Math 185 Homework 2

Lucien Chen

2024-04-16

Homework 2

Question 1

```
K <- c(5, 10, 20, 50, 100)
N <- c(2*K, 4*K, 5*K, 6*K, 8*K, 10*K)
p_mat = matrix(NA, nrow=length(K), ncol = length(N))
for (i in 1:length(K)) {
  for (j in 1:length(N)) {
    p_vals <- vector("numeric", 10000)
    for (k in 1:10000) {
      x <- sample(1:K[i], N[j], replace=TRUE)
      p <- chisq.test(x)
      p_vals[k] = p$p.value
    }
    p_mat[i, j] = mean(p_vals)
  }
}

p_mat
```

```
##              [,1]         [,2]         [,3]          [,4]          [,5]         [,6]         [,7]
## [1,] 7.132909e-01 8.193877e-01 9.178778e-01  9.884406e-01  9.994353e-01 8.182975e-01 9.156045e-01
## [2,] 2.090978e-01 1.393895e-01 6.498901e-02  9.368914e-03  4.204755e-04 1.379837e-01 6.620526e-02
## [3,] 2.670153e-02 2.592565e-03 1.465764e-05  1.611793e-12  3.215024e-25 2.608057e-03 5.001045e-05
## [4,] 8.645515e-04 3.156375e-07 6.153041e-14  5.065328e-53 1.965112e-138 7.422136e-07 1.644948e-16
## [5,] 7.523212e-05 7.547717e-14 3.487322e-41 3.679125e-161  0.000000e+00 2.479866e-14 3.980097e-48
##               [,8]          [,9]        [,10]        [,11]        [,12]         [,13]
## [1,]  9.782066e-01  9.994171e-01 9.999981e-01 8.548454e-01 9.415913e-01  9.880784e-01
## [2,]  1.754662e-02  4.662787e-04 1.139318e-06 1.135087e-01 4.668022e-02  9.149942e-03
## [3,]  3.902426e-10  4.442326e-26 1.034699e-62 8.489970e-04 2.580470e-06  3.554028e-12
## [4,]  1.094721e-43 3.262129e-149 0.000000e+00 1.123257e-09 6.036202e-26  1.558379e-58
## [5,] 3.150547e-109  0.000000e+00 0.000000e+00 3.022407e-21 1.839202e-55 5.122968e-158
##              [,14]        [,15]        [,16]        [,17]         [,18]         [,19]
## [1,]  9.998650e-01 9.999999e-01 8.785279e-01 9.581259e-01  9.938227e-01  9.999697e-01
## [2,]  1.207151e-04 2.546447e-08 9.218299e-02 3.364564e-02  5.246345e-03  2.419586e-05
## [3,]  4.335491e-35 9.589245e-75 2.972368e-04 1.682815e-07  3.354251e-14  1.112989e-41
## [4,] 1.801415e-195 0.000000e+00 6.599431e-13 2.438613e-31  3.198227e-76 7.557564e-232
## [5,]  0.000000e+00 0.000000e+00 1.647918e-29 1.379604e-86 3.033870e-192  0.000000e+00
##              [,20]        [,21]         [,22]         [,23]        [,24]         [,25]
## [1,]  1.000000e+00 9.159819e-01  9.779045e-01  9.980891e-01 9.999979e-01  1.000000e+00
## [2,]  6.268166e-09 6.469496e-02  1.742911e-02  1.481223e-03 2.363220e-06  7.130019e-12
## [3,] 1.189891e-101 5.708639e-05  2.847900e-10  5.386829e-19 3.234349e-59 2.300725e-132
## [4,]  0.000000e+00 1.538991e-18  4.695876e-42 1.895385e-104 0.000000e+00  0.000000e+00
## [5,]  0.000000e+00 2.947465e-34 1.309242e-125 2.290730e-281 0.000000e+00  0.000000e+00
##             [,26]         [,27]         [,28]        [,29]         [,30]
## [1,] 9.410604e-01  9.882351e-01  9.994060e-01 9.999999e-01  1.000000e+00
## [2,] 4.630789e-02  8.701338e-03  4.986907e-04 3.882991e-08  5.779022e-15
## [3,] 6.702134e-06  7.417604e-12  5.424228e-28 3.540246e-81 2.002435e-173
## [4,] 9.590439e-28  1.190004e-60 2.115116e-135 0.000000e+00  0.000000e+00
## [5,] 3.117667e-47 5.988393e-158  0.000000e+00 0.000000e+00  0.000000e+00
```

