Решения на задачи от Домашно 1 по ВИС практикум

Камен Димитров Младенов, FN

16.04.2024

Задача 1. Генерирайте стойности $x_1, x_2, ..., x_n$ на случайна величина $X = W_1 + W_2$, където $W_i \sim \text{Exp}(1)$ и $W_1 \perp\!\!\!\perp W_2$. Проверете нулевата хипотезата, че генерираните данни са от нормално разпределение с помощта на теста на Шапиро–Уилк (Shapiro–Wilk). Повторете 10000 пъти за $n = 10, 20, 30, 40, 50, 60$. Колко често заключението на теста е вярно? Представете чрез подходяща графика честотата на вярно заключение в зависимост от n.

Решение

Listing 1: Сорс-код

```
ex1 <- function(ns = 1:6 * 10) {
  check_null_hyp <- function(n, iters = 10000) {
    accepts <- sum(sapply(1:iters, function(i) {
      W1 <- rexp(n, 1)
      W2 <- rexp(n, 1)
      X <- W1 + W2
      return (shapiro.test(X)$p.value > 0.05)
    }))
    return (accepts / iters)
  }

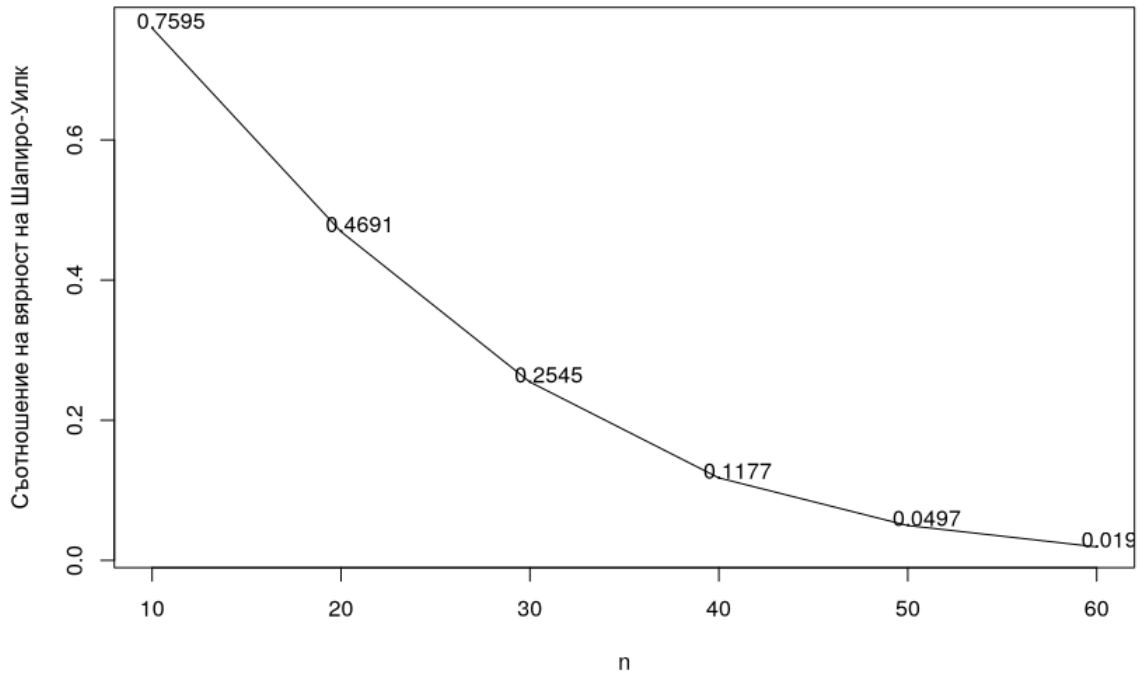
  ys <- sapply(ns, check_null_hyp)

  plot(ns, ys, type="l", xlab="n", ylab="Съотношение на вярност на Шапиро-Уилк")
  text(ns + 1, ys + 0.01, ys)
}

ex1()
```

Заключението е вярно през средно $\sim 27.8\%$ от времето.

Графика на честотата на вярно заключение в зависимост от n:



Задача 2. В кутия има 2 зелени, 3 бели и 5 жълти топки. Симулирайте n тегления на топка с връщане. Използвайте генерираните данни, за да проверите хипотезата, че вероятността да изтеглим зелена топка е $1/5$. Повторете 10000 пъти за $n = 50, 100, 200, 500$. Колко често заключението на теста е вярно? Представете чрез подходяща графика честотата на вярно заключение в зависимост от n.

Решение

Listing 2: Copc-код

```
ex2 <- function(greens = 2, whites = 3, yellows = 5, ns = c(50, 100, 200, 500)) {
  balls <- append(append(rep("green", greens), rep("white", whites)), rep("yellow", yellows))

  check_hyp <- function(n, iters=10000) {
    accepts <- sum(sapply(1:iters, function(i) {
      sampled_greens <- sum(sapply(1:n, function(s) {
        return (sample(balls, 1, replace=TRUE) == "green")
      }))

      return (prop.test(sampled_greens, n, 1/5)$p.value > 0.05)
    }))

    return (accepts / iters)
  }

  ys <- sapply(ns, check_hyp)
  plot(ns, ys, type="l", xlab="n", ylab="Съотношение на вярност на заключението")
  text(ns, ys, ys)
}

ex2()
```

Заключението е вярно през средно $\sim 96.6\%$ от времето.

Графика на честотата на вярно заключение в зависимост от n:

