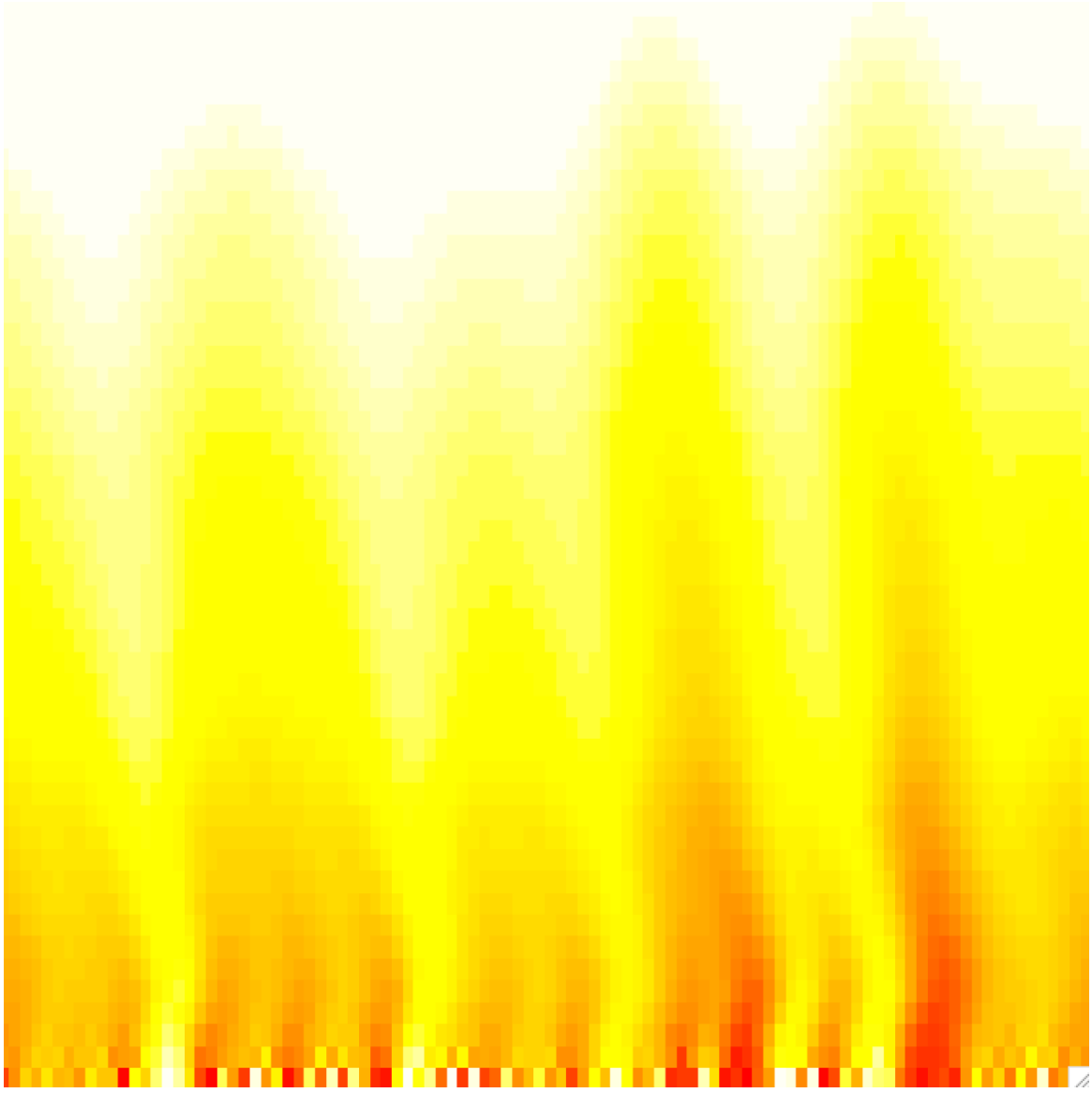
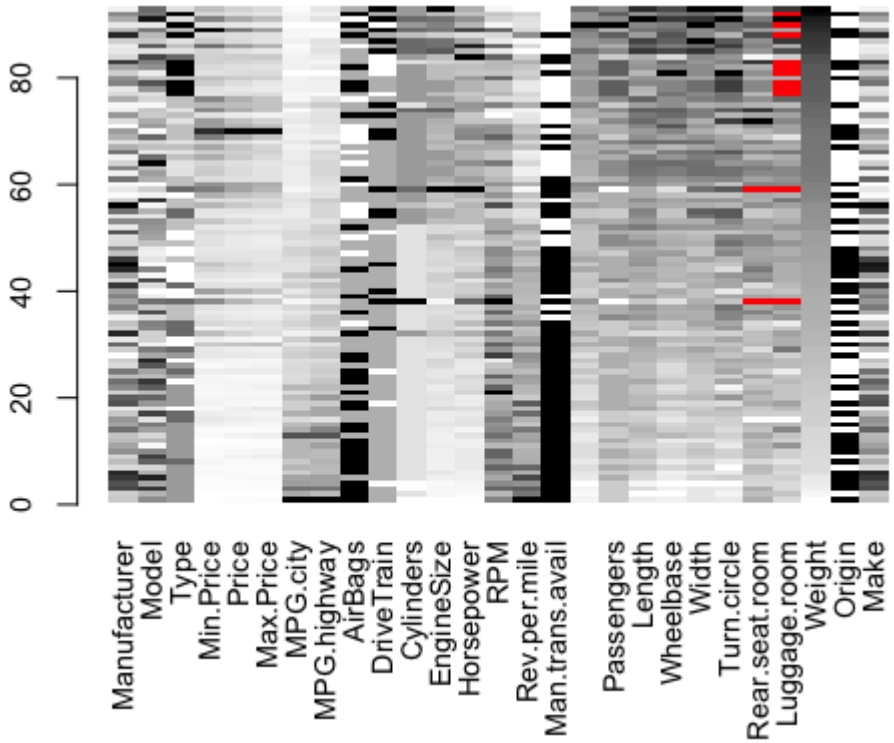
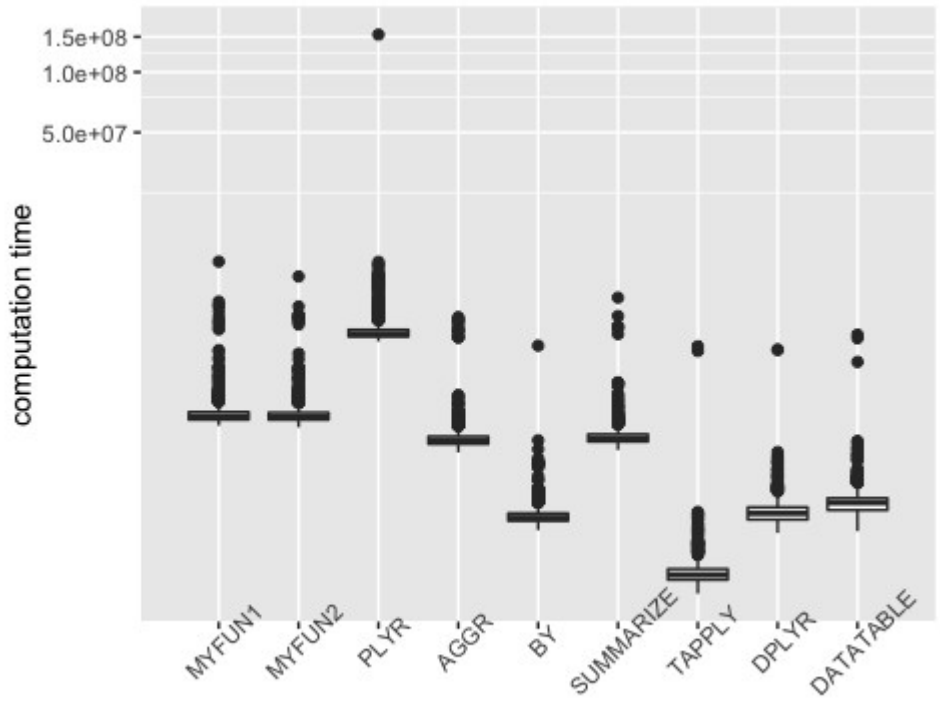
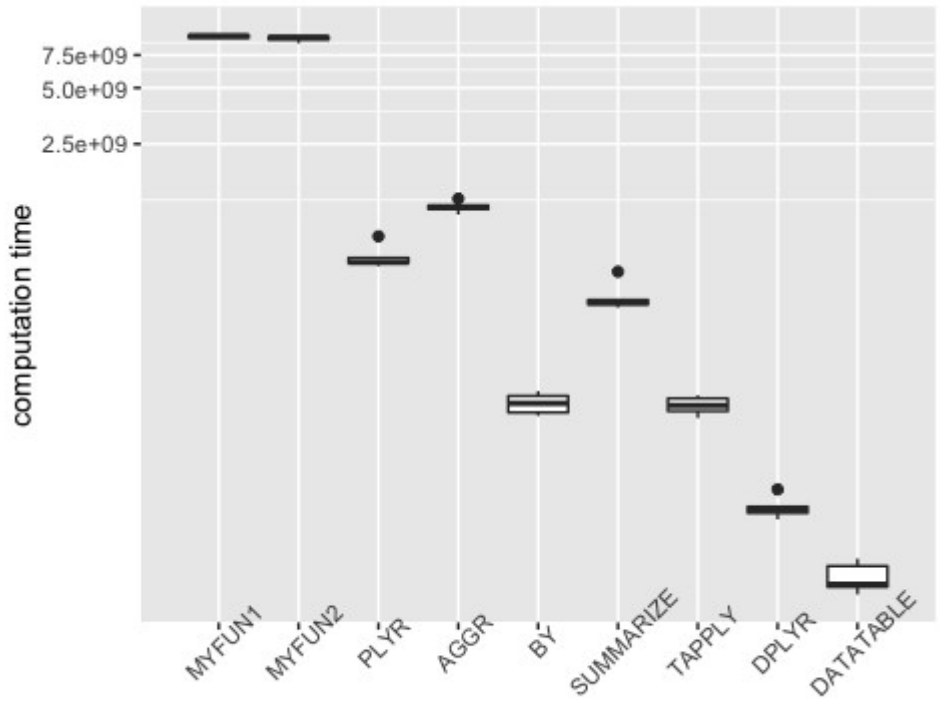
Chapter 1: Introduction

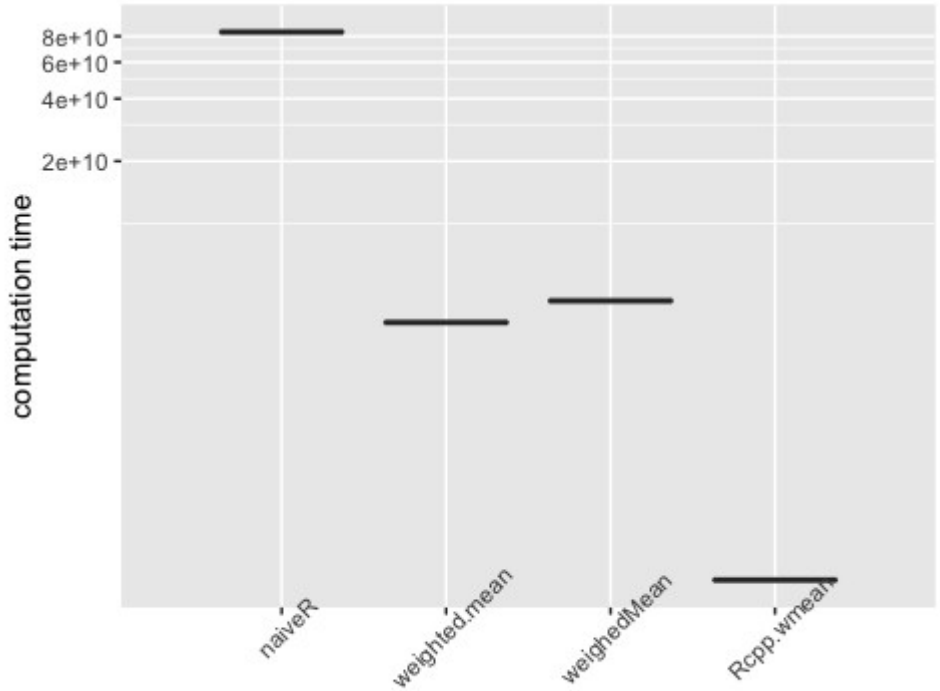


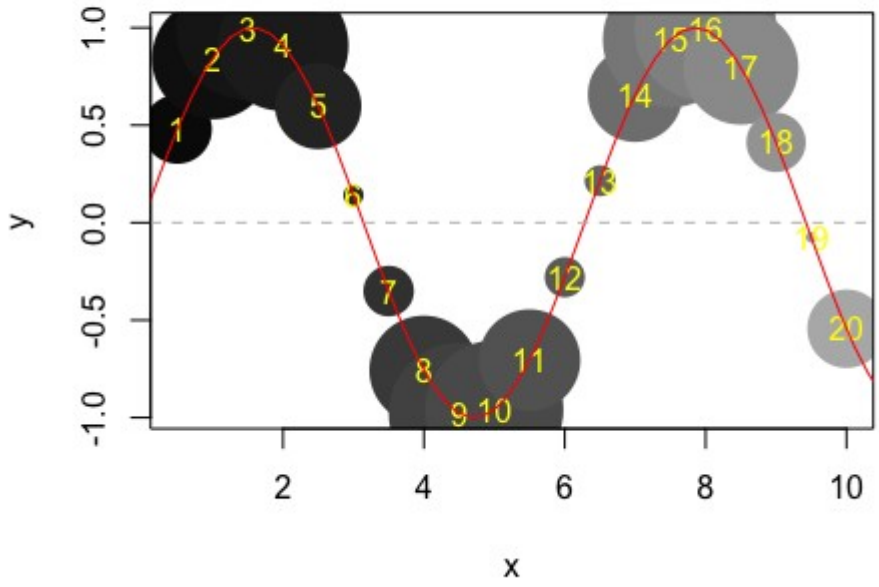
Chapter 2: R and High-Performance Computing

