Homework 2

20307100013 蒋翌坤

Solution for Problem 1

Since n exceeds 35, we can use the approximation as follows to test the hypothesis.

$$\lim_{n\to\infty} P\left(D_n \le d/\sqrt{n}\right) = L(d), where\ L(d) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2d^2}$$

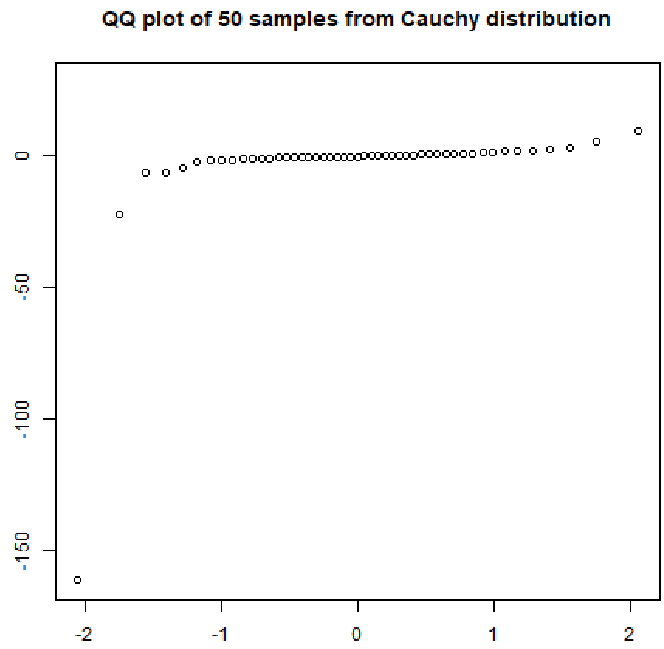
From tabulation, critical value $D_{n,\alpha} = \frac{d_\alpha}{\sqrt{n}} = \frac{1.356}{\sqrt{100}} = 0.136 > D_n = 0.04$

Therefore, we cannot reject the null hypothesis $H_0: X_1, \dots, X_n \sim F_X$ at level $\alpha = 0.05$.

Since D_n is distribution free and independent of F_X, we choose $F_X \sim U(0,1)$ in our simulation for simplicity, but it is also generalizable. To justify our answer, we use 100,000 simulations. The corresponding R code is in the appendix.

The result is as follows. In 100,000 simulations, there are 1,809 times that $D_n \le 0.04$, p-value $= \frac{100{,}000-1{,}809}{100{,}000} > \alpha$. Therefore, we cannot reject the null hypothesis $H_0: X_1, \dots, X_n \sim F_X$ at level $\alpha = 0.05$, which is the same result as the above answer.

Solution for Problem 2



The above figure is the QQ plot of 50 random numbers generated from Cauchy distribution. We can see from the figure that even though the graph resembles a straight line in the middle, points at two tails deviate and are not of a straight line. Therefore, we can say that the data does not fit a normal distribution. The corresponding R code to plot the figure is in the appendix.

Solution for Problem 3

We can use asymptotic null distribution to test whether the sequence is random.

let A be type 1 and B be type 2, $\lambda = \frac{n_1}{n} = \frac{8}{11}$

$$Z = \frac{R - 2n\lambda(1-\lambda)}{2\sqrt{n}\lambda(1-\lambda)} = \frac{5 - 2 \times 11 \times \frac{8}{11} \times \left(1 - \frac{8}{11}\right)}{2 \times \sqrt{11} \times \frac{8}{11} \times \left(1 - \frac{8}{11}\right)} \approx 0.4837$$

Assume a two tail test, $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.64 > Z$

Therefore, we cannot reject the null hypothesis of randomness. The sequence is **significantly random** at level $\alpha = 0.1$.

To justify our answer, we use 100,000 simulations. The test statistic is number of runs. The corresponding R code is in the appendix.

The result is as follows. In 100,000 simulations, there are 20,618 times that $R = 5$, p-value $= \frac{20{,}618}{100{,}000} > \alpha$. Therefore, we cannot reject the null hypothesis of randomness. The sequence is **significantly random** at level $\alpha = 0.1$, which is the same result as the above answer.

Appendix

R code for Solution for Problem 1

```
dn <- 0.04
sim_num <- 100000
a <- 0.05
n <- 100
k <- 0
i <- 0

while (i < sim_num) {
    set.seed(i)
    s <- runif(n, min = 0, max = 1)
    dn_i <- max(abs(sort(s) - seq(from = 1 / n, to = 1, by = 1 / n)))
    # cat(dn_i)
    if (dn_i < dn) {
        k <- k + 1
    }
    i <- i + 1
}
print(k)
```

R code for Solution for Problem 2

```
set.seed(2)
d <- rcauchy(50)
```

```
di <- sort(d)
p <- seq(from = 0.02, to = 1, by = 0.02)

png("h2-p2.png")
plot(qnorm(p), di, main = "QQ plot of 50 samples from Cauchy
distribution", xlab = "", ylab = "")
dev.off()
```

R code for Solution for Problem 3

```
get_r <- function(a) {
    r <- 1
    for (i in seq(length(a) - 1)) {
        if (a[i] != a[i + 1]) {
            r <- r + 1
        }
    }
    return(r)
}

sim_num <- 100000
n <- 11
r <- 5
i <- 0
k <- 0
while (i < sim_num) {
    set.seed(i)
    sim_seq <- sapply(runif(n, min = 0, max = 1), FUN = round)
    if (get_r(sim_seq) == r) {
        k <- k + 1
    }
    i <- i + 1
}
print(k)
```