As we can see from the probability matrix, where each entry $A_{i,j}$ corresponds to the p-value returned from the test with the i-th entry of K and j-th entry of N. As both k and n increase, the probability for the Pearson test approaches 0.

Question 2

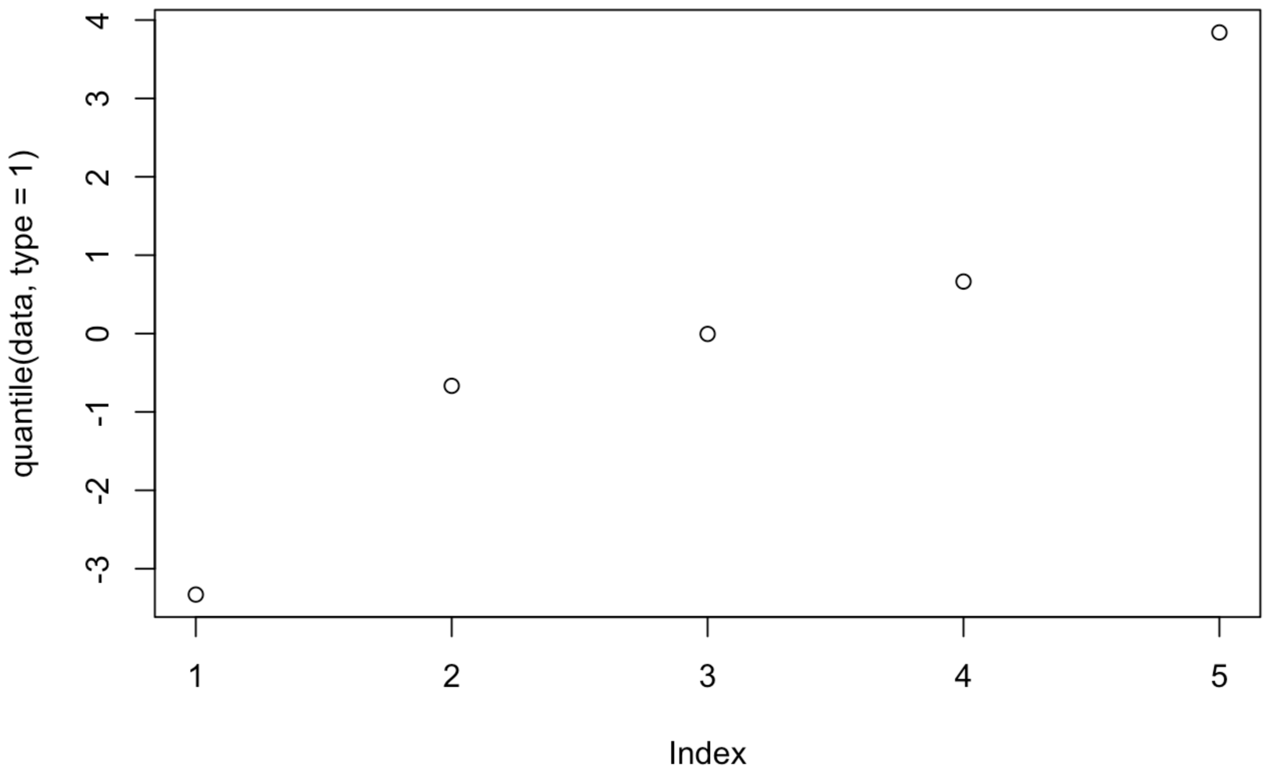
```
data <- rnorm(10000, 0, 1)
```

Type 1 computes the quantiles using the inverse of the empirical distribution function

```
quantile(data, type=1)
```

```
##           0%          25%          50%          75%         100% 
## -3.330016792 -0.667002404 -0.005033247  0.663587395  3.841937640
```

```
plot(quantile(data, type=1))
```