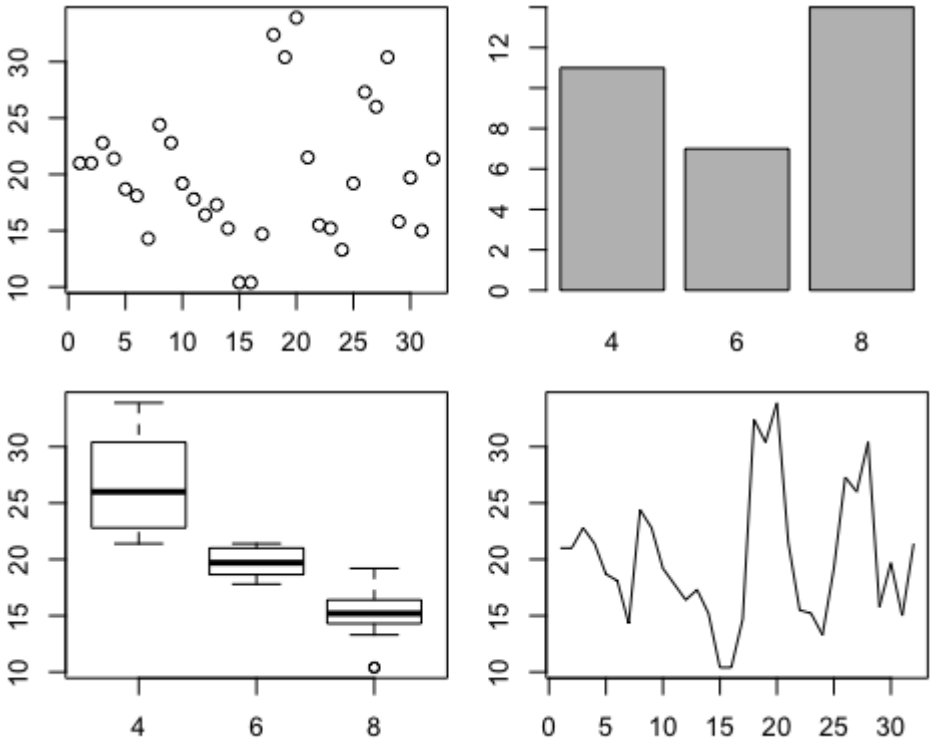


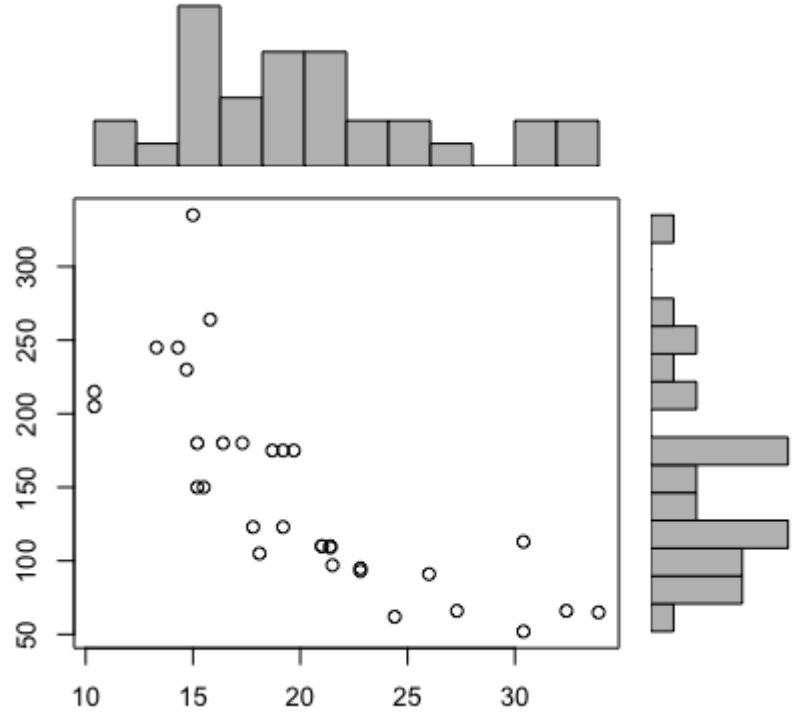


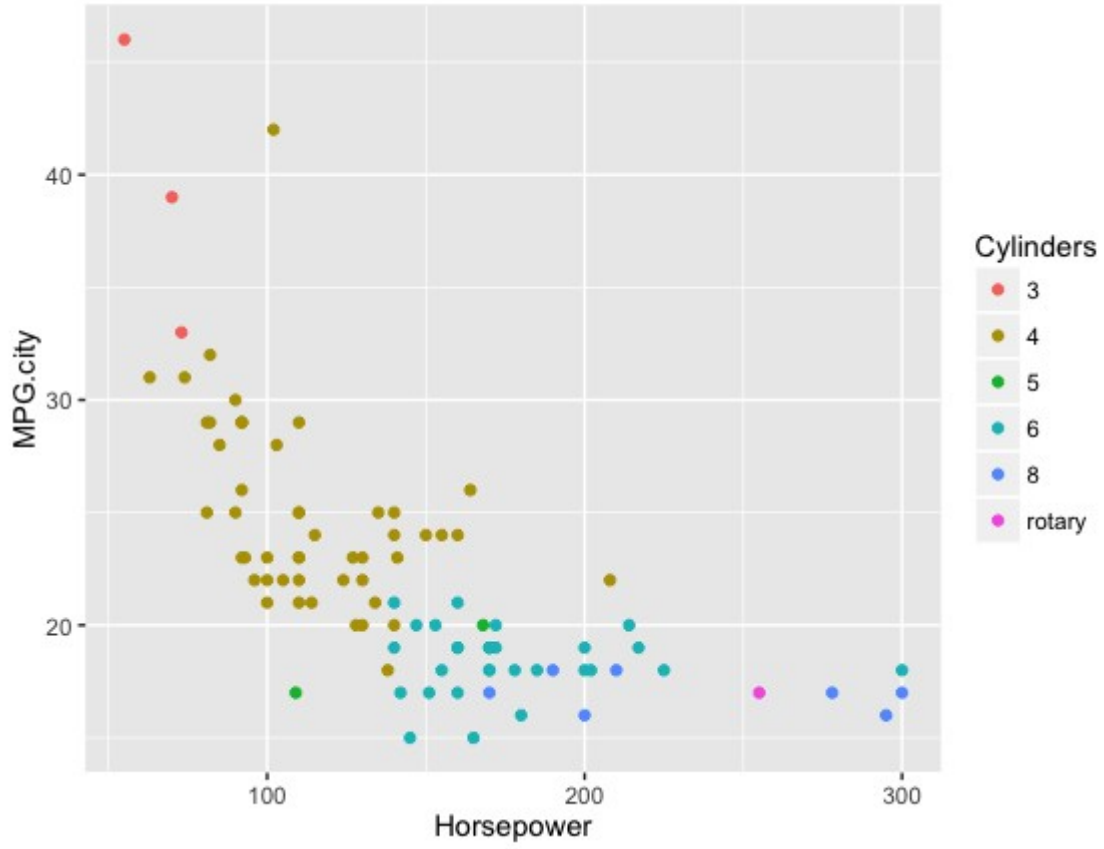

7.5e+09
5.0e+09
2.5e+09
computation time
MYFUN1
MYFUN2
PLYR
AGGR
BY
SUMMARIZE
TAPPLY
DPLYR
DATATABLE

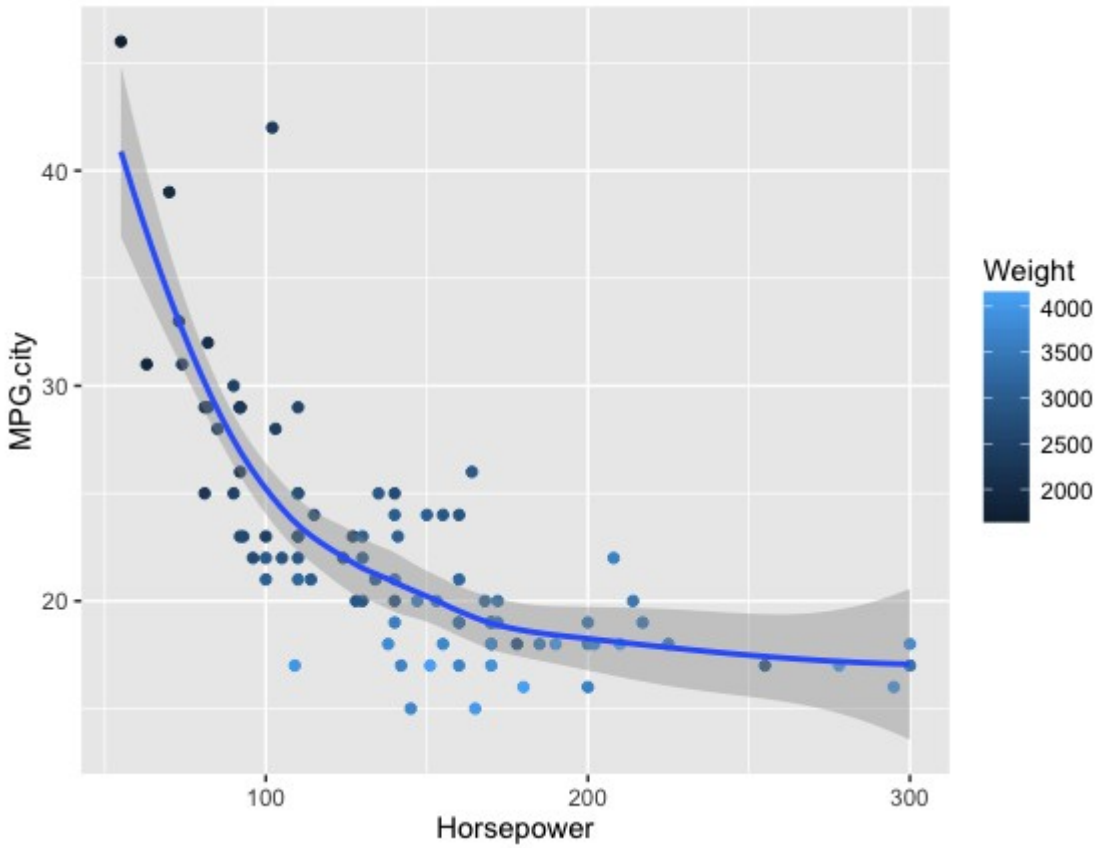

8e+10
6e+10
4e+10
2e+10
computation time
naiveR
weighted.mean
weighedMean
Rcpp.wmean

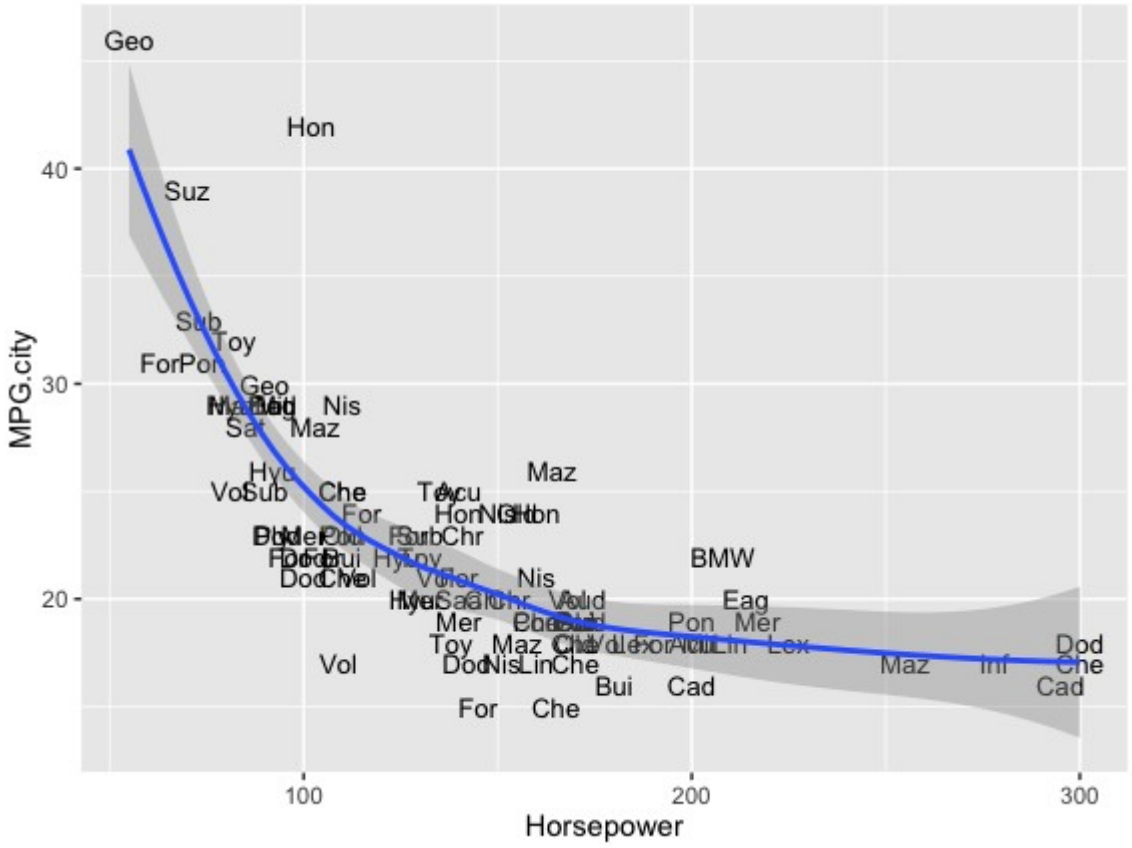

y
x
1.0
0.5
0.0
-0.5
-1.0
2
4
6
8
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

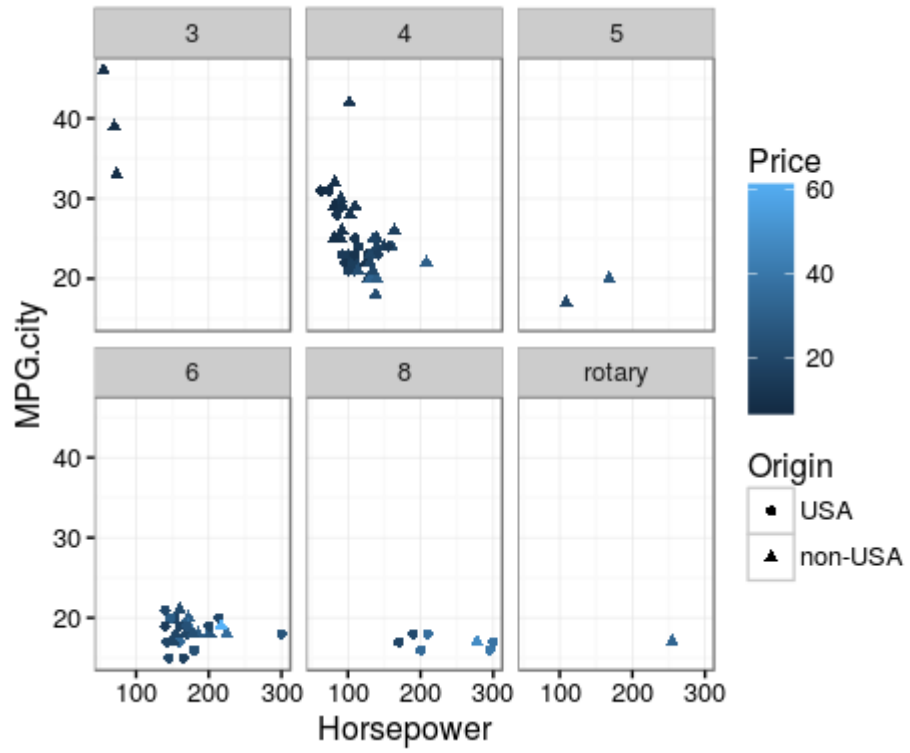

10
15
20
25
30
0
5
10
15
20
25
30
0
2
4
6
8
10
12
4
6
8
10
15
20
25
30
4
6
8
10
15
20
25
30
0
5
10
15
20
25
30


50
100
150
200
250
300
10
15
20
25
30

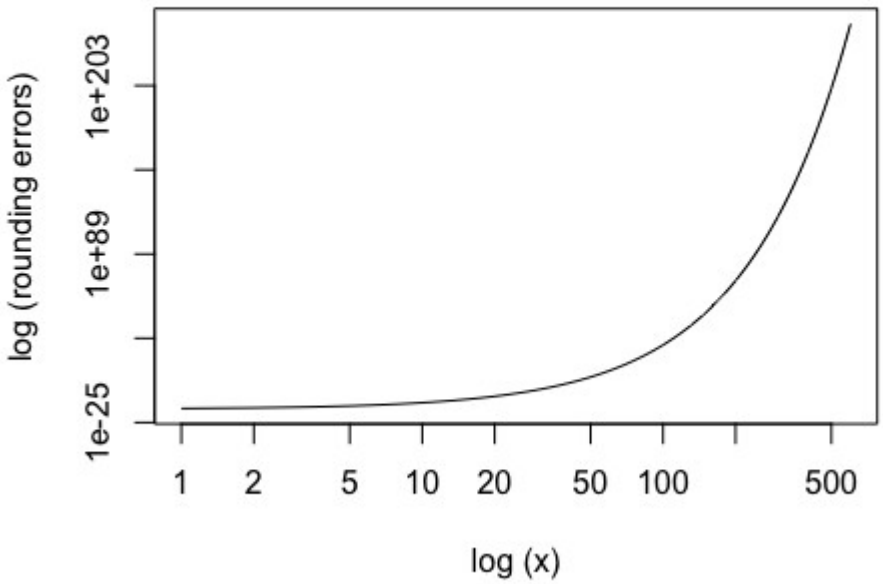

MPG.city
40
30
20
100
200
300
Horsepower
Cylinders
3
4
5
6
8
rotary

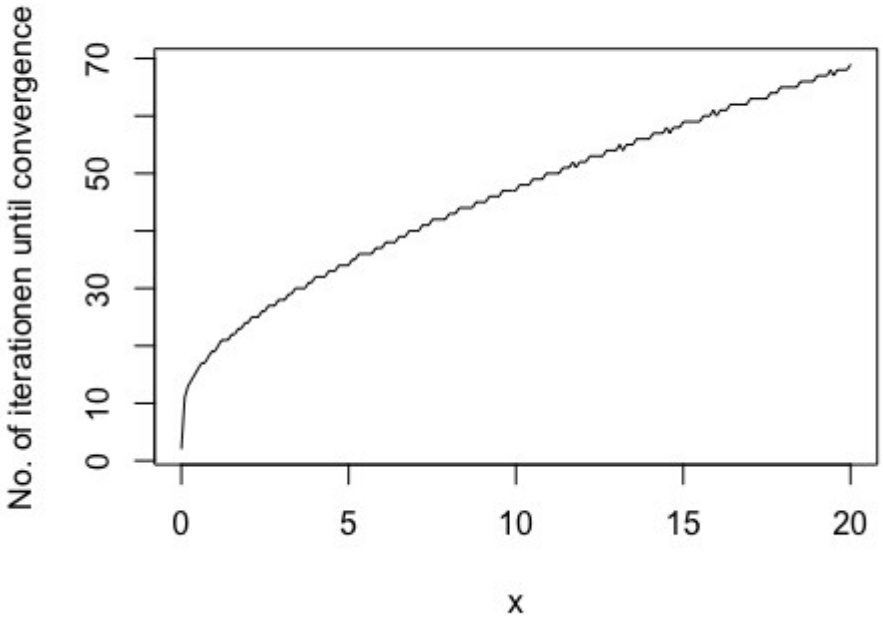

MPG.city
40
30
20
100
200
300
Horsepower
Weight
4000
3500
3000
2500
2000


MPG.city
40
30
20
100
200
300
Horsepower
Geo
Hon
Suz
Sub
Toy
Geo
Nis
Maz
Hyu
Vol
Sub
Che
Maz
BMW
Eag
Nis
Mer
Pon
Mer
Dod
Che
Vol
Maz
Inf
Cad
Bui
Cad
For
Che

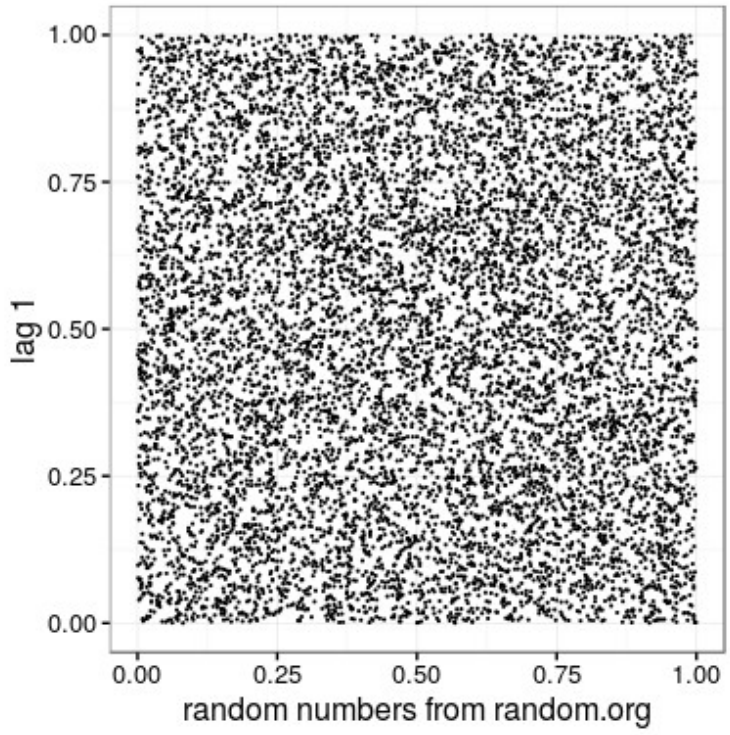


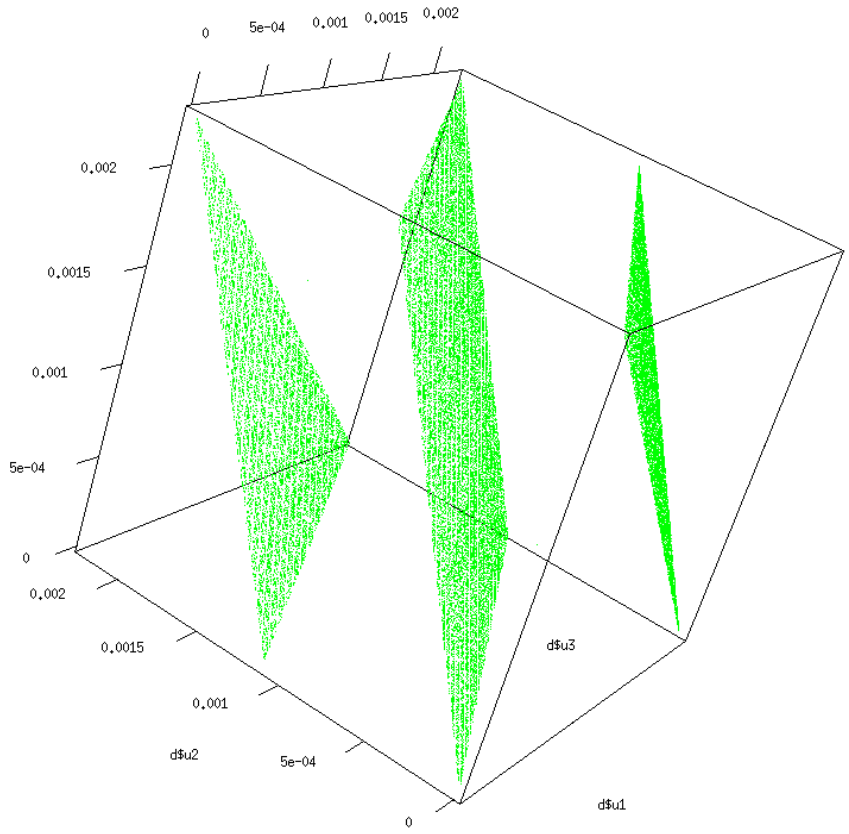
Chapter 3: The Discrepancy between Pencil-Driven Theory and Data-Driven Computational Solutions

