

# Final Project

Wenjuan Bian

2023-11-15

## Introduction

In their 2014 study, Adewara and Mbata conducted a comprehensive analysis of three resampling techniques - Bootstrap Estimation (BE), Jackknife Estimation (JE), and K-Repeated Jackknife Estimation (KJE) - for estimating survival probabilities based on an exponential lifetime distribution. This paper aims to assess the performance of these methods by comparing their mean square error (MSE) and mean percentage error (MPE) using simulated data. Our project was to replicate their study, applying our theoretical knowledge practically and validate the original findings.

However, our replication efforts using R-Software uncovered notable inconsistencies. While we successfully replicated the results in Tables 1, 2, and 3, our findings for Table 4 significantly contrasted with the paper's claims and presented data. In the original paper, Tables 4 and 5 display identical values for the Jackknife and K-Repeated Jackknife methods, which does not support the conclusion that the K-Repeated Jackknife method is superior in performance and is more effective in reducing MSE. This contradiction is a significant point of inquiry, as it suggests potential issues in either the methodology, data analysis, or reporting within the original study.

Our replication efforts, particularly for Table 4 and 5, reveal a different narrative. We observed distinct differences between the Jackknife and K-Repeated Jackknife methods across various sample sizes and lambda values, which aligns with the expected outcome if the K-Repeated Jackknife method indeed offers an improvement over the standard Jackknife method. For instance, at a sample size of 10 and lambda of 0.5, the MSE values for Bootstrap, Jackknife, and K-Repeated Jackknife were 1.28645328, 0.1779106373, and 0.0001637587, respectively. These results clearly demonstrate the variance between the methods, contradicting the original paper's findings where such variance was absent.

Recognizing these inconsistencies, we reached out to the authors for clarification but have not received a meaningful response. Despite an initial reply stating, "I will look into this and get back to you," no further explanation or response has been provided by today (11/07/2023). This lack of feedback underscores the challenges and limitations often encountered in academic research, particularly when addressing potential errors or oversights in published work.

Our project not only highlights the importance of replication in scientific research but also underscores the necessity of addressing inconsistencies in published studies. The discrepancies we discovered call for a thorough re-examination of the original paper's methodologies and conclusions. As we continue to seek clarification from the authors, our work contributes to the broader discourse on the reliability and validation of statistical research methods.

# Data in the paper

```
size <- c(10, 10, 10, 10, 20, 20, 20, 20, 30, 30, 30, 30)
lambda <- c(0.5, 1.0, 1.5, 2.0, 0.5, 1.0, 1.5, 2.0, 0.5, 1.0, 1.5, 2.0)
```

```
A1 <- matrix(
  c(
    0.38672500, 2.27994293, 0.20192160, 0.01058359,
    0.11665480, 0.71411938, 0.48301090, 2.51951031,
    0.12661390, 0.75298901, 0.12473380, 1.94495769,
    4.42286640, 2.19289233, 1.99662440, 0.67531853,
    2.00704600, 0.04442099, 0.67780240, 1.05647689,
    2.34689070, 0.35021824, 0.75981590, 0.01373747,
    0.86211020, 1.01674933, 0.87373650, 0.08979746,
    1.03118540, 0.82205630, 0.21055390, 0.73813115,
    12.74339800, 0.02251542, 1.47164790, 0.07333057,
    1.96347260, 1.01602885, 0.52020820, 1.38869338
  ),
  nrow = 10,
  byrow = TRUE
)
```

```
A2 <- matrix(
  c(
    2.62129072, 0.13858938, 0.61025422, 0.20816883,
    3.66009415, 0.23677409, 1.39562321, 0.10508682,
    0.28671927, 0.65882621, 1.64585069, 0.16343073,
    0.53614812, 2.03913598, 1.11779841, 4.47429656,
    1.67262059, 3.07310425, 0.64918675, 0.31910786,
    0.08521250, 0.85047816, 0.05778008, 0.21707275,
    0.84341716, 0.93587816, 0.04898416, 0.04360368,
    1.87988871, 3.97283307, 0.12098188, 0.11505109,
    3.13213741, 1.91406287, 0.55509004, 0.24578897,
    6.82508233, 1.04861761, 0.19238185, 1.05875641,
    0.02900937, 0.03350444, 0.26953211, 0.00355591,
    1.74900498, 1.30175357, 0.01133644, 0.05922794,
    1.24431388, 1.52538027, 0.28231025, 0.20815442,
    1.12438751, 0.06504098, 2.50193518, 1.48214568,
    1.02901637, 1.90192968, 0.52524965, 0.89694421,
    2.19818977, 0.71117769, 0.20890800, 1.53327300,
    5.75705560, 2.66119925, 0.12742349, 1.11055824,
    0.67605113, 1.38839741, 0.09933566, 0.81368075,
    0.79778248, 0.30239189, 0.08220231, 0.46919296,
    1.06989471, 0.00054309, 0.22447181, 1.03504680
  ),
  nrow = 20,
  byrow = TRUE
)
```

```
A3 <- matrix(
```

```

c(
  1.09871091, 0.17085314, 0.24591320, 0.03006484,
  1.30597223, 0.31266839, 0.78554402, 0.01316790,
  0.43742079, 0.21960578, 0.26193004, 0.08580695,
  0.57034193, 1.09603105, 0.07651001, 0.06684230,
  1.02475744, 0.54796524, 0.29620540, 0.75349995,
  1.53172531, 0.13561858, 0.58560088, 0.56383185,
  0.57173448, 0.46805576, 0.56401687, 0.10014385,
  3.18835121, 0.58675535, 0.41054238, 0.39468972,
  1.01423546, 0.41116558, 1.42498655, 0.26076246,
  0.05104055, 3.42353290, 0.03212381, 0.38563026,
  2.99399940, 1.09817464, 0.19768464, 0.96558979,
  2.95112802, 0.35962685, 0.24295821, 0.27934118,
  4.74244122, 1.98862082, 0.16056365, 0.01547041,
  0.04628853, 1.11964839, 0.24140637, 0.34343548,
  5.35809191, 0.48539163, 0.06956623, 0.80092480,
  4.26185504, 1.09365222, 1.49188159, 1.49780414,
  0.04367701, 0.73949713, 0.22345808, 0.90349970,
  0.05851474, 0.34345758, 1.56937093, 0.18701462,
  0.46455153, 0.29243694, 0.59195204, 0.00288255,
  0.09371455, 0.25904558, 0.42176436, 0.74330912,
  2.96970220, 1.79180600, 1.11983745, 0.25455636,
  0.54133977, 0.01809066, 0.28416911, 0.81341579,
  1.29204462, 2.94010580, 0.65638599, 0.10230827,
  0.02923826, 0.79327459, 0.12515747, 2.24807427,
  2.28724168, 1.13577062, 2.26479793, 0.08718112,
  3.62406597, 0.05051664, 1.73361763, 1.15865739,
  0.55831489, 0.53652729, 0.36225826, 0.02484439,
  4.24494133, 1.08361205, 1.41035274, 0.48756216,
  0.13192788, 0.25488727, 0.93471466, 0.25596331,
  0.39449119, 1.92176066, 0.27655395, 0.09251206
),
nrow = 30,
byrow = TRUE
)

data.paper <- list(A1[,1], A1[,2], A1[,3], A1[,4], A2[,1], A2[,2], A2[,3], A2[,4], A3[,1], A3[,
2], A3[,3], A3[,4])

# data.paper