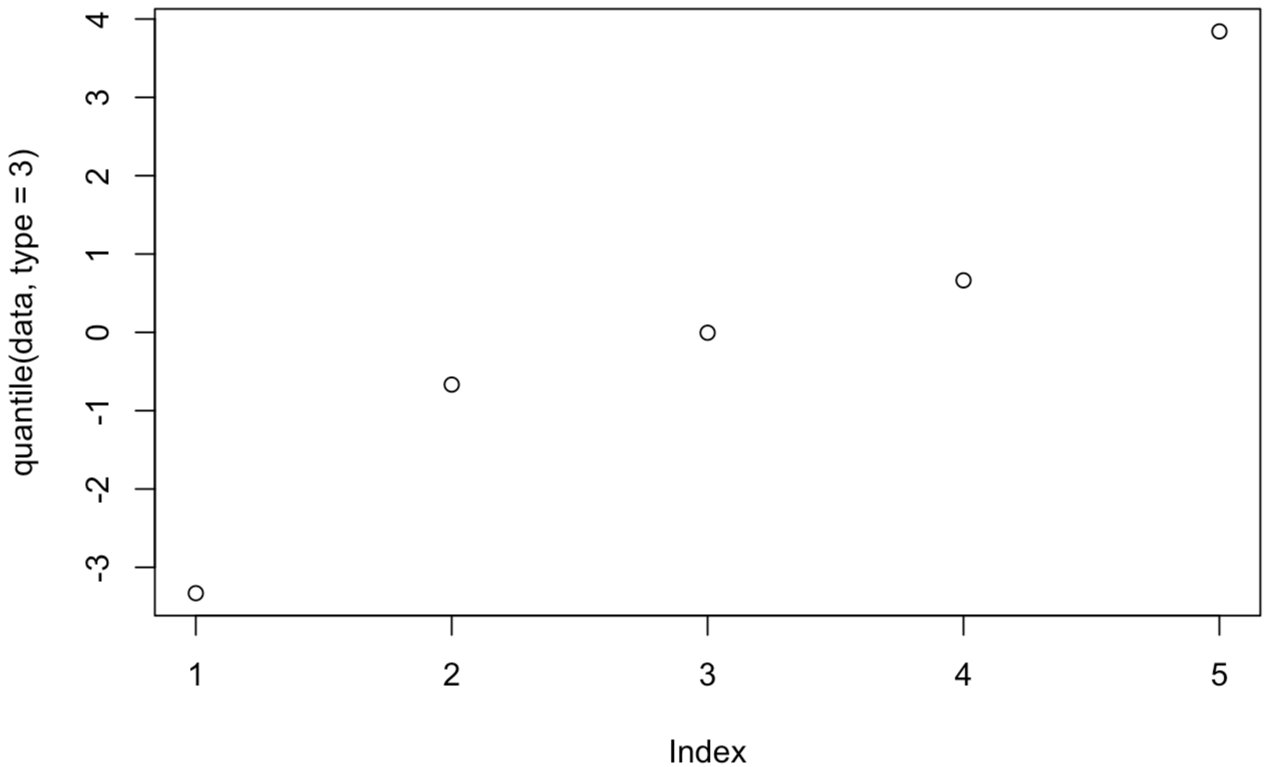


Type 3 computes the quantiles using the nearest even order statistic.

```
quantile(data, type=3)
```

```
##           0%          25%          50%          75%         100% 
## -3.330016792 -0.667002404 -0.005033247  0.663587395  3.841937640
```

```
plot(quantile(data, type=3))
```