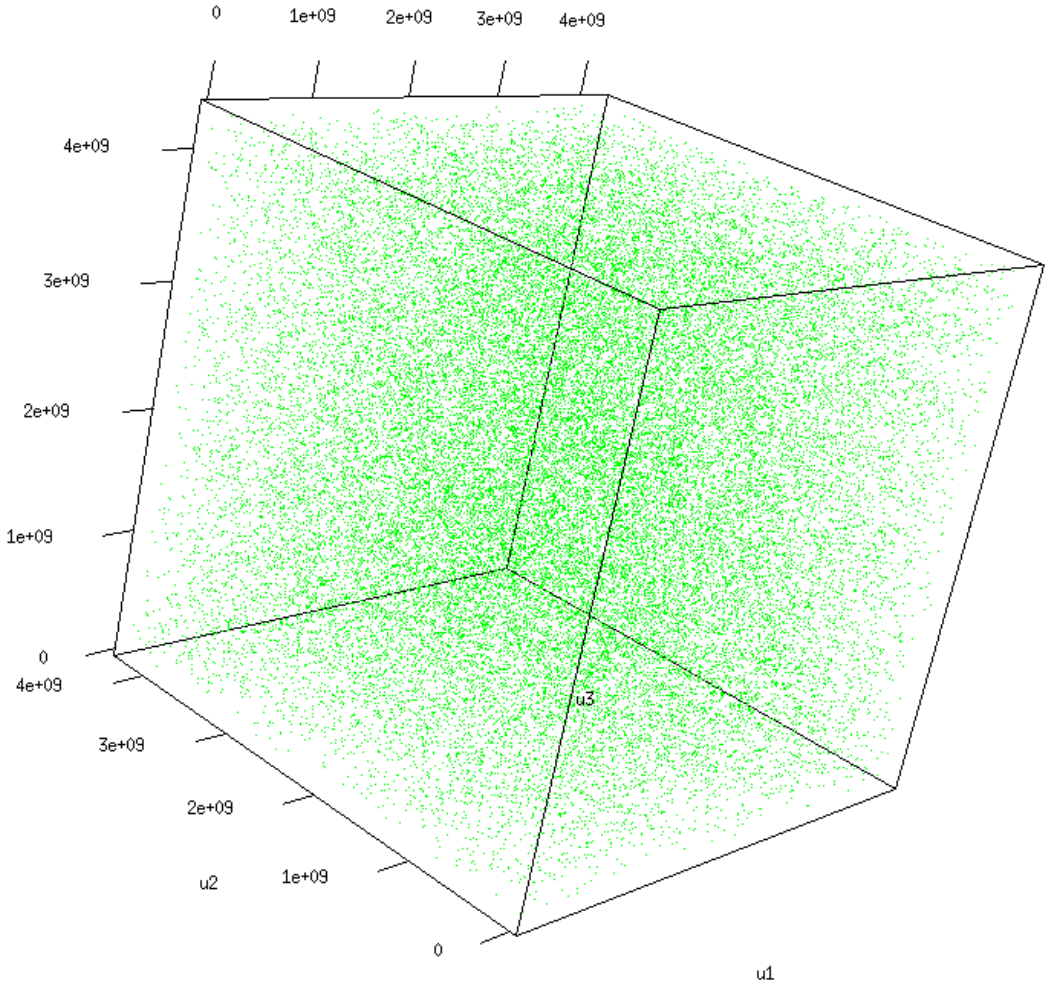


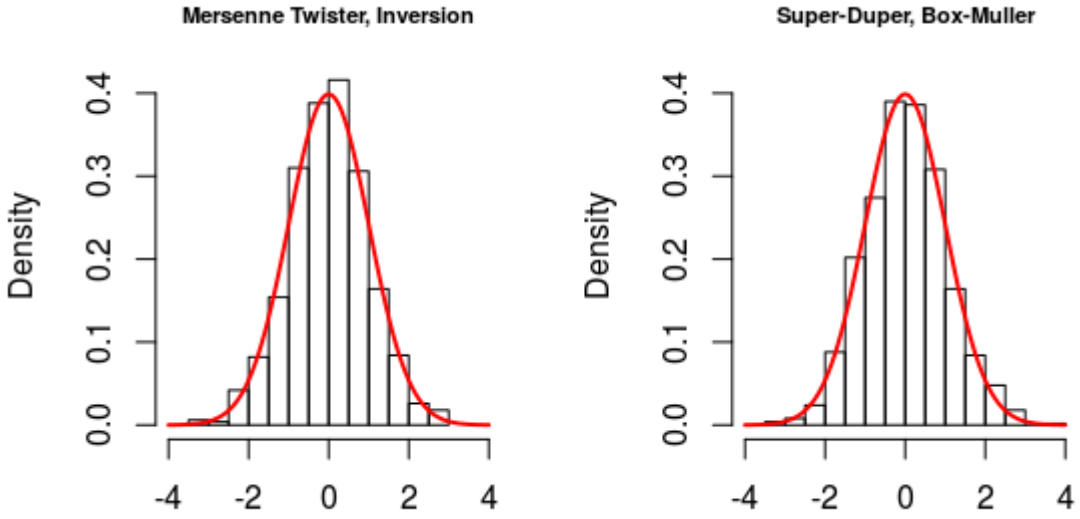


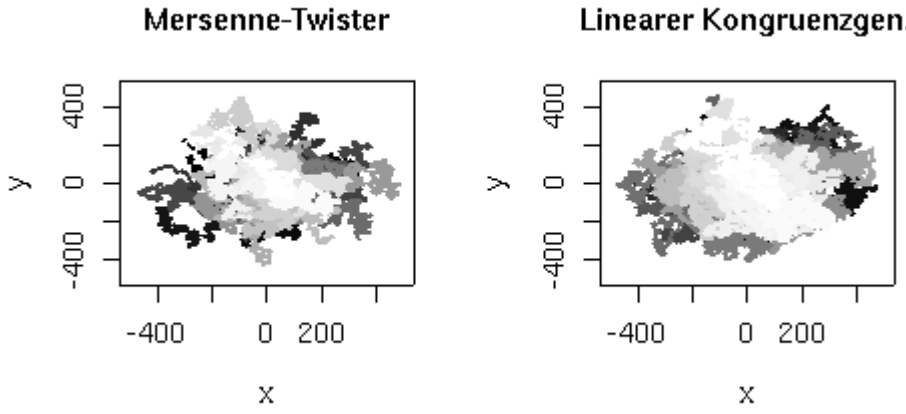
Chapter 4: Simulation of Random Numbers

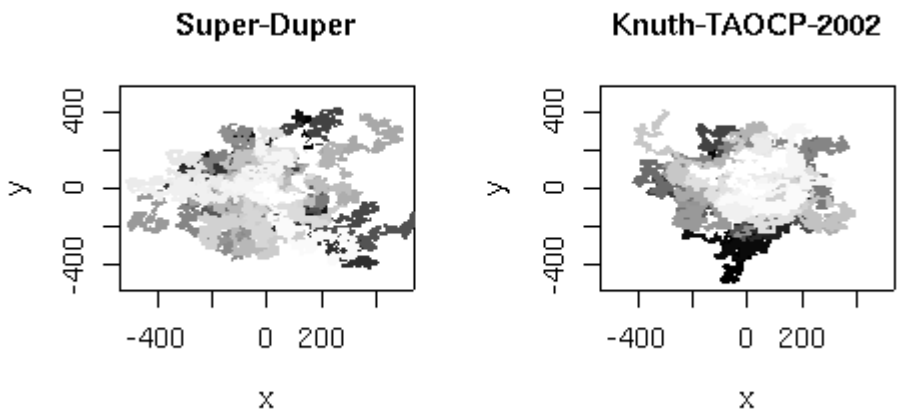


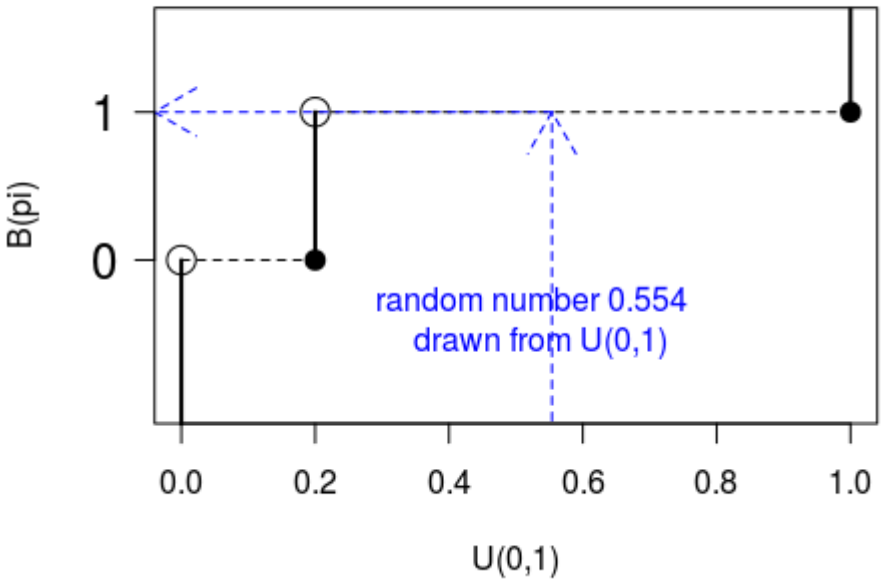

0
5e-04
0.001
0.0015
0.002
0.002
0.0015
0.001
5e-04
0
0.002
0.0015
0.001
d$u2
5e-04
0
d$u3
d$u1

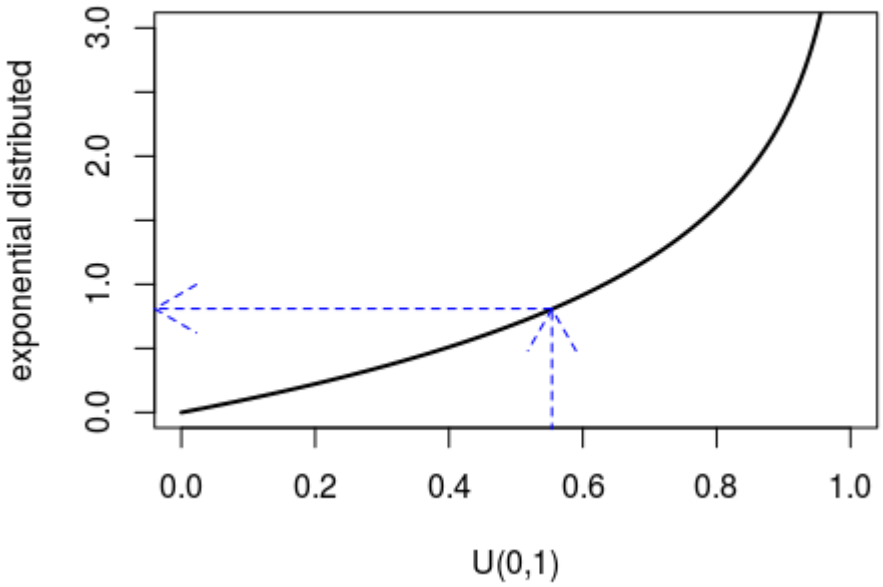

0
1e+09
2e+09
3e+09
4e+09
4e+09
3e+09
2e+09
1e+09
0
4e+09
3e+09
2e+09
u2
1e+09
0
u3
u1

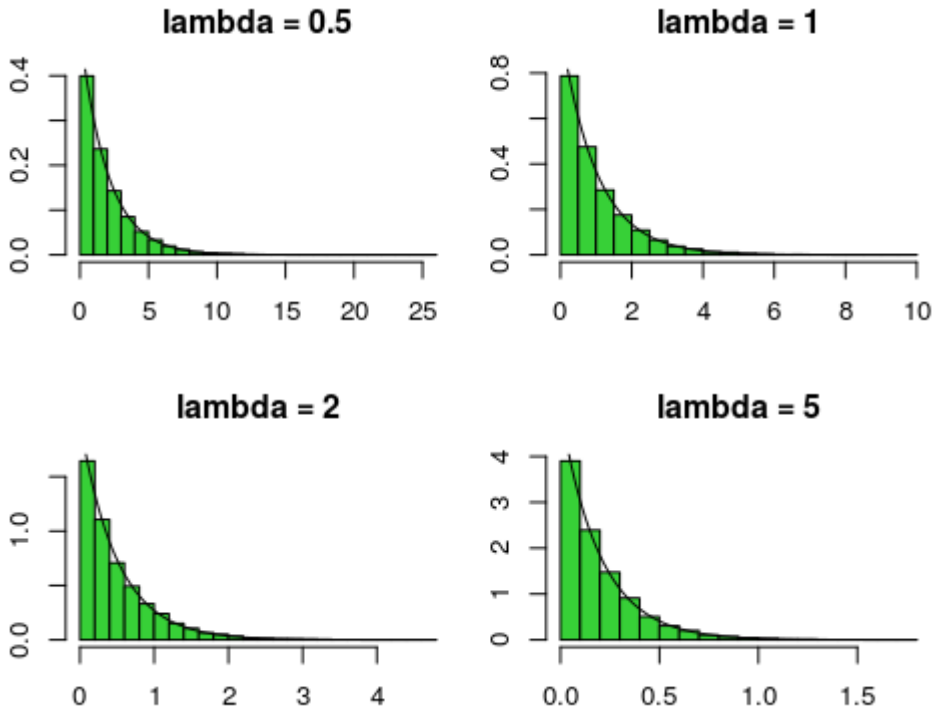

Mersenne Twister, Inversion
Super-Duper, Box-Muller
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4

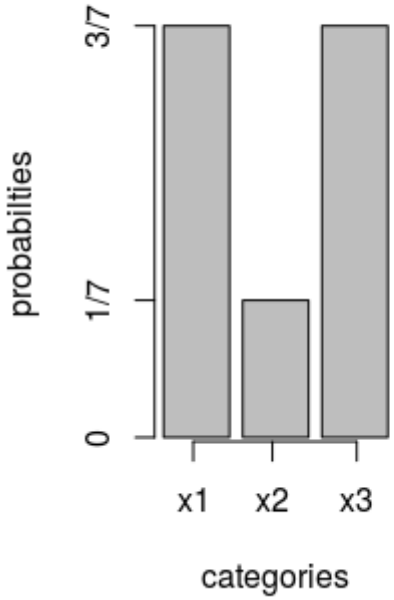

Mersenne-Twister
Linearer Kongruenzgen.
y
x
400
0
-400
-400
0
200

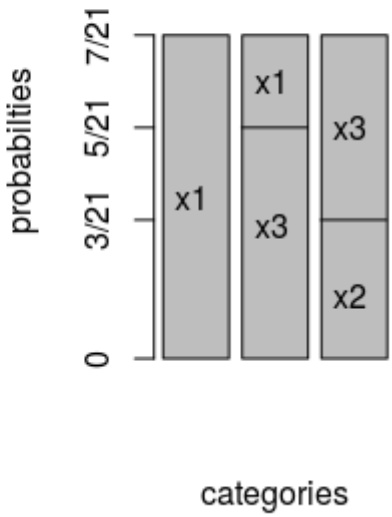

Super-Duper
Knuth-TAOCP-2002
y
x
400
0
-400
-400
0
200

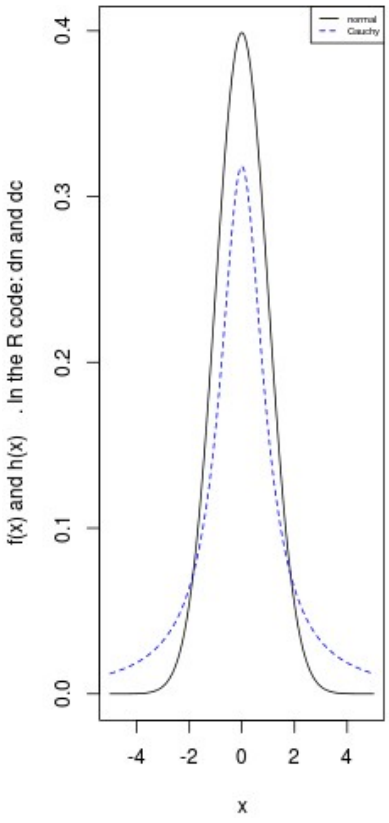

B(pi)
1
0
random number 0.554
drawn from U(0,1)
0.0
0.2
0.4
0.6
0.8
1.0
U(0,1)