```

## Table 2

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

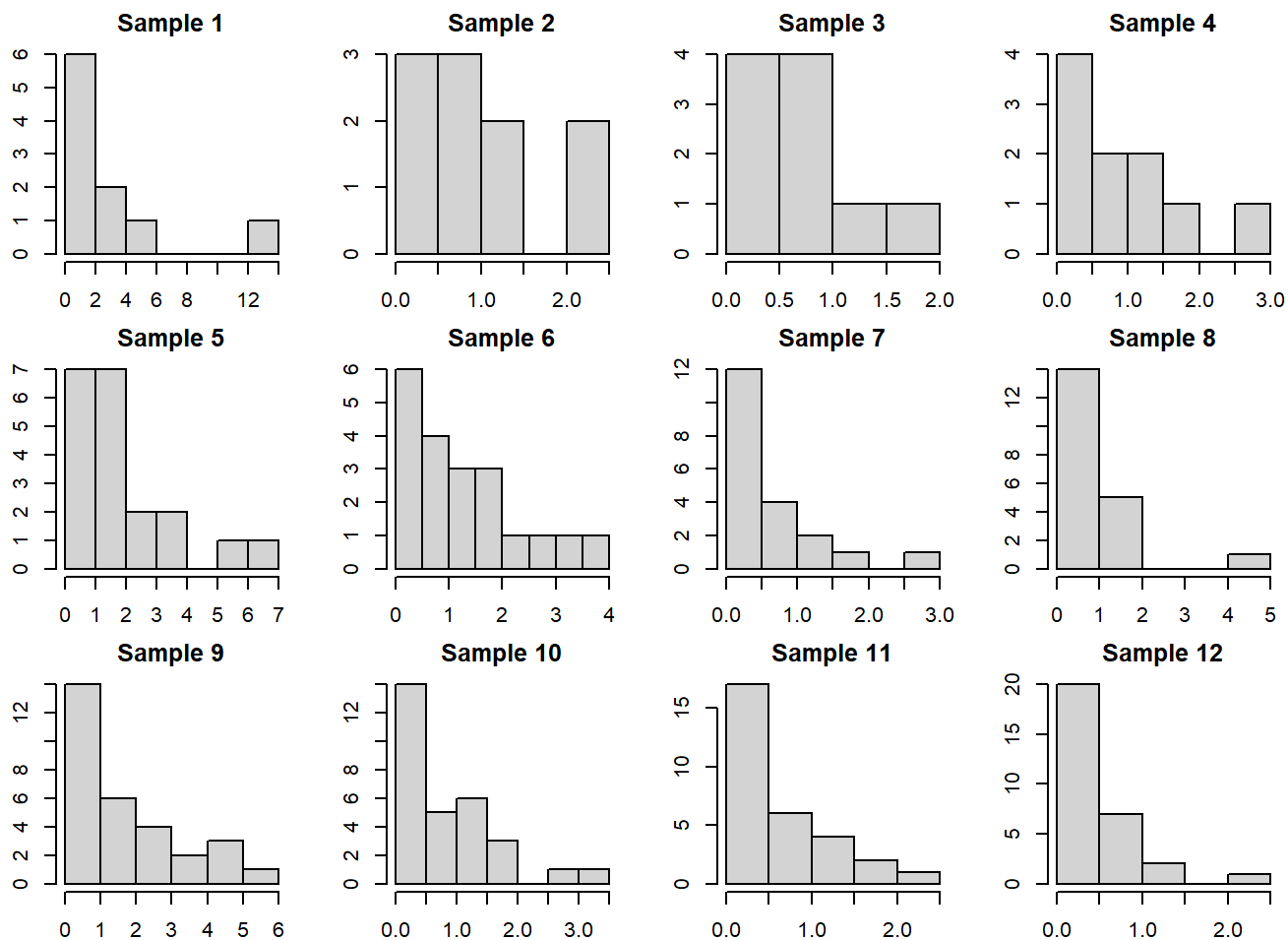
```
library(purrr)
```

```
par(mfrow = c(3, 4), mar = c(2, 2, 2, 2))
```

```
for (i in 1:length(data.paper)) {
```

```
  hist(data.paper[[i]], main = paste("Sample", i), xlab = "Value", ylab = "Frequency")
```

```
}
```



```

calculate_stats <- function(sample) {
  c(mean = mean(sample), median = median(sample), sd = sd(sample))
}

stats <- lapply(data.paper, calculate_stats)

df.stats <- cbind(N= size, lambda=lambda, do.call(rbind, stats))
rownames(df.stats) <- paste("Sample", 1:length(data.paper))

print(df.stats)

```

##		N	lambda	mean	median	sd
##	Sample 1	10	0.5	2.6006963	1.4973290	3.7961509
##	Sample 2	10	1.0	0.9211933	0.7875227	0.7792381
##	Sample 3	10	1.5	0.7320056	0.5990053	0.5961606
##	Sample 4	10	2.0	0.8510537	0.7067248	0.8785794
##	Sample 5	20	0.5	1.8608658	1.1843507	1.8059576
##	Sample 6	20	1.0	1.2379809	0.9922479	1.0967927
##	Sample 7	20	1.5	0.5363318	0.2470020	0.6540289
##	Sample 8	20	2.0	0.7281072	0.2824484	1.0112265
##	Sample 9	30	0.5	1.5960620	1.0194965	1.6278563
##	Sample 10	30	1.0	0.8559385	0.5422463	0.8296031
##	Sample 11	30	1.5	0.6353941	0.3864003	0.5880883
##	Sample 12	30	2.0	0.4639596	0.2700518	0.5150784