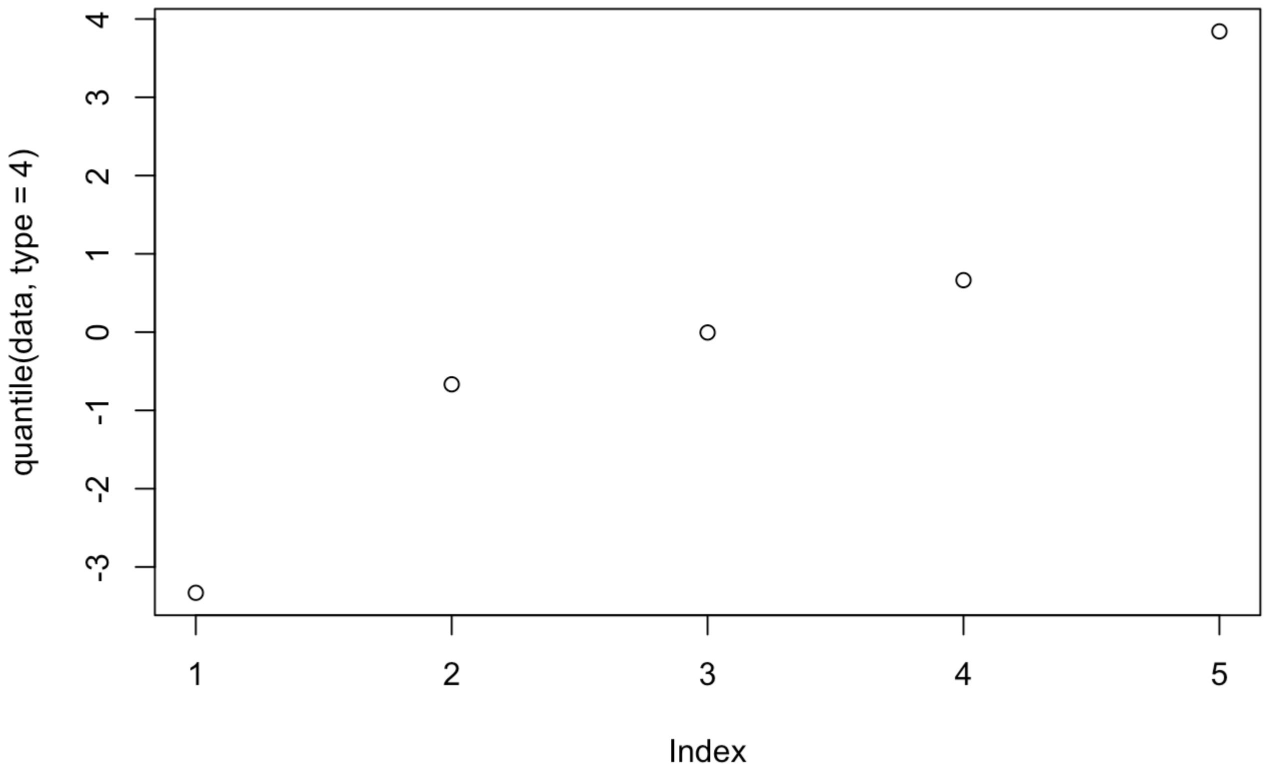
In this case, the results are identical to that of type 1.