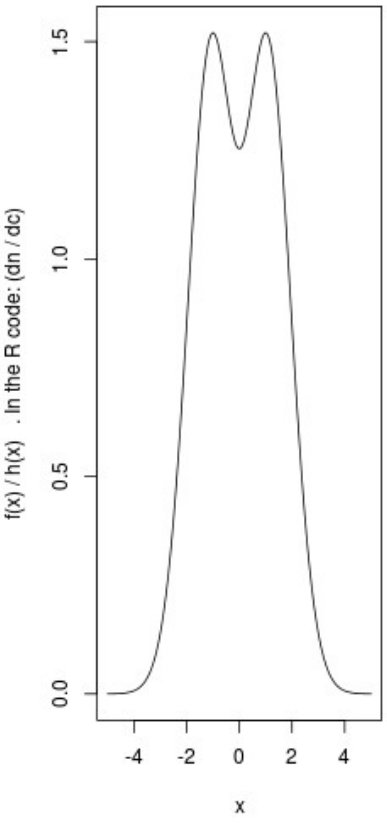

exponential distributed
0.0
1.0
2.0
3.0
0.0
0.2
0.4
0.6
0.8
1.0
U(0,1)

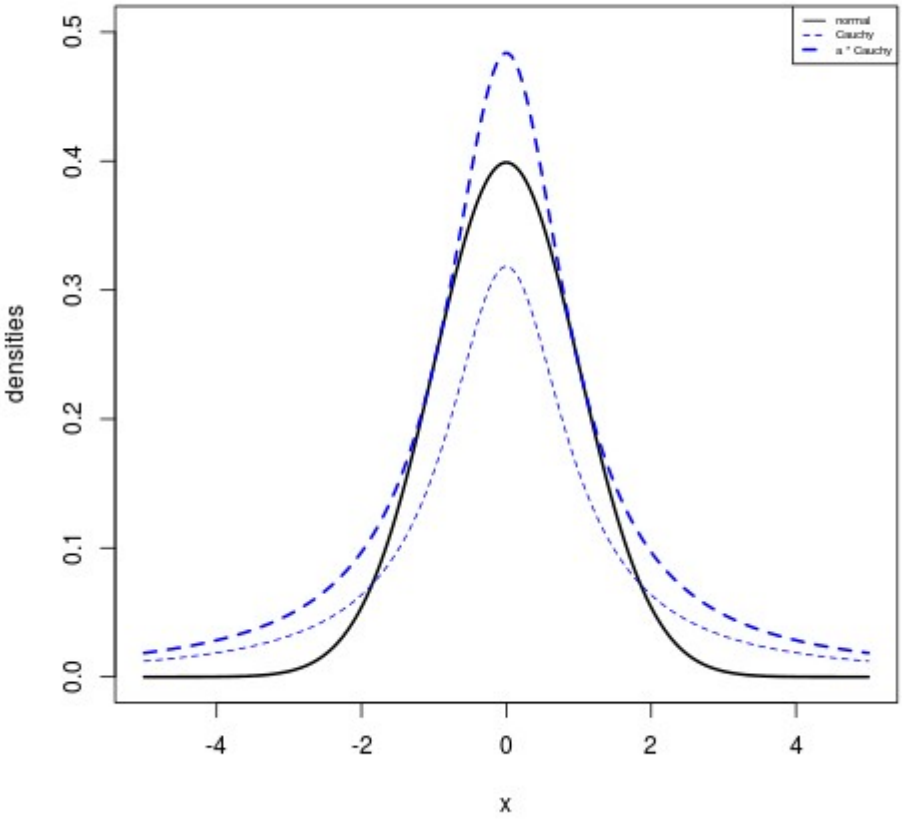

lambda = 0.5
lambda = 1
lambda = 2
lambda = 5

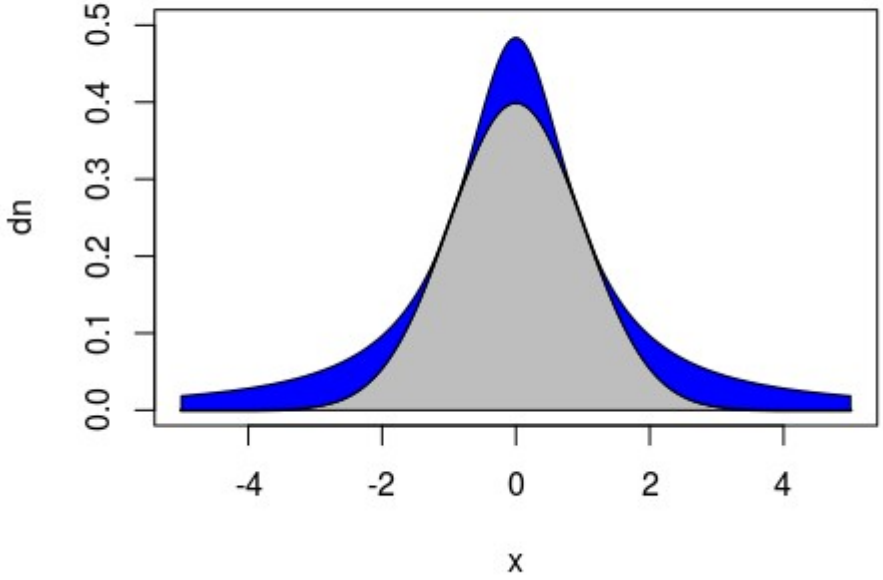

probabilties
0
1/7
3/7
x1
x2
x3
categories

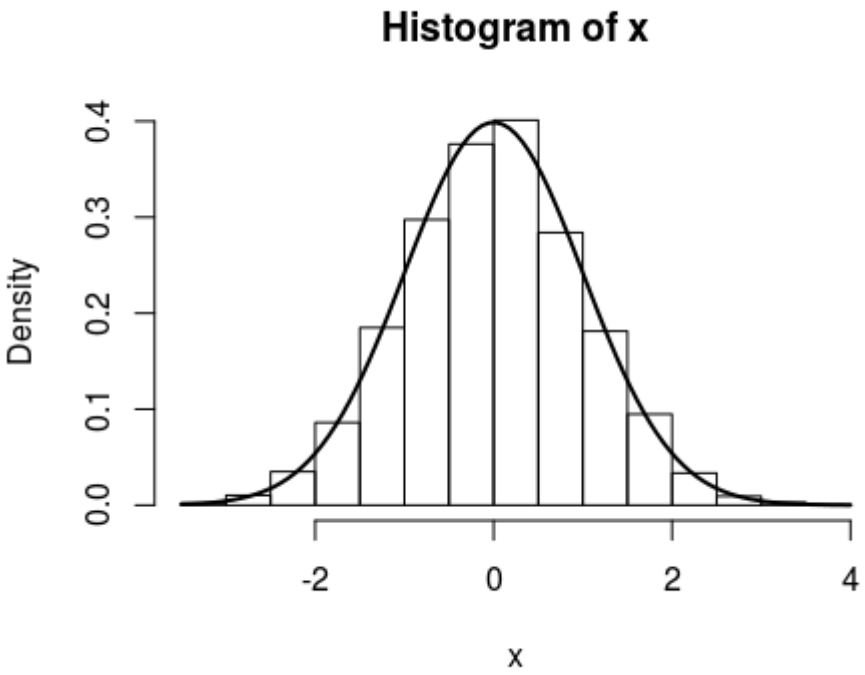

probabilties
0
3/21
5/21
7/21
x1
x1
x3
x3
x2
categories

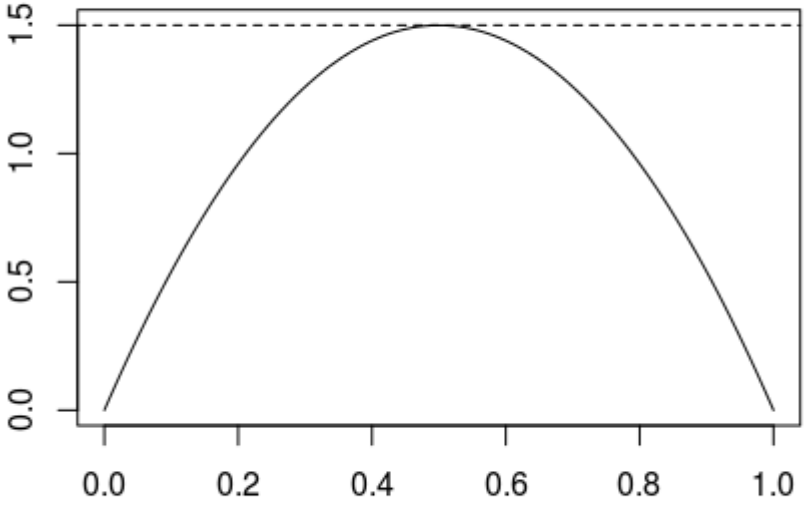

f(x) and h(x) . In the R code: dn and dc
normal
Cauchy
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x

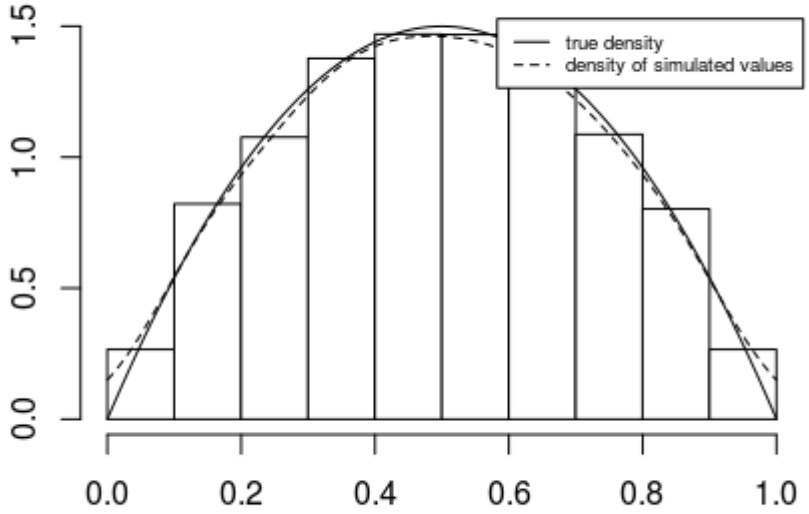

f(x) / h(x) . In the R code: (dn / dc)
0.0
0.5
1.0
1.5
-4
-2
0
2
4
x

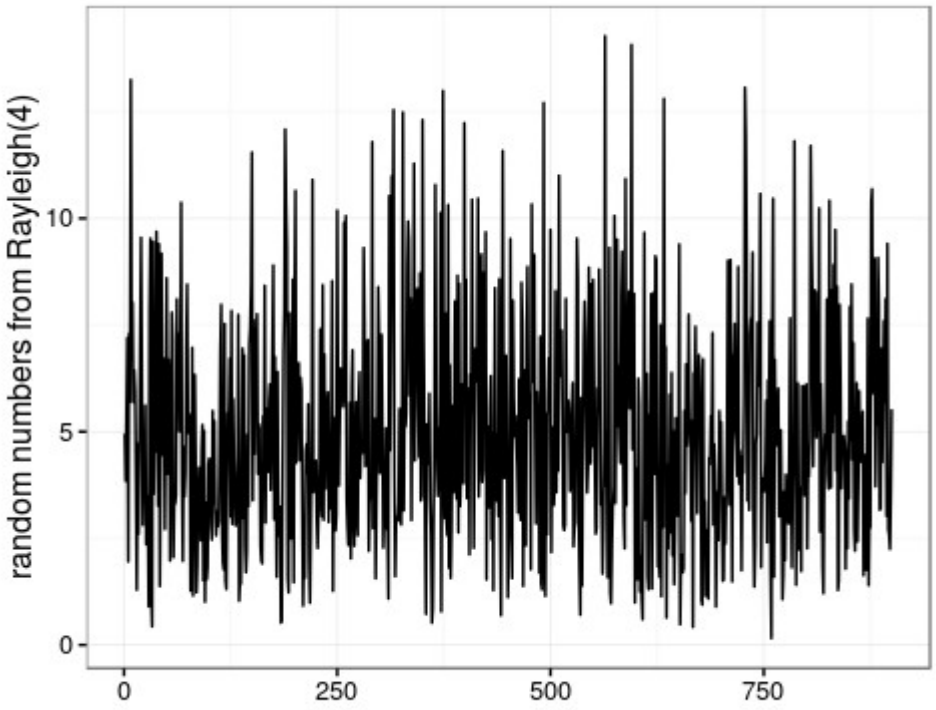

normal
Cauchy
a * Cauchy
densities
0.0
0.1
0.2
0.3
0.4
0.5
-4
-2
0
2
4
x

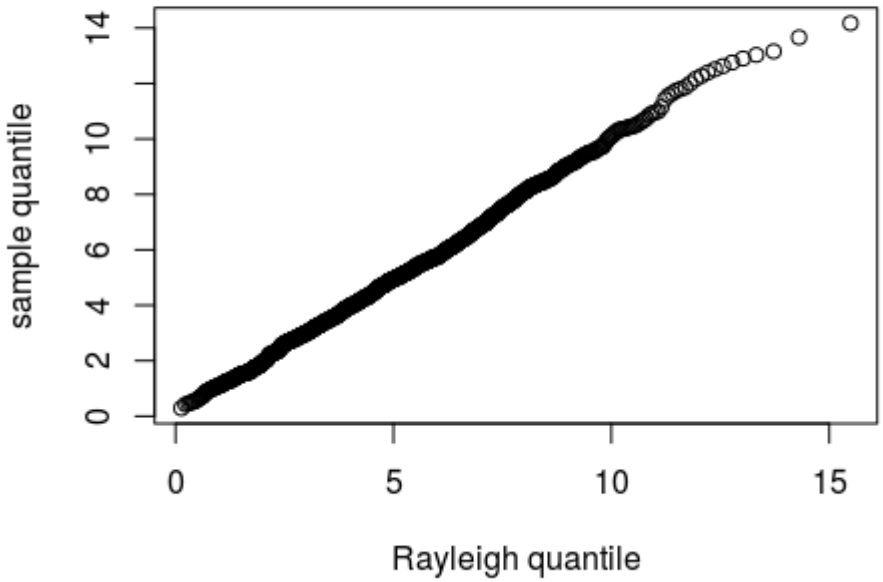

dn
0.0
0.1
0.2
0.3
0.4
0.5
-4
-2
0
2
4
x

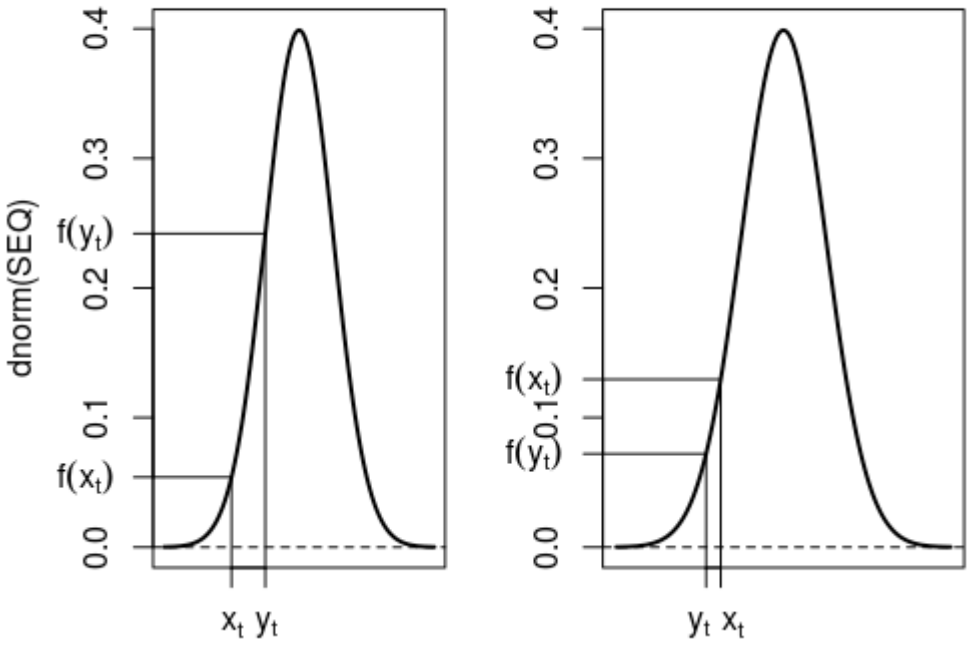

Histogram of x
Density
0.0
0.1
0.2
0.3
0.4
-2
0
2
4
x

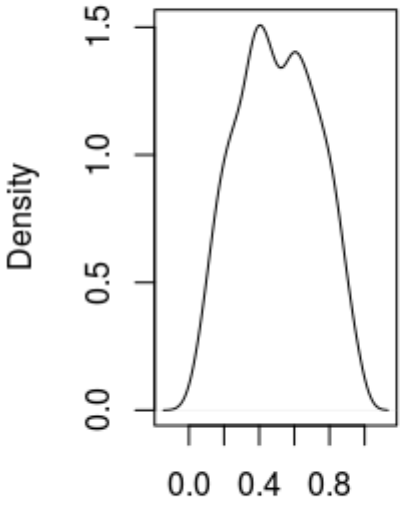

0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0

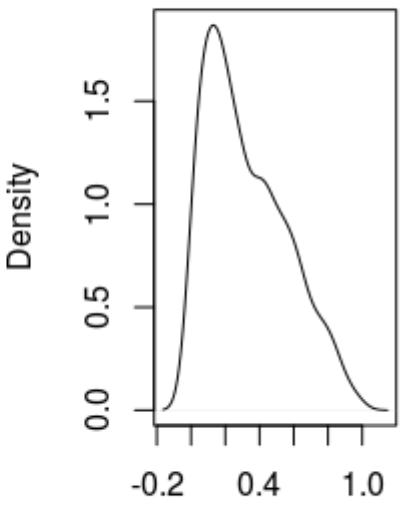

true density
density of simulated values
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0

