Creating invariant test for exponential distribution

Kacper Kowalczyk

2023-02-14

Testing problem

My goal is to develop a way of testing if set sample is from exponential distribution. Let $\mathcal{X} = (X_1, \dots, X_n)$ be iid sample.

$$H_0 : \exists \lambda : X \sim EXP(\lambda) \qquad\qquad H_1 : \quad \sim H_0$$

Construcion of the test

Hypothesis H_0 is invariant under the group

$$G = \{g : \quad g(\mathcal{X}) = (aX_1, \dots, aX_n), \quad a > 0\}$$

becouse if $X \sim Exp(\lambda)$ then $aX \sim Exp(\lambda/a)$.

Maximal invariant of this testing problem is

$$s(\mathcal{X}) = \left(\frac{X_1}{\bar{X}}, \dots, \frac{X_n}{\bar{X}}\right)$$

becouse if we assume $s(\mathcal{X}) = s(\mathcal{Y})$ then

$$\mathcal{Y} = (Y_1, \dots, Y_n) = \left(\frac{\bar{Y}}{\bar{X}}X_1, \dots, \frac{\bar{Y}}{\bar{X}}X_n\right) = g(\mathcal{X})$$

Test is invariant if our test statistic is a function of maximal invariant.

It is easy to see that, under the null, $Z = s(\mathcal{X})$ has approximately exponential distribution with parameter $\lambda = 1$. Thats, becouse maximum likelihood estimator of λ is $\hat{\lambda} = \bar{X}^{-1}$.

$$Z = \mathcal{X}/\bar{X} = \hat{\lambda}\mathcal{X} \sim Exp(\lambda/\hat{\lambda}) \approx Exp(1)$$

Furthermore $F(Z) \sim U(0, 1)$. Then our we will define our test statistic as follows:

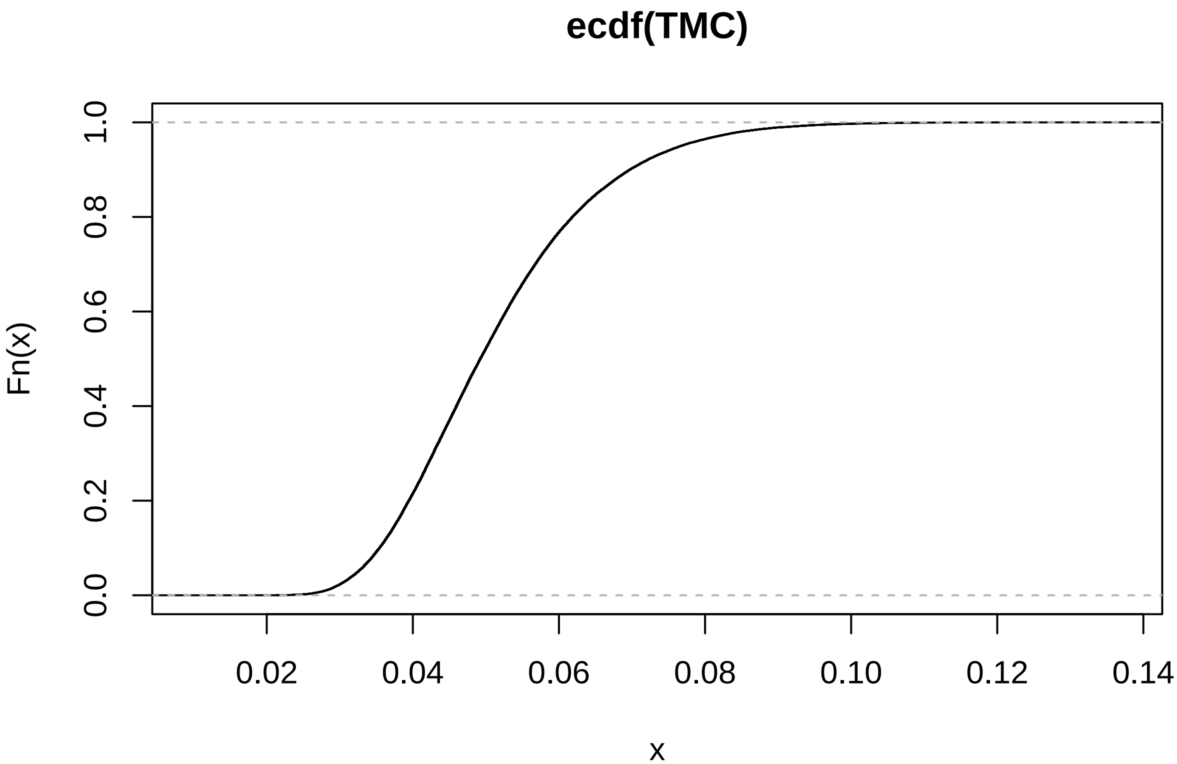
$$Q(\mathcal{X}) = \sup_{t \in \mathbb{R}} |\hat{F}_Z(t) - F_{Exp(1)}(t)| = \sup_{t \in (0,1)} |\hat{F}_{F_{Exp(1)}(Z)} - t|$$

To find supremum it is sufficient to check only $2 * n$ points.

$$Q(\mathcal{X}) = \max_{i=1,\dots,n} \max \left\{ \left| \frac{i-1}{n} - F_{Exp(1)}(Z_{(i)}) \right|, \quad \left| \frac{i}{n} - F_{Exp(1)}(Z_{(i)}) \right| \right\}$$