Type 4 calculates the quantiles via a linear interpolation of the empirical cdf.

```
quantile(data, type=4)
```

```
##           0%          25%          50%          75%         100%
## -3.330016792 -0.667002404 -0.005033247  0.663587395  3.841937640
```

```
plot(quantile(data, type=4))
```

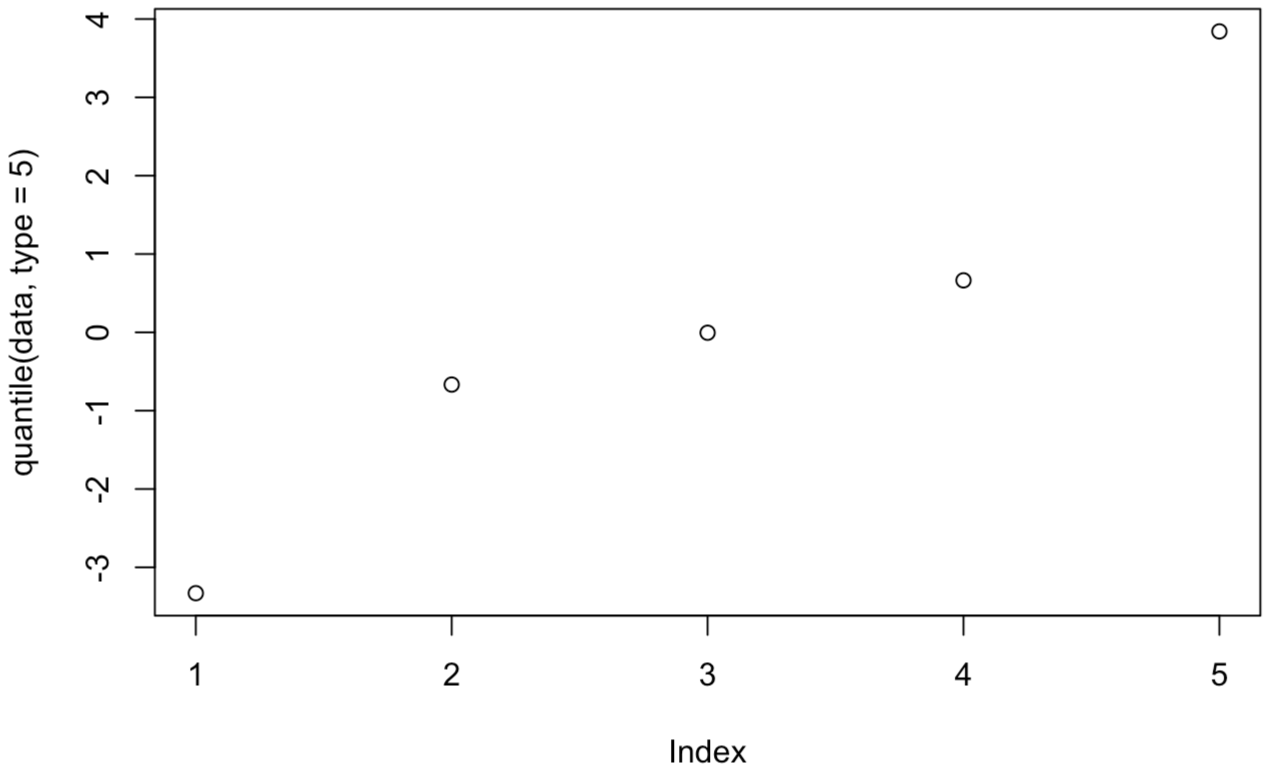


Type 5 calculates the quantiles using a piecewise linear function where the knots are values midway through steps of the empirical cdf.

```
quantile(data, type=5)
```

```
##          0%         25%         50%         75%        100% 
## -3.33001679 -0.66679789 -0.00501535  0.66361528  3.84193764
```

```
plot(quantile(data, type=5))
```