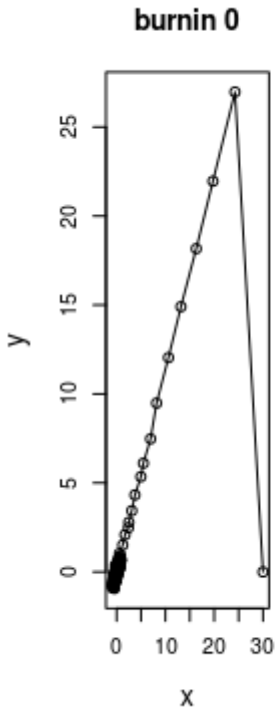

random numbers from Rayleigh(4)
0
5
10
0
250
500
750

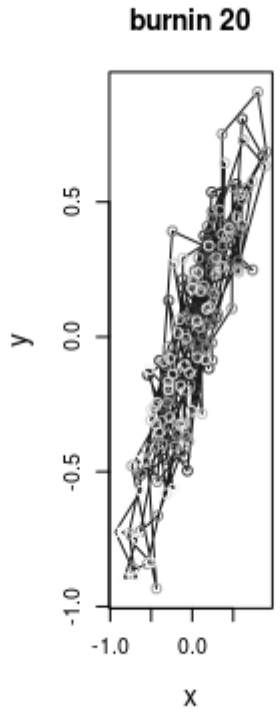

sample quantile
0
2
4
6
8
10
14
0
5
10
15
Rayleigh quantile

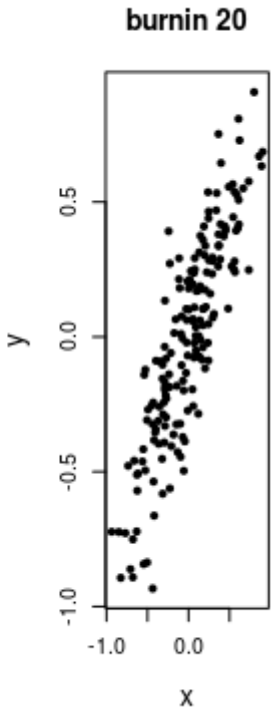

dnorm(SEQ)
0.0
0.1
0.2
0.3
0.4
f(y_t)
f(x_t)
x_t y_t
f(x_t)
f(y_t)
y_t x_t

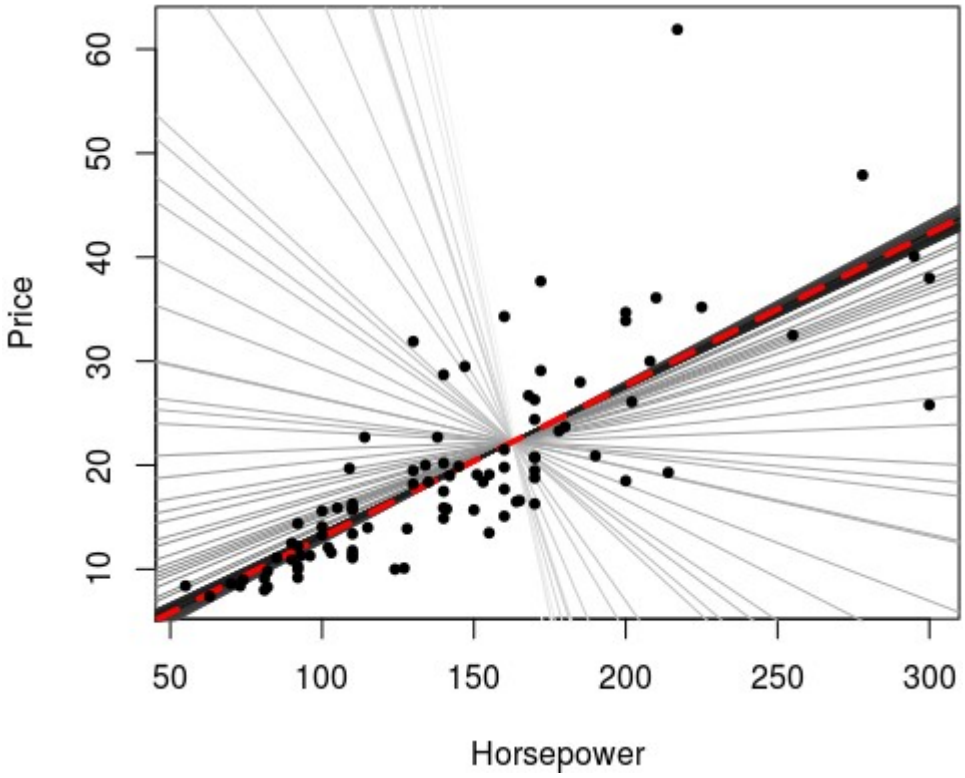

Density
0.0
0.5
1.0
1.5
0.0
0.4
0.8

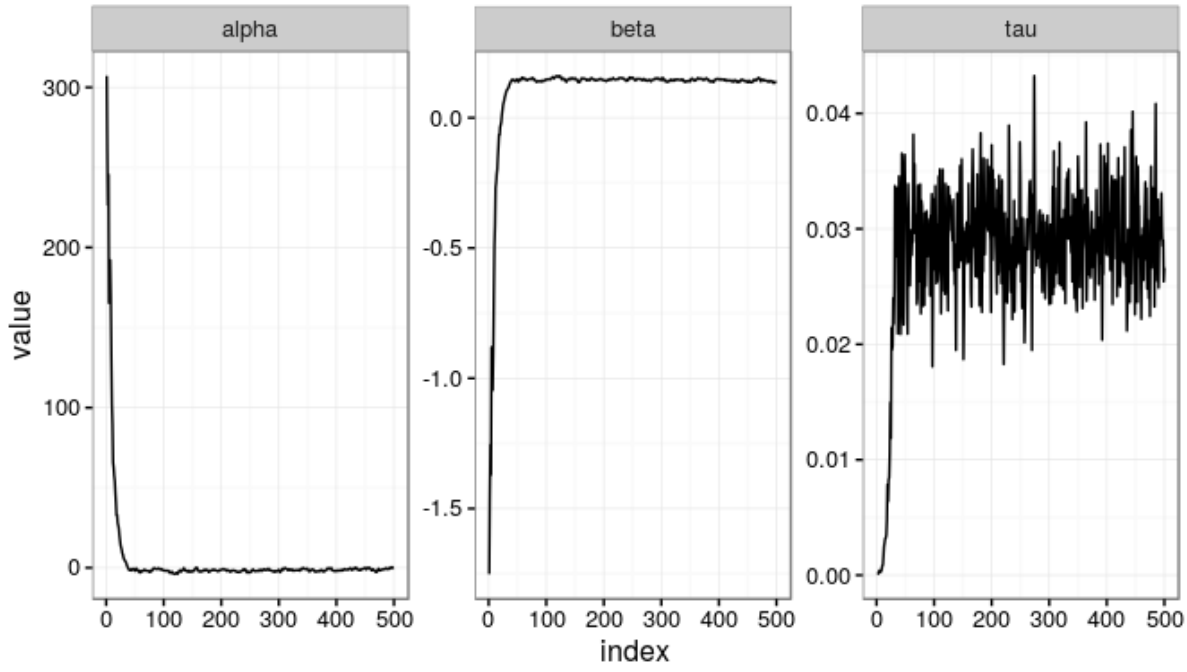

Density
0.0
0.5
1.0
1.5
-0.2
0.4
1.0

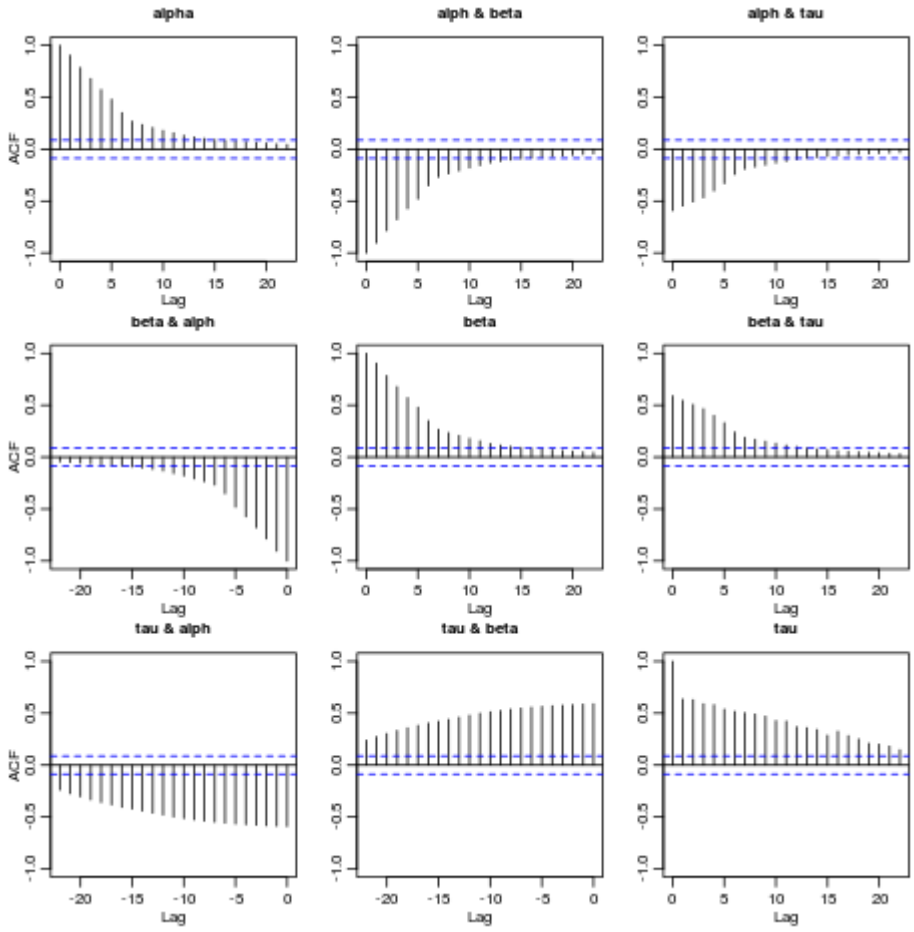

burnin 0
y
0
5
10
15
20
25
0
10
20
30
x

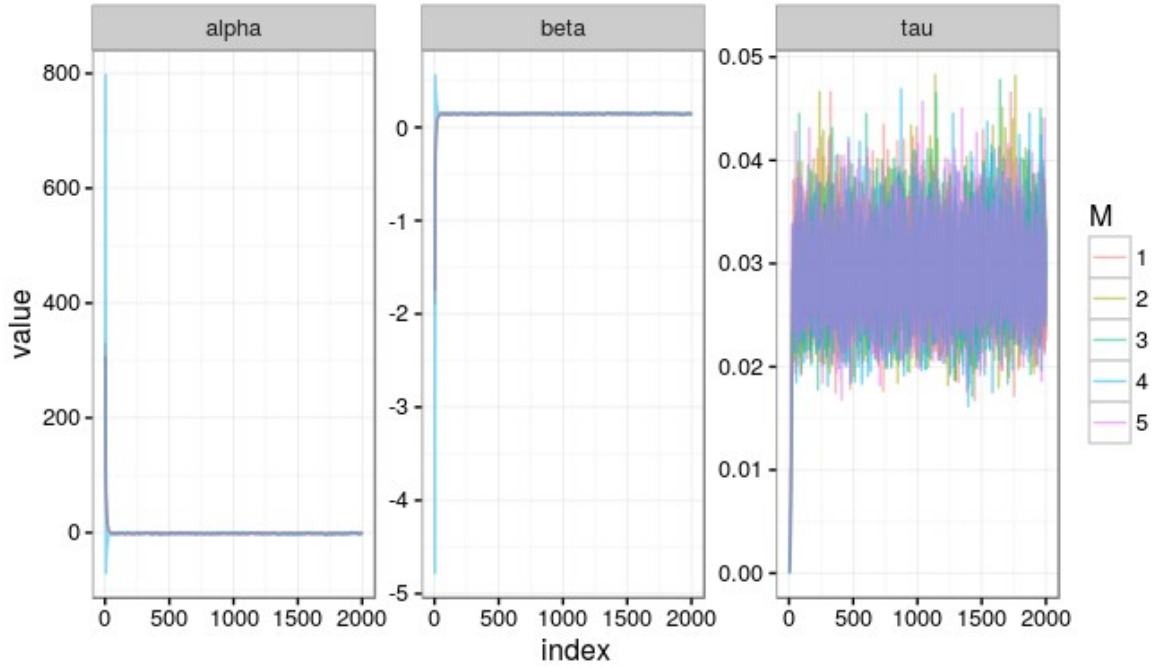

burnin 20
y
-1.0
-0.5
0.0
0.5
-1.0
0.0
x

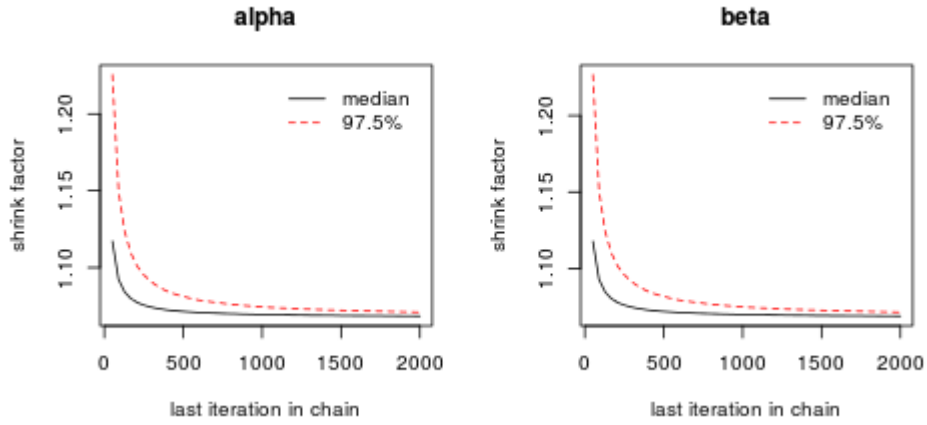

burnin 20
y
-1.0
-0.5
0.0
0.5
-1.0
0.0
x


Price
10
20
30
40
50
60
50
100
150
200
250
300
Horsepower

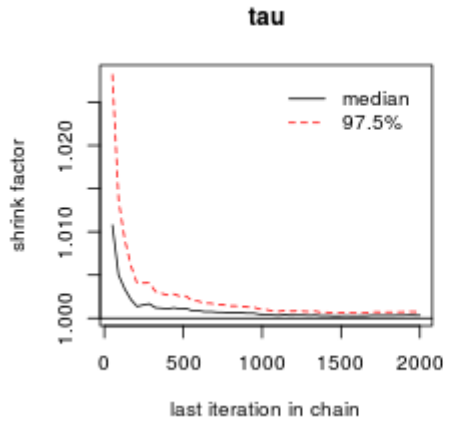

alpha
beta
tau
value
300
200
100
0
0.0
-0.5
-1.0
-1.5
0.04
0.03
0.02
0.01
0.00
0
100
200
300
400
500
index

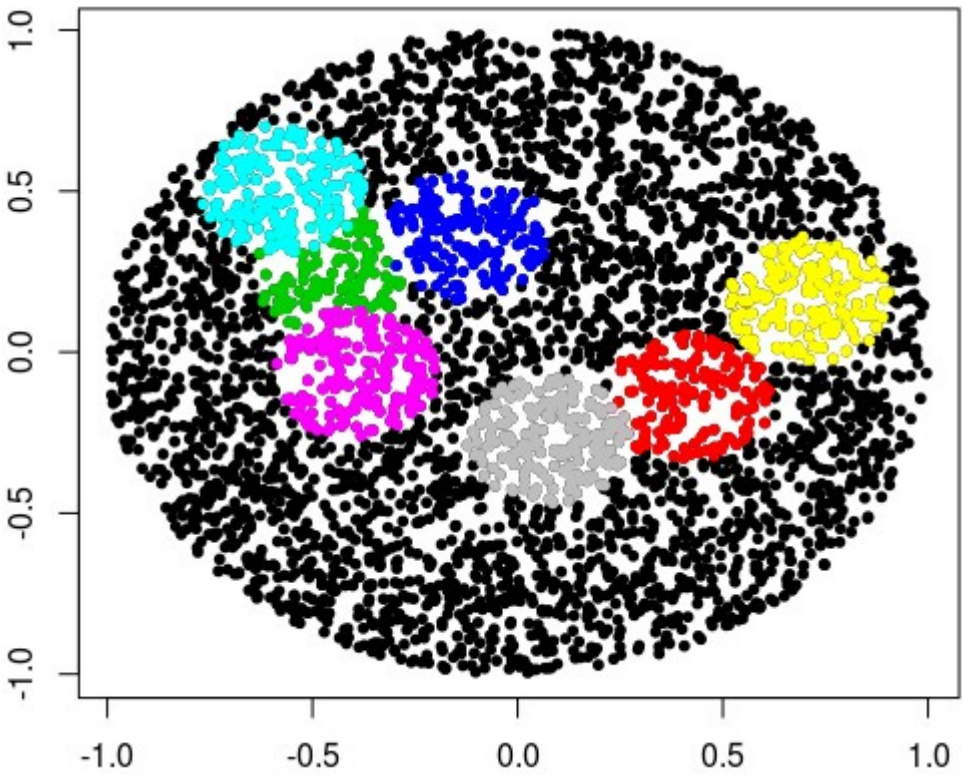

alpha
alph & beta
alph & tau
beta & alph
beta
beta & tau
tau & alph
tau & beta
tau
ACF
Lag

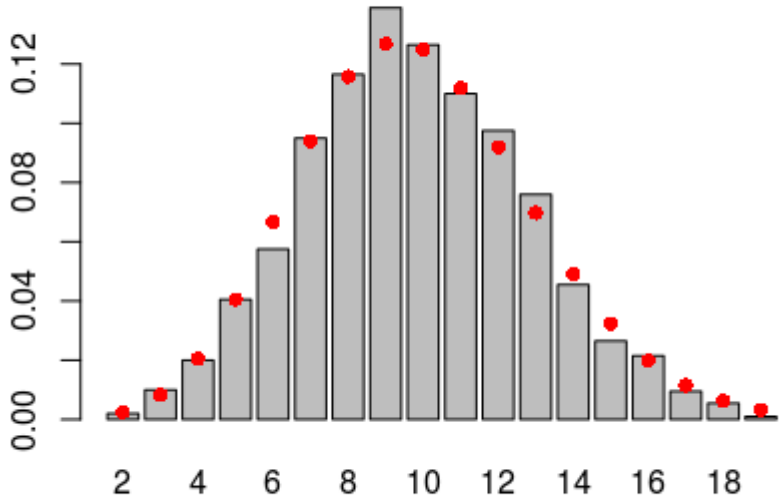

alpha
beta
tau
value
index
M
1
2
3
4
5


alpha
median
97.5%
shrink factor
1.10
1.15
1.20
0
500
1000
1500
2000
last iteration in chain