## Table 3 and 4

```
library(boot)
```

```
## Warning: package 'boot' was built under R version 4.3.2
```

```

sur.f <- function(t, lambda) {
  exp(-lambda * t)
}

boot.f <- function(data, indices) {
  sample1 <- data[indices]
  theta.b <- mean(sample1)
  return(theta.b)
}

# Create empty lists/vectors to store the results
boot.re <- numeric(length(data.paper))
boot.mse <- numeric(length(data.paper))
boot.sur <- numeric(length(data.paper))
boot.mpe <- numeric(length(data.paper))

# Iterate over each dataset in data.paper and apply boot.f
for (i in 1:length(data.paper)) {
  boot.obj <- boot(data = data.paper[[i]], statistic = boot.f, R = 2000)
  y <- boot.obj$t
  boot.re[i] <- mean(y)
  boot.sur[i] <- mean(sur.f(data.paper[[i]], lambda = 1 / boot.re[i]))
  boot.mse[i] <- sum((y - boot.re[i])^2) / (length(y) - 1)
  boot.mpe[i] <- mean(abs(y - boot.re[i])) / boot.re[i]
}

boot.results <- data.frame(Sample = 1:length(data.paper), Mean = boot.sur, MSE = boot.mse, MPE=boot.mpe)

# Jackknife and Repeated jackknife
jk.re.an <- function(data) {

  jk.re1 <- function(data1) {
    n <- length(data1)
    jk1 <- numeric(n)

    for (i in 1:n) {
      sample1 <- data1[-i]
      jk1[i] <- mean(sample1)
    }
    return(jk1)
  }

  jk.ret1 <- jk.re1(data)

  mean.jk <- mean(jk.ret1)
  jk.1f <- mean(sur.f(data, 1 / mean.jk))
  mse.jk1 <- sum((jk.ret1 - mean.jk)^2) / (length(jk.ret1) - 1)
  mpe.jk1 <- mean(abs(jk.ret1 - mean.jk)) / mean.jk

  #Repeated k jackknife

```

```

k=length(data)

jk.results <- vector("list", k)

jk.results[[1]] <- jk.ret1

for (i in 2:k) {
  jk.results[[i]] <- jk.re1(jk.results[[i - 1]])
}

jkre <- unlist(jk.results)
length.jkre <- length(jkre)
jkre.mean <- mean(jkre)
jkre.mse <- var(jkre)/(length(jk.results)-1) #sum((jkre - jkre.mean)^2) / ((Length(jkre) - 1)*(Length(jkre))) #var(jkre)
thetahatmse <- var(jkre)/(length(jk.results)-1)
thetahatmpe <- mean(abs(jkre-jkre.mean)/jkre.mean)
jkrkf <- mean(sur.f(data, 1/jkre.mean))
return(list(jk.1f=jk.1f, mse.jk1=mse.jk1, mpe.jk1=mpe.jk1, jkrkf = jkrkf, jkre.mse = jkre.mse, length.jkre= length.jkre, thetahatmse=thetahatmse, thetahatmpe=thetahatmpe))
}

jk.re.results <- lapply(data.paper, jk.re.an)

jk.sur <- sapply(jk.re.results, function(x) x$jk.1f)
mse.jk <- sapply(jk.re.results, function(x) x$mse.jk1)
mpe.jk <- sapply(jk.re.results, function(x) x$mpe.jk1)
jk.re.sur <- sapply(jk.re.results, function(x) x$jkrkf)
mse.jk.re <- sapply(jk.re.results, function(x) x$jkre.mse)
jk.re.length <- sapply(jk.re.results, function(x) x$length.jkre)
jk.re.thetamse <- sapply(jk.re.results, function(x) x$thetahatmse)
jk.re.thetampe <- sapply(jk.re.results, function(x) x$thetahatmpe)
jk.re.mpe <- jk.re.thetampe

```

## table 3

```

df2 <- data.frame(
  SampleSize = size,
  Lambda = lambda,
  Bootstrap = boot.results[,2],
  Jackknife = jk.sur,
  JackknifeRe = jk.re.sur
)

df2