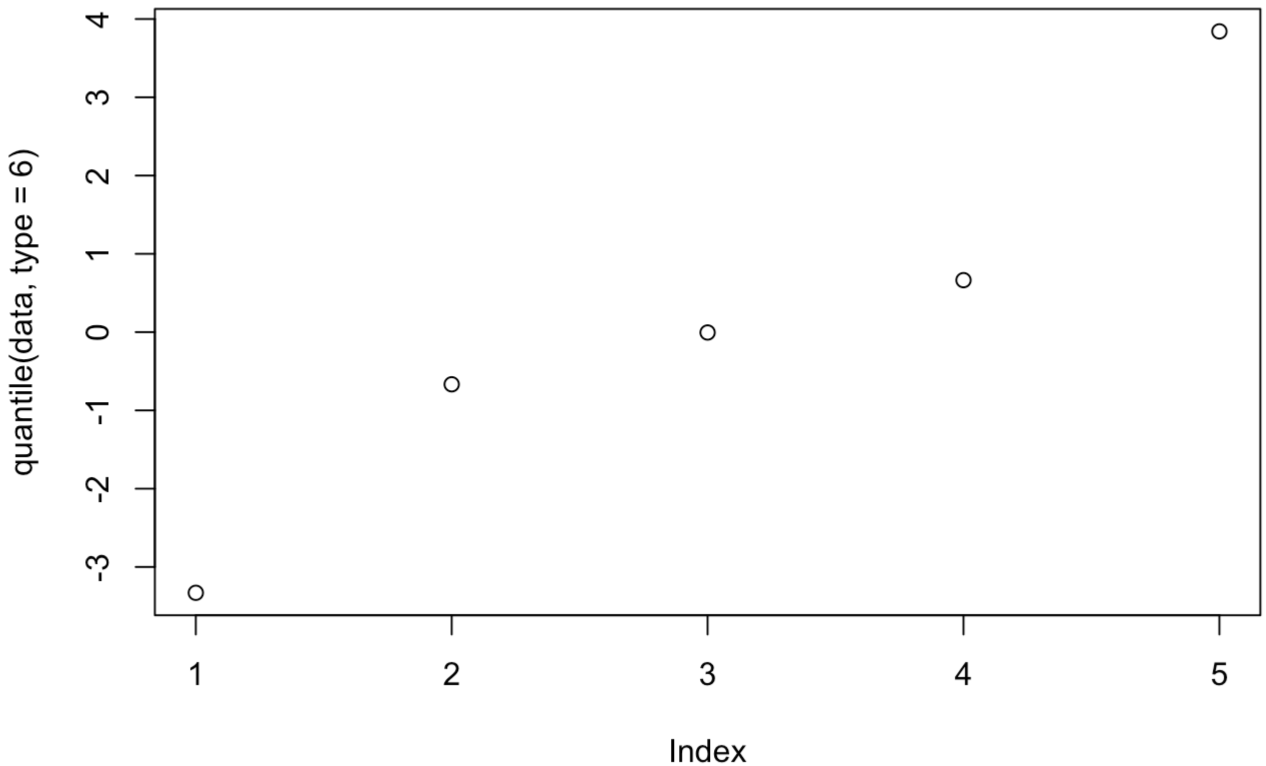
Here we start to see some differences in the calculations of some of the quantiles. Note, however, that these differences are quite small (median is .0058 vs .0057 in other cases).

Type 6 calculates quantiles using the expectation of the cdf.

```
quantile(data, type=6)
```

```
##          0%         25%         50%         75%        100% 
## -3.33001679 -0.66690015 -0.00501535  0.66362922  3.84193764
```

```
plot(quantile(data, type=6))
```



In this case, we see that the results are the same as those from using the type 5 method for calculating quantiles. In that regard, we can see that this case is also different from the first three methods we went over.

Question 3

```
uniform.test <- function(x, B=10000) {
  a = min(x)
  b = max(x)
  D_0 = b - a
  tests = vector("numeric", B)
  for (b in 1:B) {
    sample_dist = sample(x, replace=TRUE)
    a_b = min(sample_dist)
    b_b = max(sample_dist)
    D_b = b_b - a_b
    tests[b] = D_b
  }
  p_value = (as.integer(tests >= D_0) + 1)/(B + 1)
  return(p_value)
}
```

In my function uniform.test, I define the test stat to be the difference of max and min of our original data. Then, I conduct a bootstrap where I repeatedly sample the data and create a boostrap statistic, the difference between the max and min of the sample data. Then I compute a p-value and return it.