowanych danych przez tę metodę.

3 Zadanie 3

Estymatorami rozkładu normalnego o nieznannej średniej i wariancji są podane wzorem

$$\hat{\mu} = \bar{X}_n, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (8)$$

Do wyznaczenia przedziałów ufności można wykorzystać wzory z prezentacji wykładowych

$$\mu \in \left[\hat{\mu} - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \hat{\mu} + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right] \quad (9)$$

$$\sigma^2 \in \left[\frac{(n-1)s^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)s^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right] \quad (10)$$

gdzie

$t_{n-1, 1-\frac{\alpha}{2}}$ – rozkład Studenta o $n-1$ stopniach swobody

$\chi_{n-1, 1-\frac{\alpha}{2}}$ – rozkład χ^2 o $n-1$ stopniach swobody

Dla każdego z poziomów ufności zostały wyznaczone przedziały ufności. Wyniki zawiera tabela 4.

Poziom ufności	μ	σ^2
90%	[-1.064291, -0.4556072]	[0.7104926, 1.657718]
95%	[-1.126038, -0.3938608]	[0.6626684, 1.822352]
99%	[-1.252499, -0.2673994]	[0.5810942, 2.210698]

Tabela 4: Przedziały ufności dla μ i σ^2

Do wyznaczenia przedziałów ufności dla średniej została wykorzystana funkcja *qt*, a dla wariancji funkcja *qchisq*. Funkcje te pozwalają na wyznaczanie odpowiednio rozkładów Studenta i χ^2 z zadanymi parametrami prawdopodobieństwa (ufności) i stopni swobody.

Estymatorami wyznaczonymi na podstawie dostępnych danych są $\hat{\mu} = -0.7599493$ i $\hat{\sigma}^2 = 1.015394$. Tak jak poprzednio, wartości te leżą blisko środków przedziałów ufności. Ponadto widać, że im większy poziom ufności tym dłuższy przedział ufności.