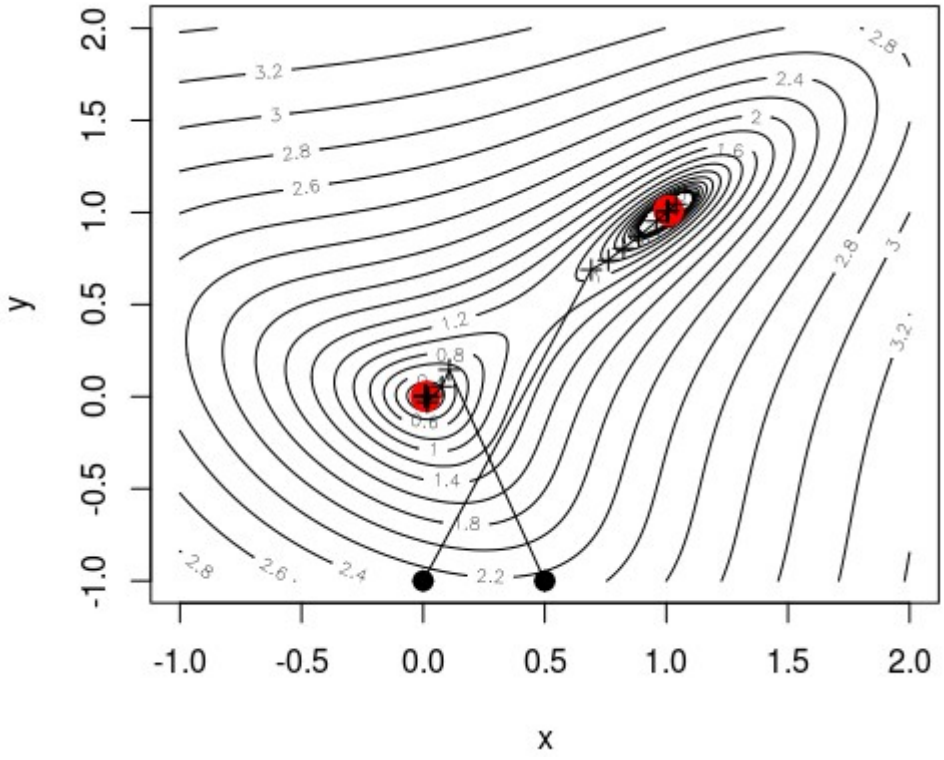
beta
median
97.5%
shrink factor
1.10
1.15
1.20
0
500
1000
1500
2000
last iteration in chain

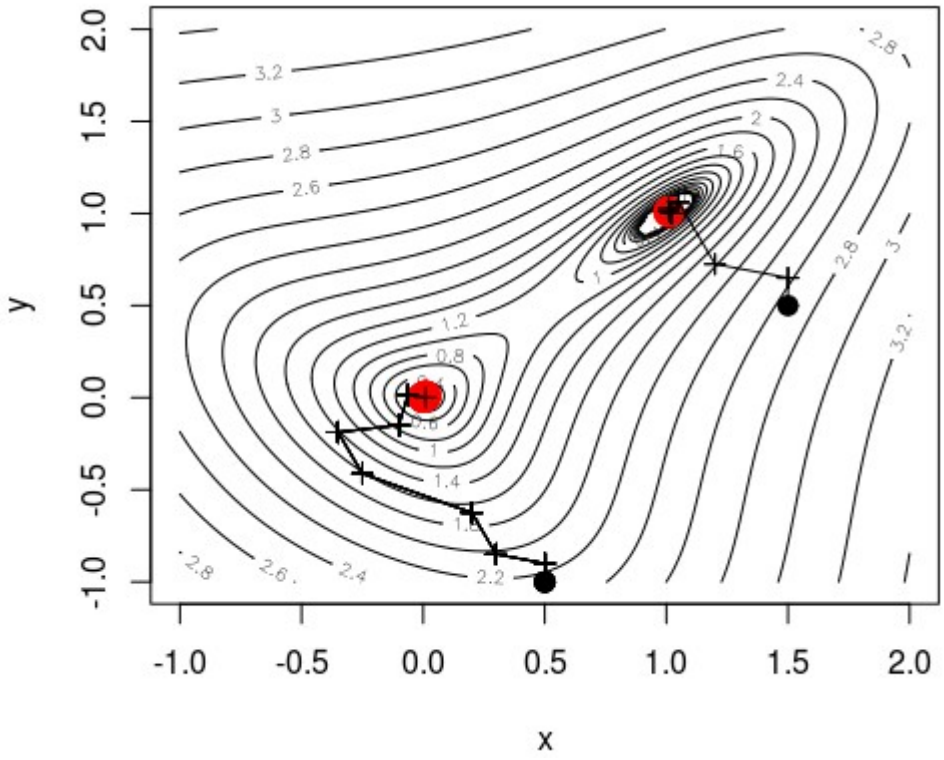

tau
median
97.5%
shrink factor
1.000
1.010
1.020
0
500
1000
1500
2000
last iteration in chain

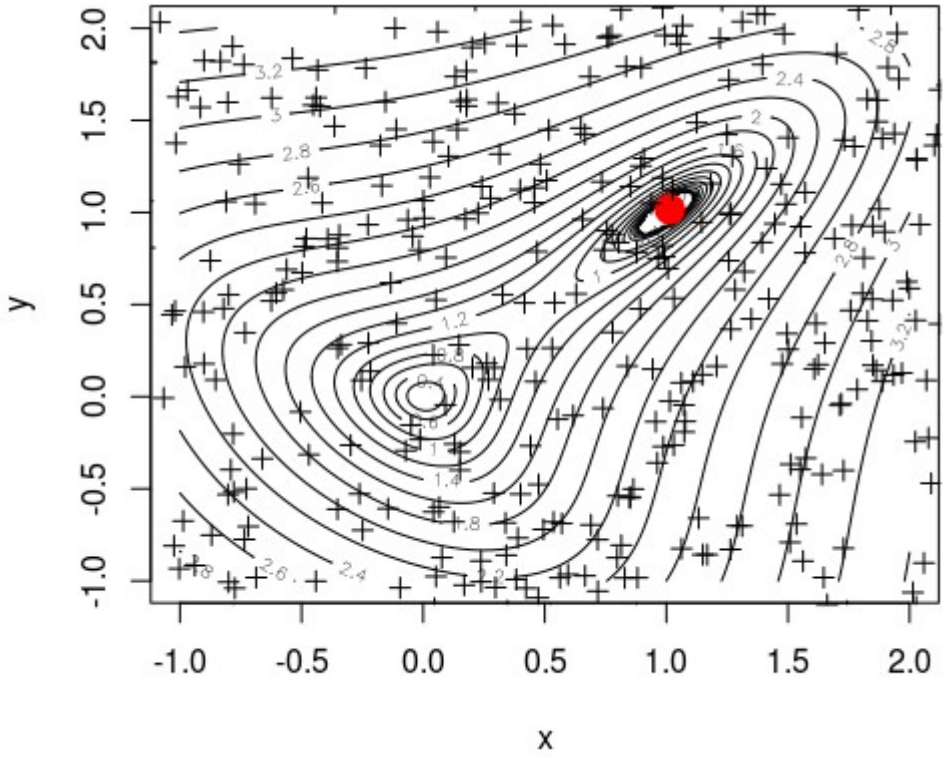

1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0

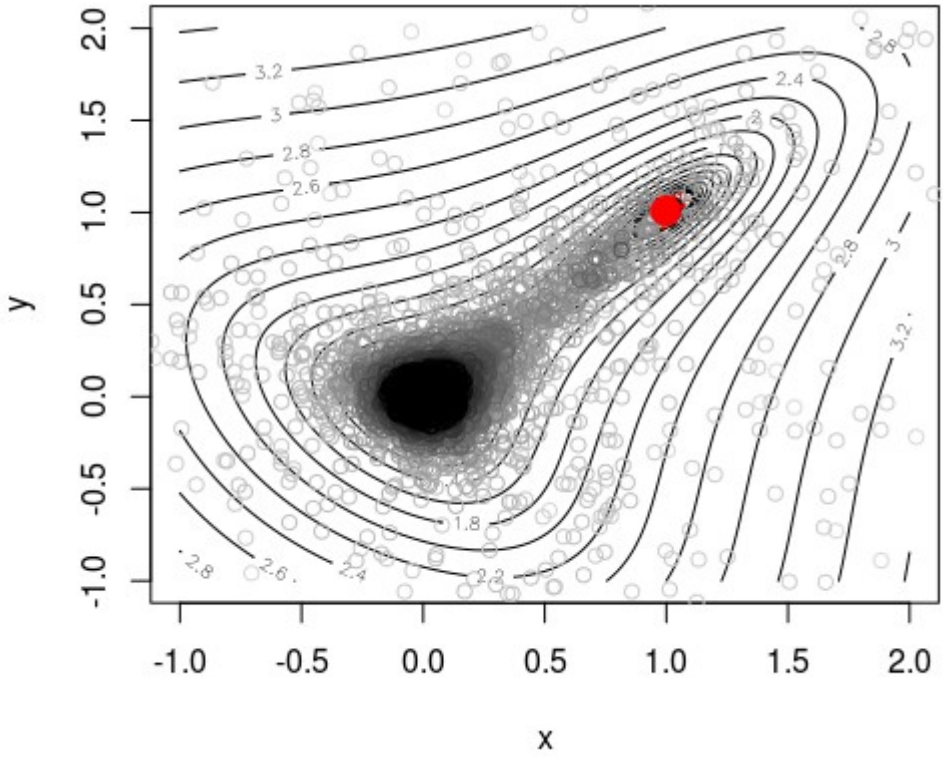


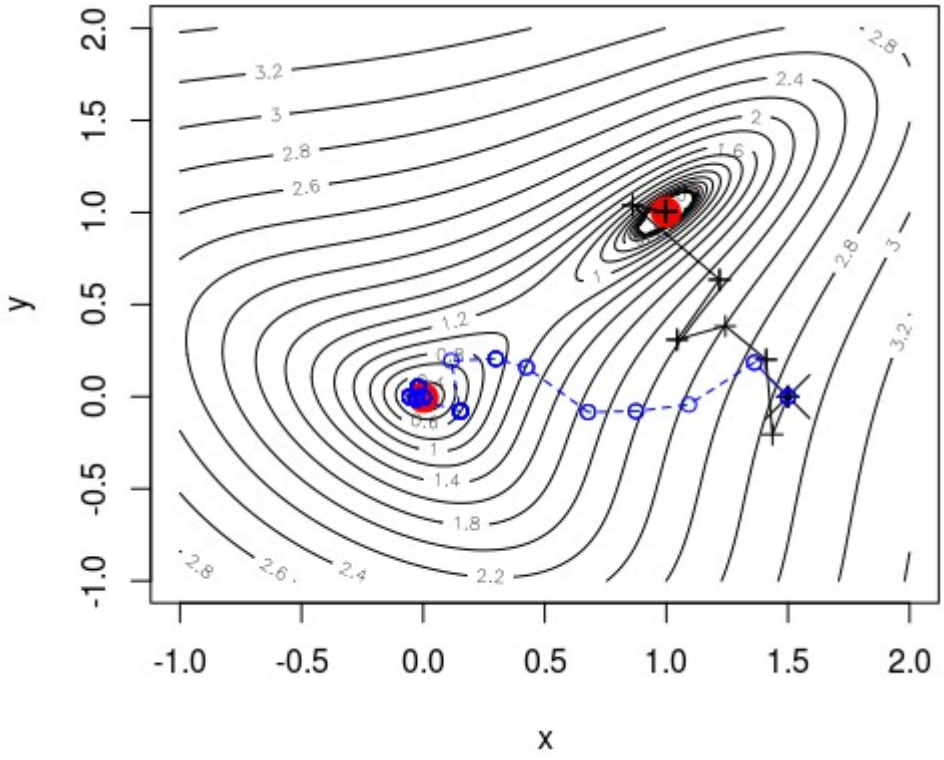
Chapter 5: Monte Carlo Methods for Optimization Problems

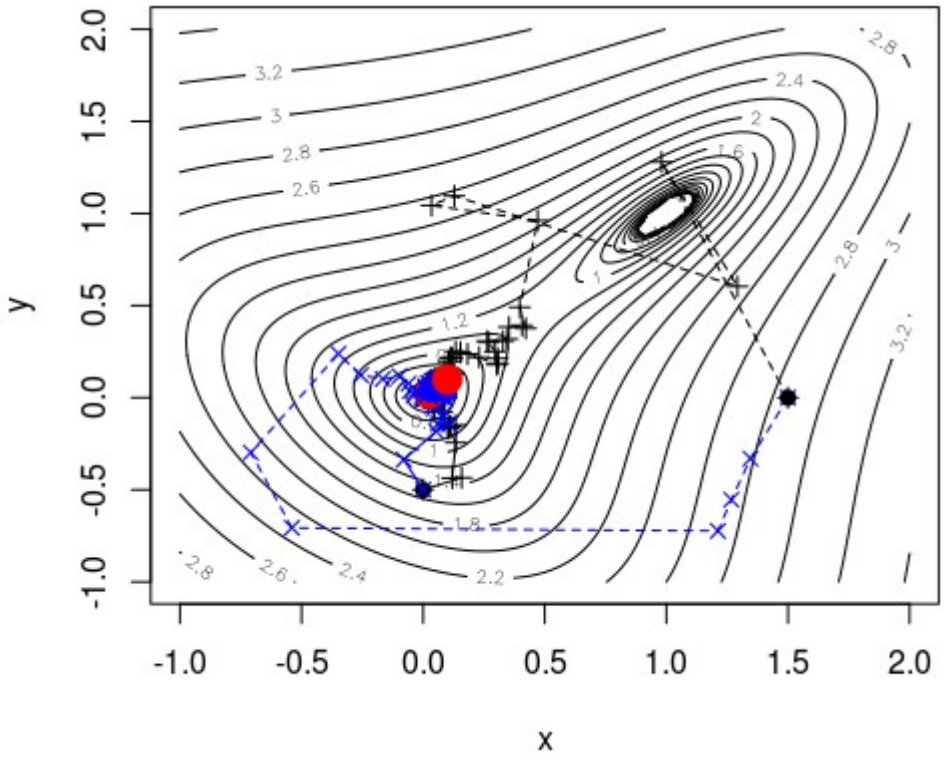

y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
3.2
3
2.8
2.6
2.4
2
1.6
1.2
0.8
1
1.4
1.8
2.2

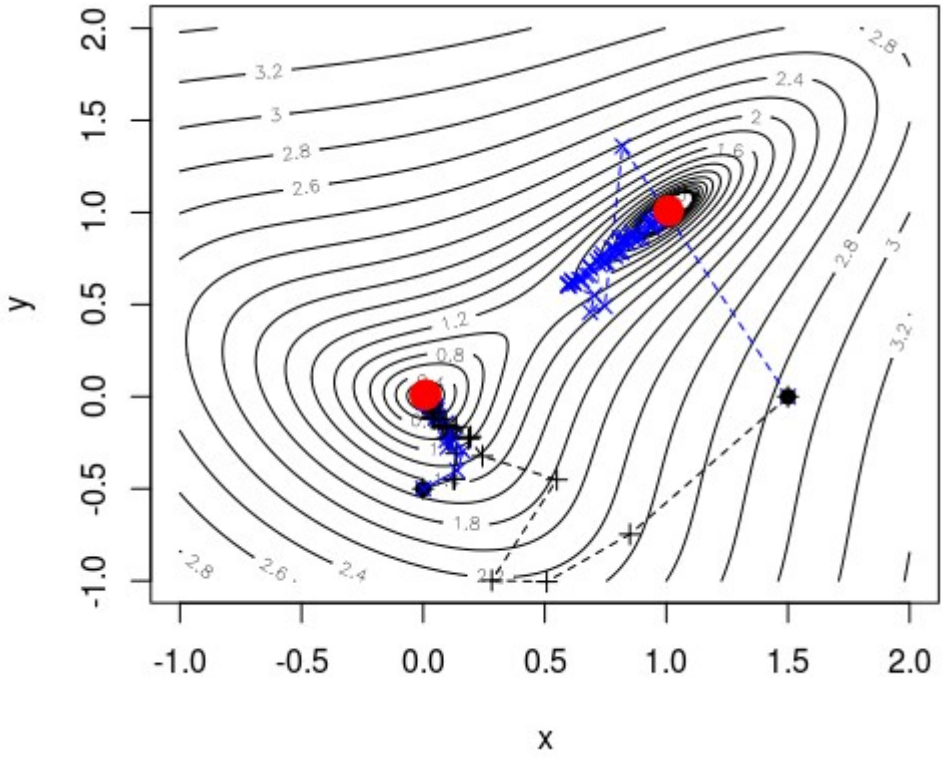

y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
3.2
3
2.8
2.6
2.4
2
1.6
1.2
0.8
1
1.4
2.2

