6 Przedziały ufności

6.1 Przykład

Przykład. Badano czas oczekiwania na tramwaj, który kursuje w jednakowych odstępach czasu. Plik czas_oczek_tramwaj.RData zawiera dane dotyczące czasu oczekiwania na tramwaj (wyrażonego w minutach) 100 osób wybranych losowo. Zmienna X to czas oczekiwania na tramwaj. Jest to zmienna ilościowa ciągła.

- model: rozkład jednostajny
- $\mathcal{P} = \{U(0,b) : b \in (0,\infty)\}$
- $\Theta = (0,\infty)$ oraz $\theta = b$

Niech $\mathbf{X} = (X_1,\ldots,X_n)^\top$ będzie próbą prostą z populacji o rozkładzie jednostajnym $U(0,b)$. $100(1-\alpha)\%$ przedziałem ufności dla parametru b jest przedział losowy postaci:

$$\left(\frac{\max\{X_1,\ldots,X_n\}}{\sqrt[n]{1-\frac{\alpha}{2}}}, \frac{\max\{X_1,\ldots,X_n\}}{\sqrt[n]{\frac{\alpha}{2}}}\right).$$

```
load(url("http://ls.home.amu.edu.pl/data_sets/czas_oczek_tramwaj.RData"))
# estymator b
(b_est <- max(czas_oczek_tramwaj))
```

```
## [1] 13.92
```

```
# przedział ufności dla b
b_conf_int <- function(x, conf_level = 0.95) {
  alpha <- 1 - conf_level
  n <- length(x)
  l <- max(x) / (1 - alpha / 2)^(1 / n)
  u <- max(x) / (alpha / 2)^(1 / n)
  return(c(l, u))
}
b_conf_int(czas_oczek_tramwaj)
```

```
## [1] 13.92352 14.44308
```

6.2 Zadania

Zadanie 1. Przebadano 200 losowo wybranych 5-sekundowych okresów pracy centrali telefonicznej. Rejestrowano liczbę zgłoszeń. Wyniki są zawarte w pliku Centrala.RData. Wykorzystując przyjęty wcześniej model statystyczny dla tych danych, wyznacz (trzema metodami) przedział ufności dla parametru rozkładu teoretycznego.

```
##      LCL      UCL
## 1.561968 1.932765
```

```
##      LCL      UCL
## 1.561968 1.932765
```

```
##      LCL      UCL
## 1.557187 1.922813
```

Zadanie 2. Zmienna w pliku awarie.txt opisuje wyniki 50 pomiarów czasu bezawaryjnej pracy danego urządzenia (w godzinach). Wykorzystując przyjęty na wykładzie model statystyczny dla tych danych wyznacz granice przedziału ufności dla wartości oczekiwanej i wariancji rozkładu teoretycznego.

```
##       UCL       LCL
##  850.0693 1483.8742
```

```
##       UCL       LCL
##  722617.9 2201882.5
```

Zadanie 3. Niech $\mathbf{X} = (X_1, \ldots, X_n)^\top$ będzie próbą prostą z populacji o rozkładzie Rayleigha $R(\lambda)$, $\lambda > 0$.

1. Napisz funkcję `median_cint()` , która implementuje następujący przybliżony przedział ufności dla mediany $\sqrt{\lambda \ln 2}$ tego rozkładu:

$$\left(\sqrt{\ln(2)\frac{1}{n}\sum_{i=1}^{n} X_i^2\left(1-\frac{z(1-\alpha/2)}{\sqrt{n}}\right)}, \sqrt{\ln(2)\frac{1}{n}\sum_{i=1}^{n} X_i^2\left(1+\frac{z(1-\alpha/2)}{\sqrt{n}}\right)}\right),$$

   gdzie $z(\beta)$ oznacza kwantyl rzędu β z rozkładu normalnego $N(0,1)$. Funkcja ta powinna mieć dwa argumenty: `x` - wektor zawierający dane, `conf_level` - poziom ufności. Funkcja zwraca obiekt typu `list` klasy `confint` o następujących elementach: `title` - nazwa estymowanej funkcji parametrycznej, `est` - wartość ENW funkcji parametrycznej, `l` - lewy kraniec przedziału ufności, `r` - prawy kraniec przedziału ufności, `conf_level` - poziom ufności.

2. Następujące dane to pomiary średniej szybkości wiatru w odstępach 15 minutowych odnotowane wokół nowo powstającej elektrowni wiatrowej:

0.9	6.2	2.1	4.1	7.3
1.0	4.6	6.4	3.8	5.0
2.7	9.2	5.9	7.4	3.0
4.9	8.2	5.0	1.2	10.1
12.2	2.8	5.9	8.2	0.5

   Teoretyczny rozkład średniej szybkości wiatru to rozkład Rayleigha $R(\lambda)$, $\lambda > 0$. Używając funkcji `median_cint()` , oblicz wartość ENW i krańce 95% przedziału ufności dla mediany średniej szybkości wiatru. **Wskazówka:** Przed wywołaniem funkcji `median_cint()` , najpierw załaduj następujące funkcje przeciążone `print()` i `summary()` :

```r
print.confint <- function(x) {
  cat(x$conf_level * 100, "percent confidence interval:", "\n")
  cat(x$l, " ", x$r, "\n")
}

summary.confint <- function(x) {
  cat("\n", "Confidence interval of", x$title, "\n", "\n")
  cat(x$conf_level * 100, "percent confidence interval:", "\n")
  cat(x$l, " ", x$r, "\n")
  cat("sample estimate", "\n")
  cat(x$est, "\n")
}
```

```
## 95 percent confidence interval:
## 3.863593   5.845955
```

```
##
##  Confidence interval of mediana
##
## 95 percent confidence interval:
## 3.863593   5.845955
## sample estimate
## 4.954924
```

Zadanie 4. Dla danego wektora obserwacji i poziomu ufności napisz funkcję określającą granice przedziału ufności na poziomie ufności $1-\alpha$, $\alpha \in (0,1)$ dla wartości oczekiwanej w rozkładzie normalnym. Domyślny poziom ufności powinien wynosić $0{,}95$. Następnie przeprowadź symulacje (z liczbą powtórzeń `nr = 1000`) sprawdzając prawdopodobieństwo pokrycia tego przedziału ufności (tj. prawdopodobieństwo, że ten przedział ufności zawiera wartość oczekiwaną) dla rozkładów $N(1,3)$, $\chi^2(3)$ i $Ex(3)$ osobno. Rozważ liczby obserwacji $n = 10, 50, 100$. Zinterpretuj wyniki. **Wskazówka:** Symulacja powinna przebiegać według następujących kroków:

1. Przyjmij poziom istotności, n, `nr`, rozkład generowanych danych oraz `temp = 0`.
2. Wygeneruj n obserwacji z zadanego rozkładu.
3. Wyznacz granice przedziału ufności dla danych wygenerowanych w kroku 2.

4. Jeśli teoretyczna wartość oczekiwana należy do przedziału otrzymanego w kroku 3, zwiększ `temp` o jeden.
5. Powtórz kroki 2-4 `nr` razy.
6. Wyznacz `temp / nr`.

```
## n = 10
```

```
## [1] 0.959
```

```
## [1] 0.901
```

```
## [1] 0.899
```

```
## n = 50
```

```
## [1] 0.941
```

```
## [1] 0.944
```

```
## [1] 0.941
```

```
## n = 100
```

```
## [1] 0.946
```

```
## [1] 0.942
```

```
## [1] 0.946
```