As we can see this statistic is a function of our maximal invariant, therefore this test is invariant. To calculate p-value we will use Monte Carlo method.

```
MC <- 100000 # amount of Monte Carlo iterations needed to calculate
# empirical distribution of test statistics that is used to calculate p-values

stat <- function(X){   #function that calculates test statistics

  n <- length(X)

  Z <- X/mean(X)

  Z_ord <- sort(Z)

  s <- c()

  for(i in 1:n){

    one <- abs( i/n - pexp(Z_ord[i]) )

    two <- abs( (i-1)/n - pexp(Z_ord[i]) )

    s[i] <- max(one, two)
  }

  return (max(s));

}

n <- 200 # sample size

TMC <- c() # list with values of test statistics under the null

for(i in 1:MC){

  X <- rexp(n)

  TMC[i] <- stat(X)


}

plot(ecdf(TMC))
```

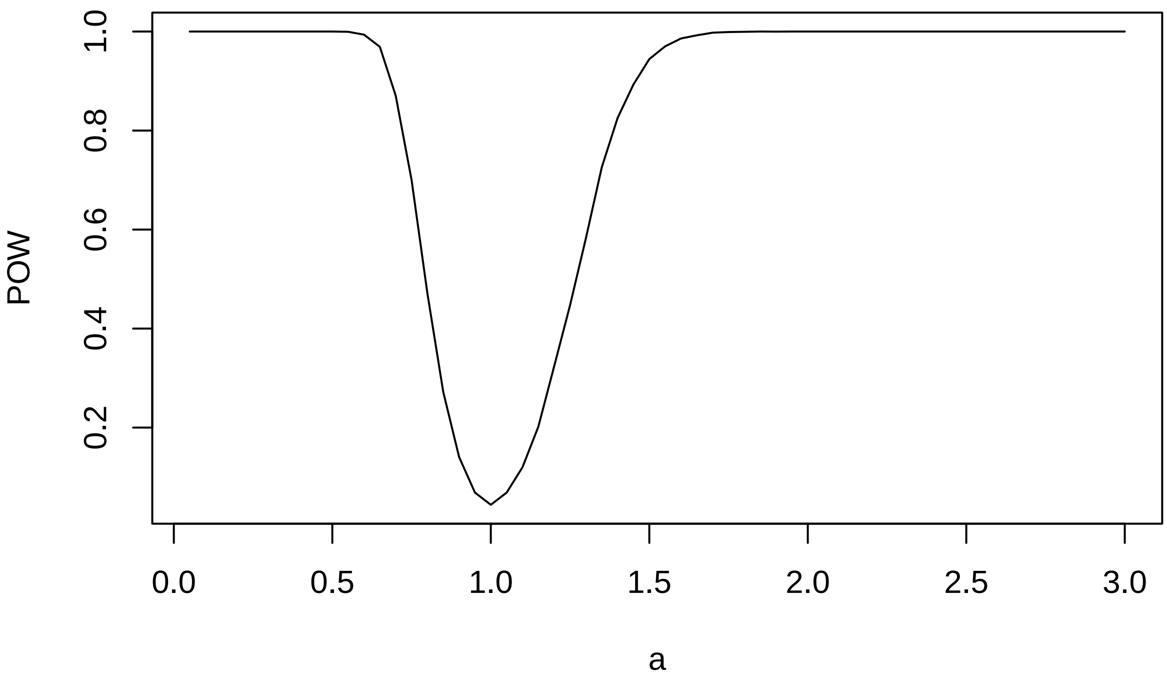


Empirical power of the test

Power of the test is defined as probability of rejection the null. We will calculate the empirical power given samples from $Gamma(a, 1)$ distribution with changing parameter $a = 0.05, 0, 1, \ldots, 3$. It is important to see that $Gamma(1, 1) = Exp(1)$. One would hope for the power of the test when drawing from exponential distribution to be equal to significanse level.

```
m <- 10000  # amount of test conducted to calculate empirical power of the test

a <- c(seq(0.05, 3, 0.05)) # list with different parameters "a"

POW <- c() # list with different values of empirical power for different "a"'s

alpha = 0.05 #significanse level

for(j in 1:60){

  M <- 0

  for(i in 1:m){

    X <- rgamma(n,a[j],1)  #distribution on which we calculate the power
                           #for a = 1 it is exponential distribution
    p <- (sum(stat(X)<TMC))/MC #p-value calculated using MC method

    if(p < alpha){M <- M + 1}

  }

  POW[j] <- M/m
```

```
}
plot(a,POW, type = 'l')  #plot of empirical power
```




```
POW
```

```
##  [1] 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
## [11] 0.9995 0.9937 0.9691 0.8704 0.7002 0.4708 0.2719 0.1404 0.0687 0.0441
## [21] 0.0687 0.1200 0.2019 0.3242 0.4472 0.5825 0.7257 0.8255 0.8930 0.9444
## [31] 0.9701 0.9860 0.9924 0.9977 0.9990 0.9995 1.0000 0.9998 1.0000 1.0000
## [41] 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
## [51] 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
```

As we can see power is lowest when we are close to the null hypothesis, but rises very quickly no matter how we change the parameter a.