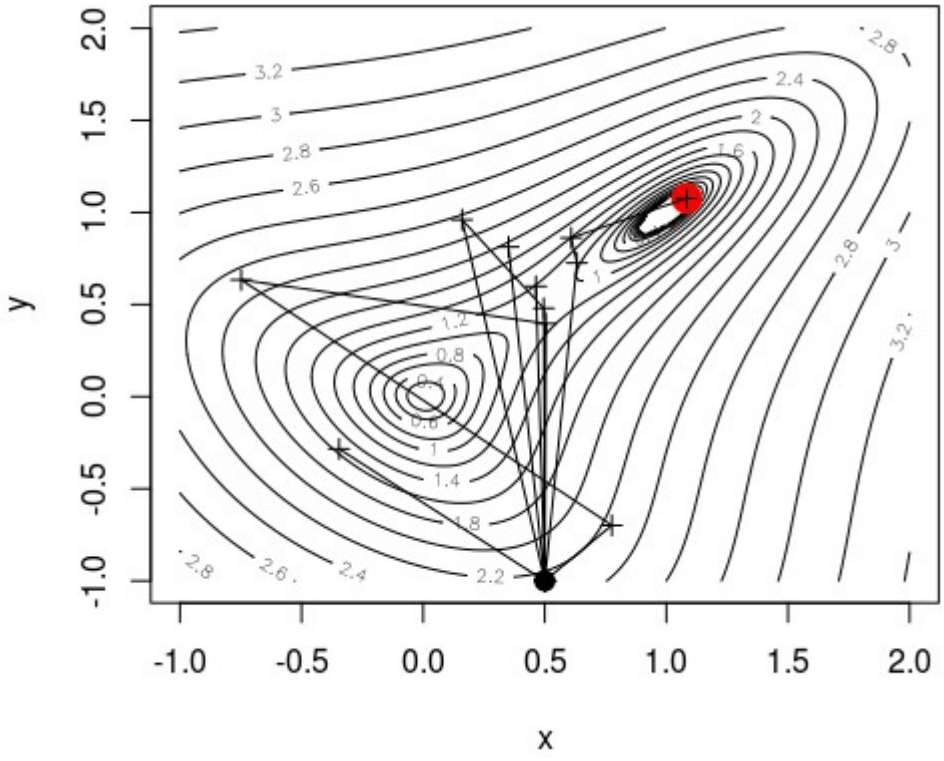

y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0


y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0

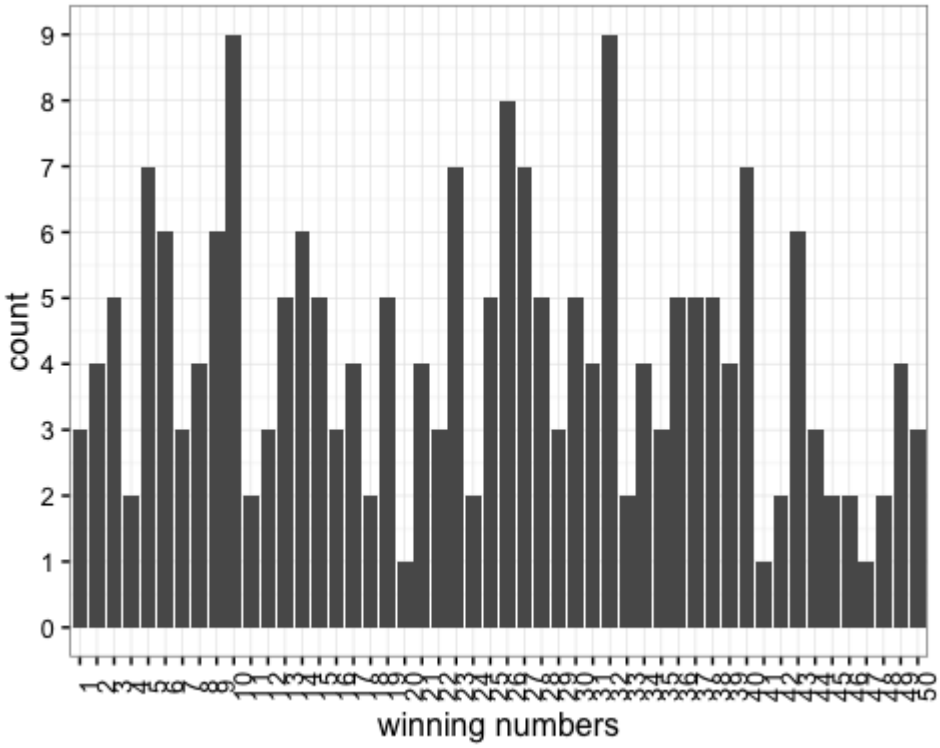

y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
3.2
3
2.8
2.6
2.4
2
1.6
1.2
1
1.4
1.8
2.2
3.2

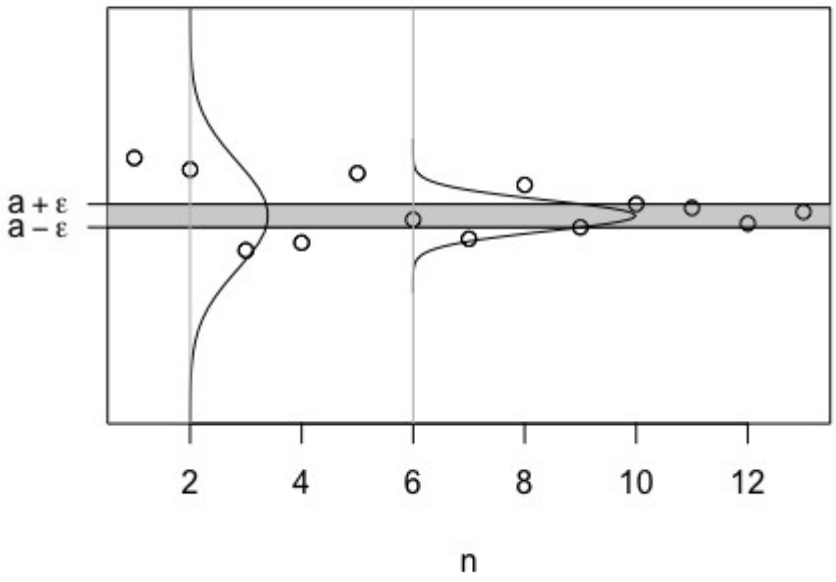

y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
3.2
3
2.8
2.6
2.4
2
1.6
1.2
1
1.8
2.2
3.2

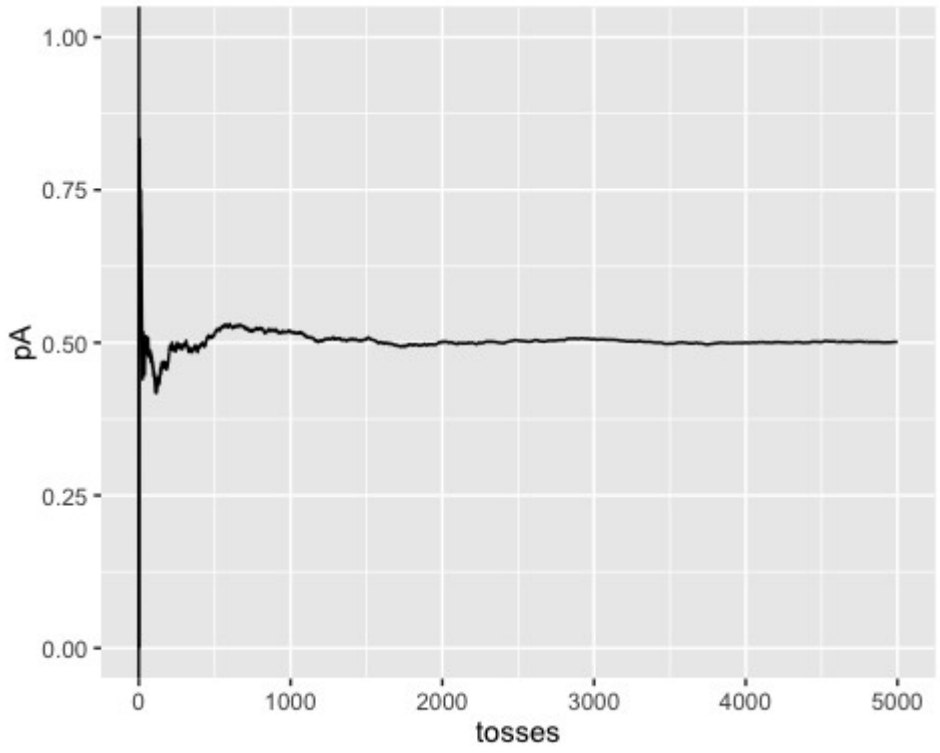

y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
3.2
3
2.8
2.6
2.4
2
1.6
1.2
0.8
1.8

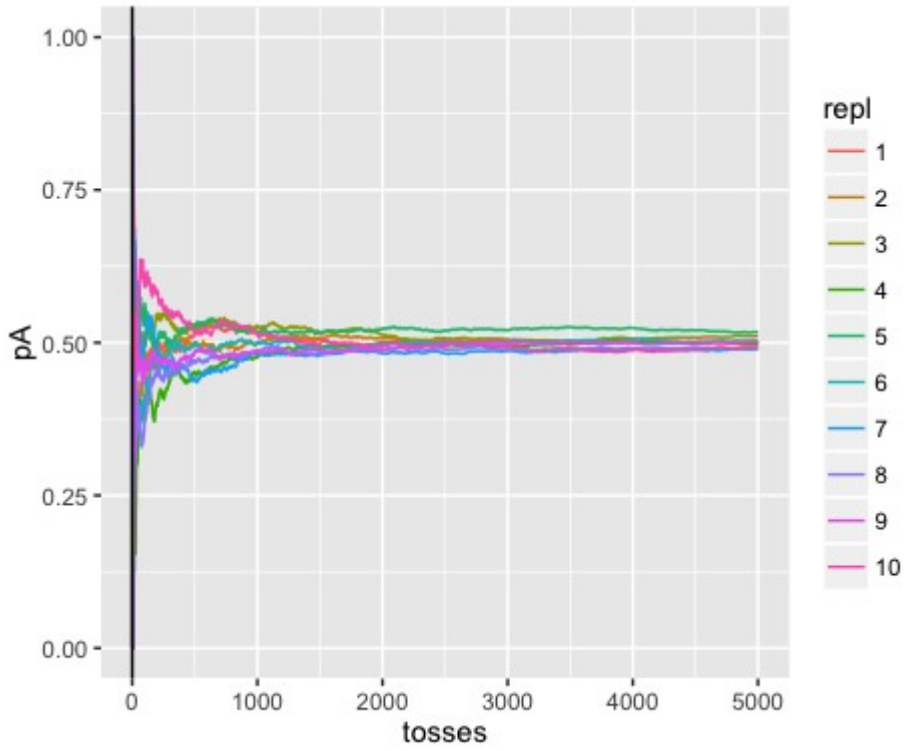

y
x
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
3.2
3
2.8
2.6
2.4
2
1.6
1.2
0.8
1.4
1.8
2.2

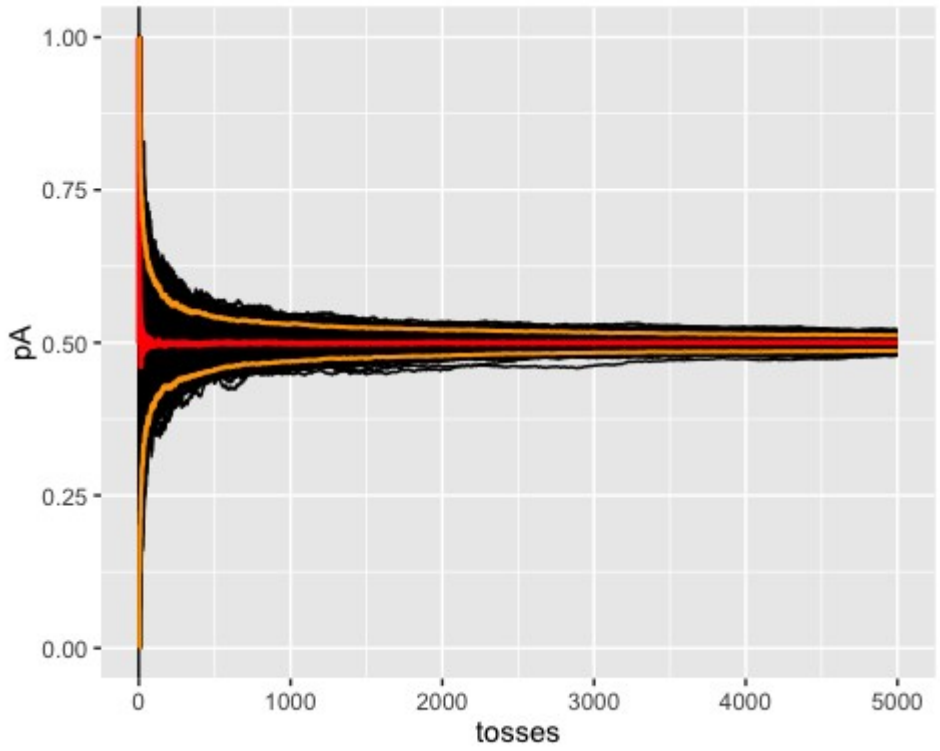
Chapter 6: Probability Theory Shown by Simulation