```

```
##      SampleSize Lambda Bootstrap Jackknife JackknifeRe
## 1          10      0.5 0.5711603 0.5688799 0.5688799
## 2          10      1.0 0.4763098 0.4764569 0.4764569
## 3          10      1.5 0.4610719 0.4613285 0.4613285
## 4          10      2.0 0.5300484 0.5299378 0.5299378
## 5          20      0.5 0.4894883 0.4917777 0.4917777
## 6          20      1.0 0.4914196 0.4902400 0.4902400
## 7          20      1.5 0.5442241 0.5449304 0.5449304
## 8          20      2.0 0.5567579 0.5535801 0.5535801
## 9          30      0.5 0.5261816 0.5274456 0.5274456
## 10         30      1.0 0.4923180 0.4918826 0.4918826
## 11         30      1.5 0.4913832 0.4910998 0.4910998
## 12         30      2.0 0.5279396 0.5281186 0.5281186
```

```
# Print the data frame
df2
```

```
##      SampleSize Lambda Bootstrap Jackknife JackknifeRe
## 1          10      0.5 0.5711603 0.5688799 0.5688799
## 2          10      1.0 0.4763098 0.4764569 0.4764569
## 3          10      1.5 0.4610719 0.4613285 0.4613285
## 4          10      2.0 0.5300484 0.5299378 0.5299378
## 5          20      0.5 0.4894883 0.4917777 0.4917777
## 6          20      1.0 0.4914196 0.4902400 0.4902400
## 7          20      1.5 0.5442241 0.5449304 0.5449304
## 8          20      2.0 0.5567579 0.5535801 0.5535801
## 9          30      0.5 0.5261816 0.5274456 0.5274456
## 10         30      1.0 0.4923180 0.4918826 0.4918826
## 11         30      1.5 0.4913832 0.4910998 0.4910998
## 12         30      2.0 0.5279396 0.5281186 0.5281186
```

## table 4

```
df3 <- data.frame(
  SampleSize = size,
  Lambda = lambda,
  Bootstrap = boot.results[,3],
  Jackknife = mse.jk,
  JackknifeRe = mse.jk.re
)

df3
```



##	SampleSize	Lambda	Bootstrap	Jackknife	JackknifeRe
## 1	10	0.5	1.291648236	0.1779106373	1.819541e-03
## 2	10	1.0	0.057181645	0.0074964448	7.666819e-05
## 3	10	1.5	0.031491751	0.0043877462	4.487468e-05
## 4	10	2.0	0.069757120	0.0095296511	9.746234e-05
## 5	20	0.5	0.151923761	0.0090345787	2.270595e-05
## 6	20	1.0	0.057916643	0.0033322834	8.374786e-06
## 7	20	1.5	0.020045212	0.0011849137	2.977958e-06
## 8	20	2.0	0.049415782	0.0028326288	7.119041e-06
## 9	30	0.5	0.083794068	0.0031509108	3.509079e-06
## 10	30	1.0	0.022800519	0.0008183606	9.113847e-07
## 11	30	1.5	0.011054816	0.0004112341	4.579796e-07
## 12	30	2.0	0.008512397	0.0003154647	3.513240e-07

## table 5

```
df4 <- data.frame(
  SampleSize = size,
  Lambda = lambda,
  Bootstrap = boot.results[,3],
  Jackknife = mpe.jk,
  JackknifeRe = jk.re.mpe
)
df4
```

##	SampleSize	Lambda	Bootstrap	Jackknife	JackknifeRe
## 1	10	0.5	1.291648236	0.10223648	0.0115016042
## 2	10	1.0	0.057181645	0.06804798	0.0076553976
## 3	10	1.5	0.031491751	0.06599229	0.0074241326
## 4	10	2.0	0.069757120	0.09153159	0.0102973039
## 5	20	0.5	0.151923761	0.03690325	0.0019476713
## 6	20	1.0	0.057916643	0.03671500	0.0019377359
## 7	20	1.5	0.020045212	0.04633243	0.0024453228
## 8	20	2.0	0.049415782	0.04756272	0.0025102547
## 9	30	0.5	0.083794068	0.02975887	0.0010273897
## 10	30	1.0	0.022800519	0.02491690	0.0008602264
## 11	30	1.5	0.011054816	0.02546179	0.0008790379
## 12	30	2.0	0.008512397	0.02889973	0.0009977289