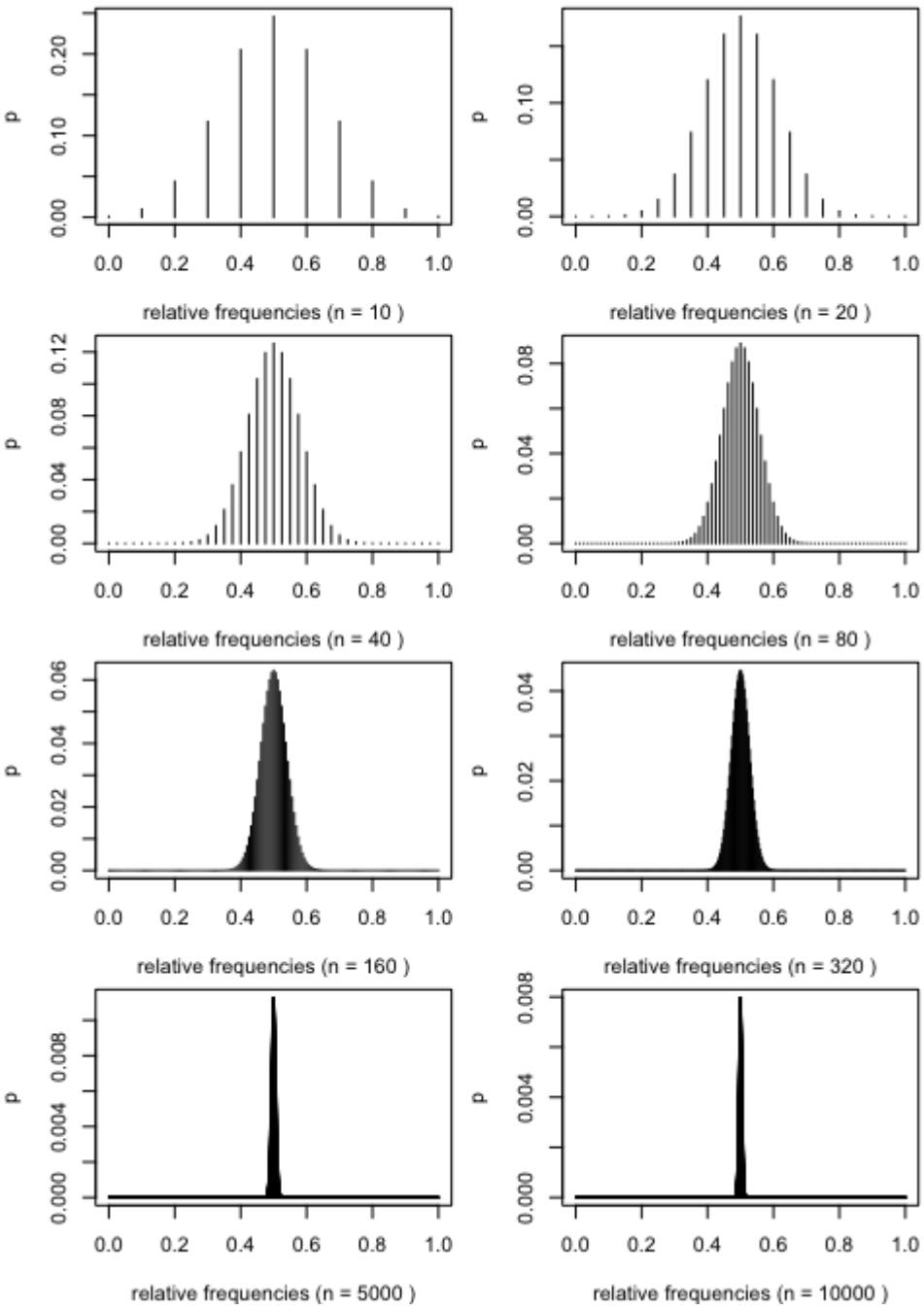


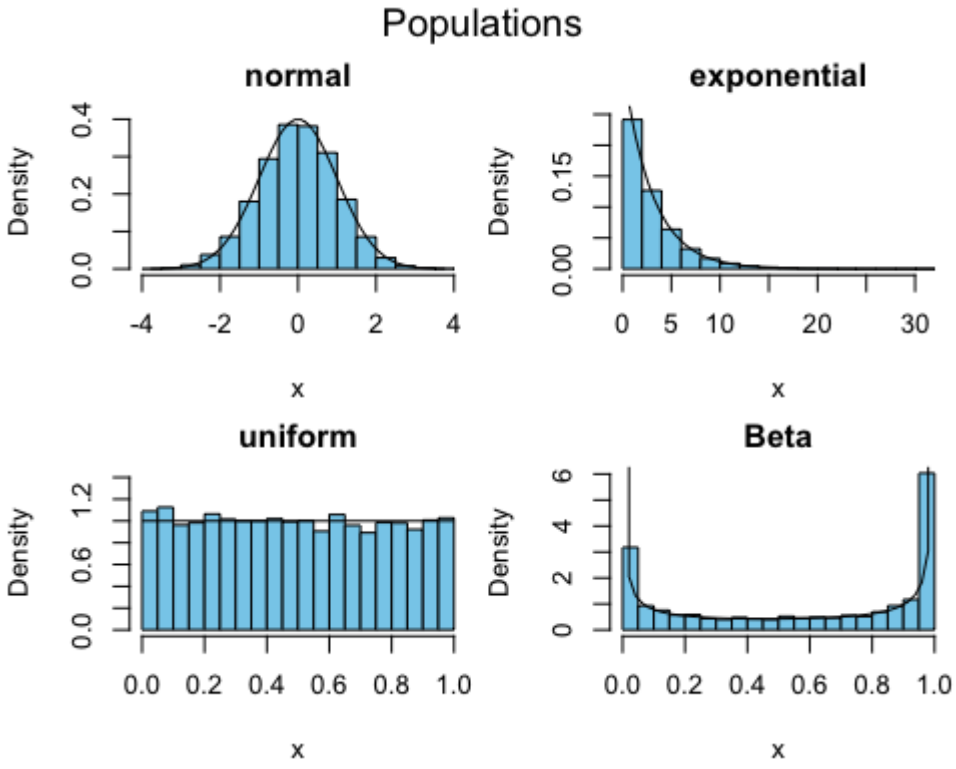

a + ε
a − ε
2
4
6
8
10
12
n

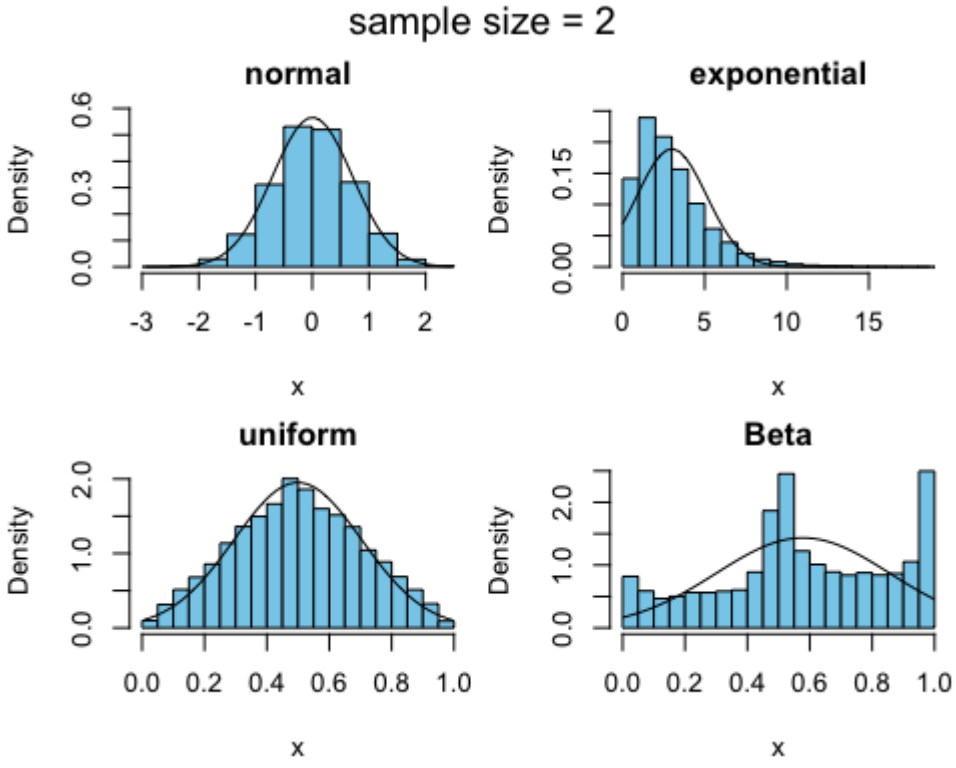

1.00
0.75
0.50
0.25
0.00
pA
0
1000
2000
3000
4000
5000
tosses

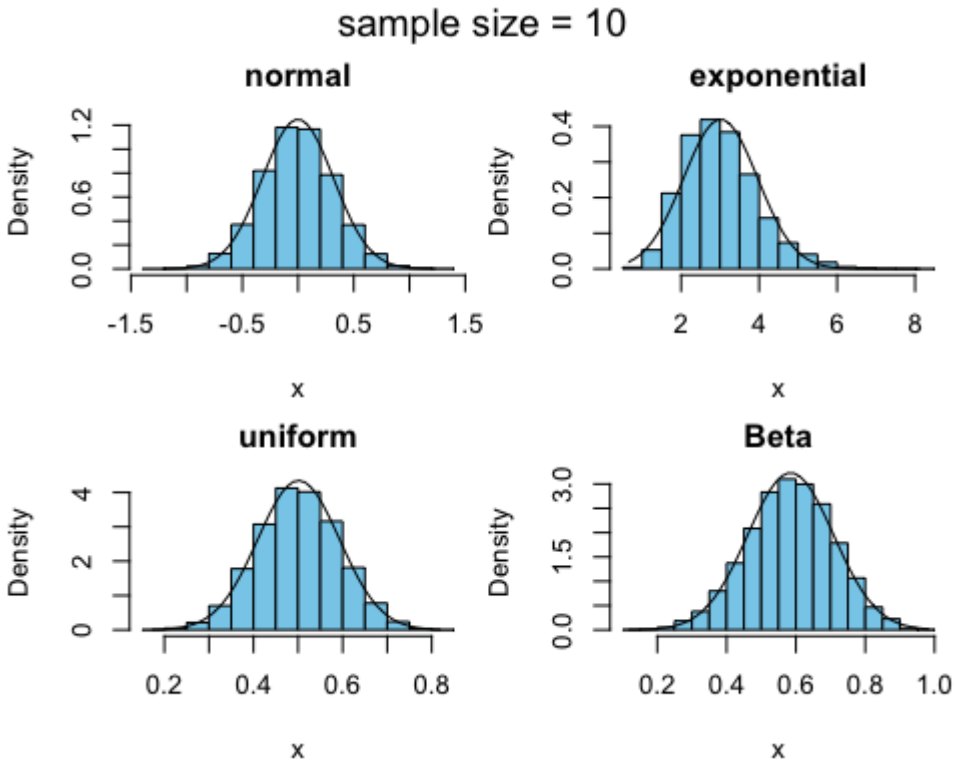

1.00
0.75
0.50
0.25
0.00
pA
0
1000
2000
3000
4000
5000
tosses
repl
1
2
3
4
5
6
7
8
9
10


1.00
0.75
0.50
0.25
0.00
pA
0
1000
2000
3000
4000
5000
tosses

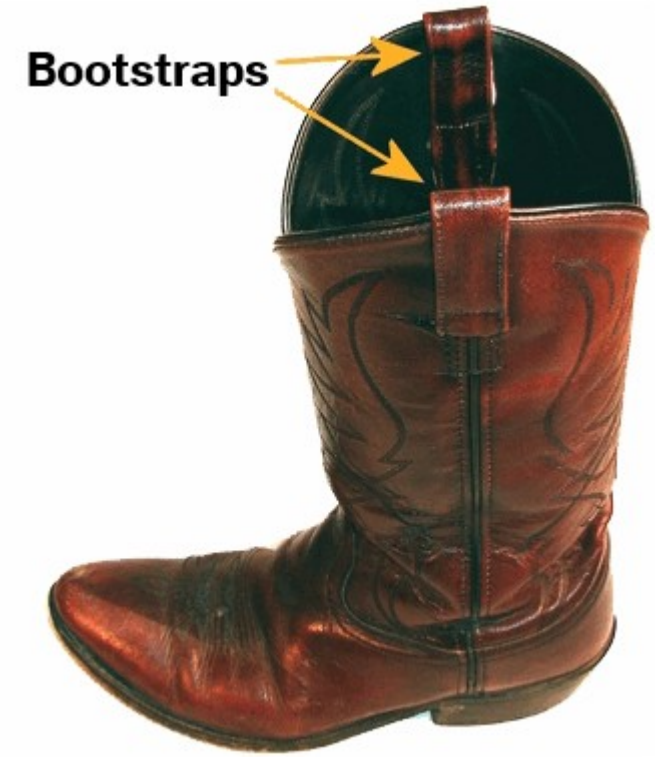

p
0.00
0.10
0.20
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 10)
p
0.00
0.10
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 20)
p
0.00
0.04
0.08
0.12
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 40)
p
0.00
0.04
0.08
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 80)
p
0.00
0.02
0.04
0.06
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 160)
p
0.00
0.02
0.04
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 320)
p
0.000
0.004
0.008
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 5000)
p
0.000
0.004
0.008
0.0
0.2
0.4
0.6
0.8
1.0
relative frequencies (n = 10000)

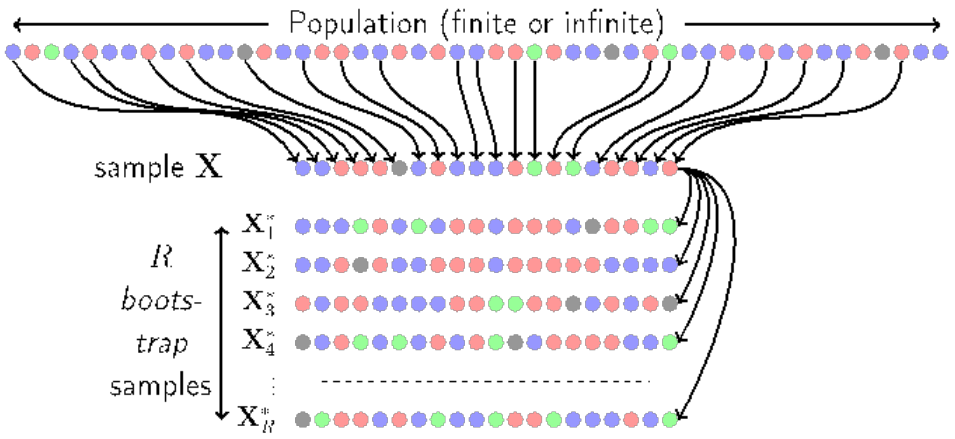

Populations
normal
exponential
uniform
Beta
Density
x

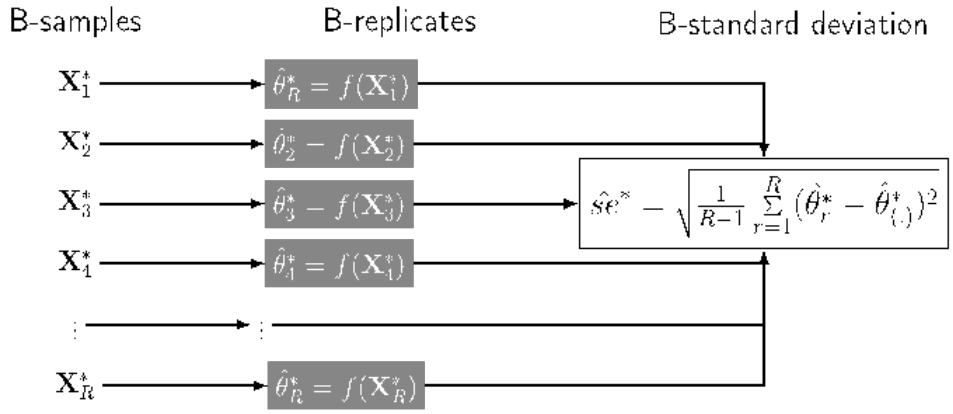

sample size = 2
normal
exponential
uniform
Beta
Density
x

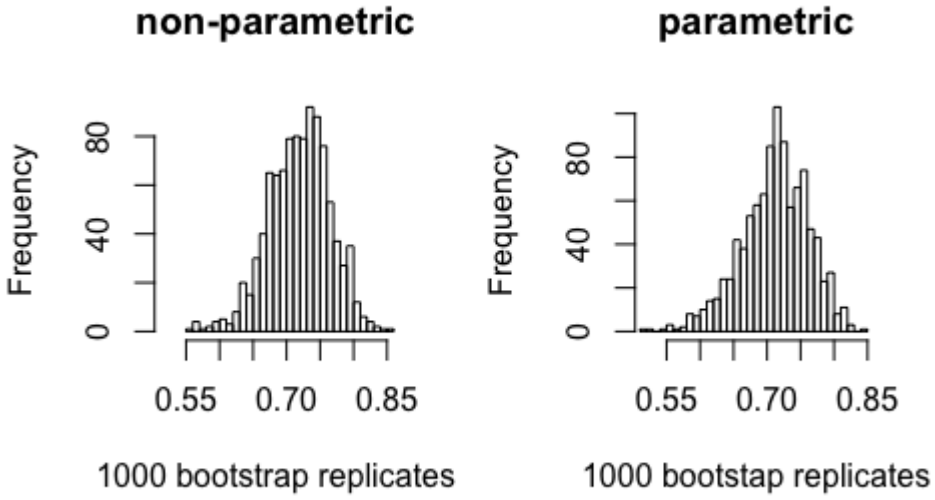

sample size = 10
normal
exponential
uniform
Beta
Density
x
0.0
0.6
1.2
-1.5
-0.5
0.5
1.5
0.2
0.4
2
4
6
8
0
2
4
0.6
0.8
1.5
3.0
1.0

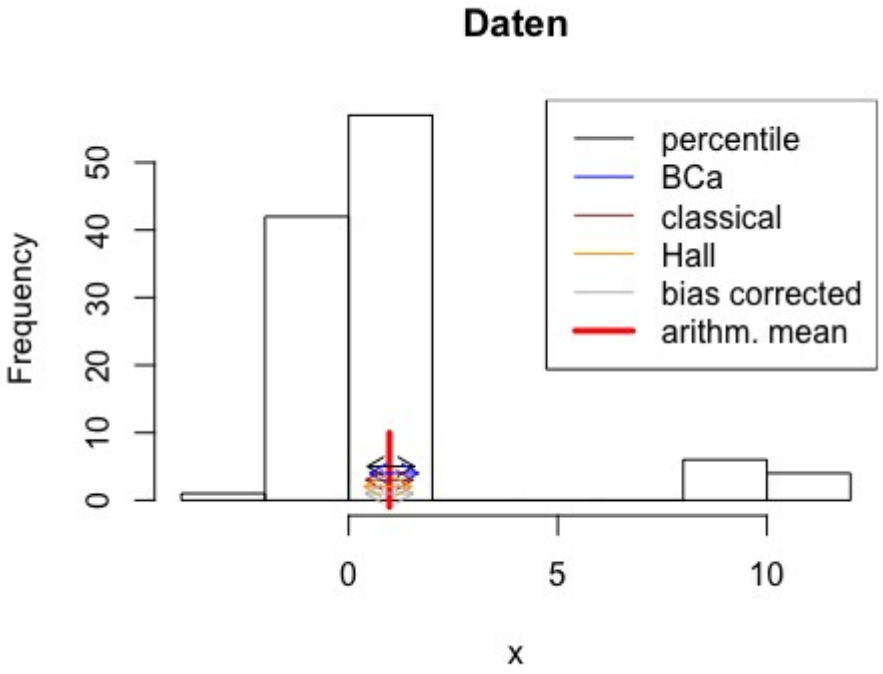
Chapter 7: Resampling Methods

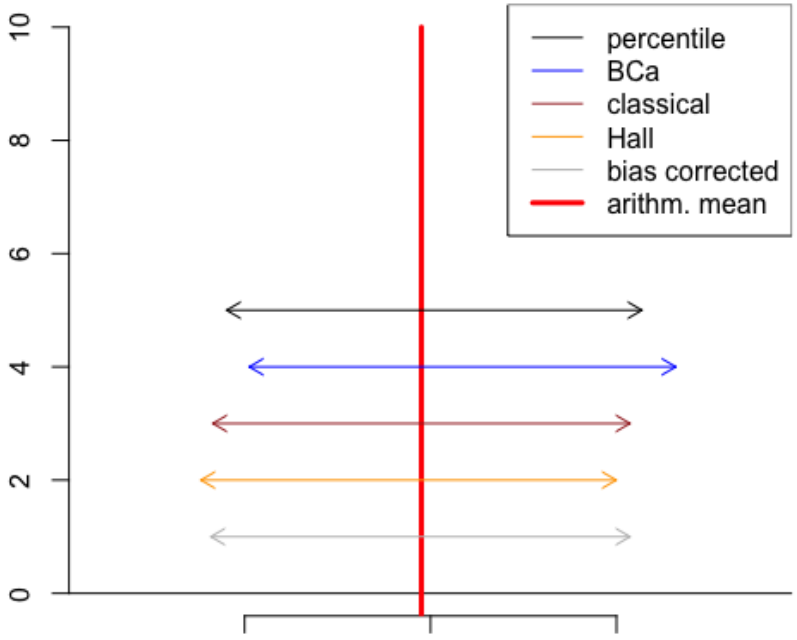


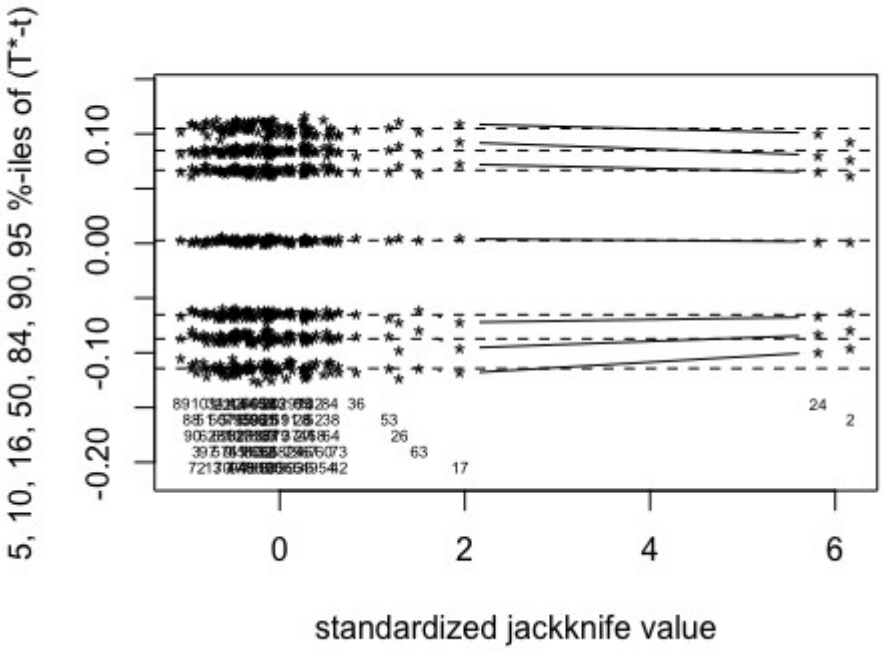

Population (finite or infinite)
sample X
R boots-trap samples

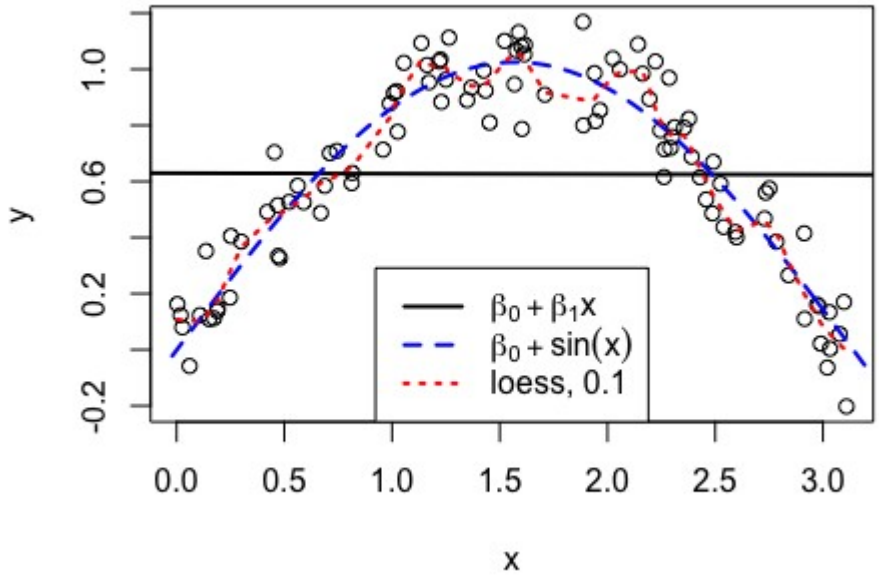

B-samples
B-replicates
B-standard deviation

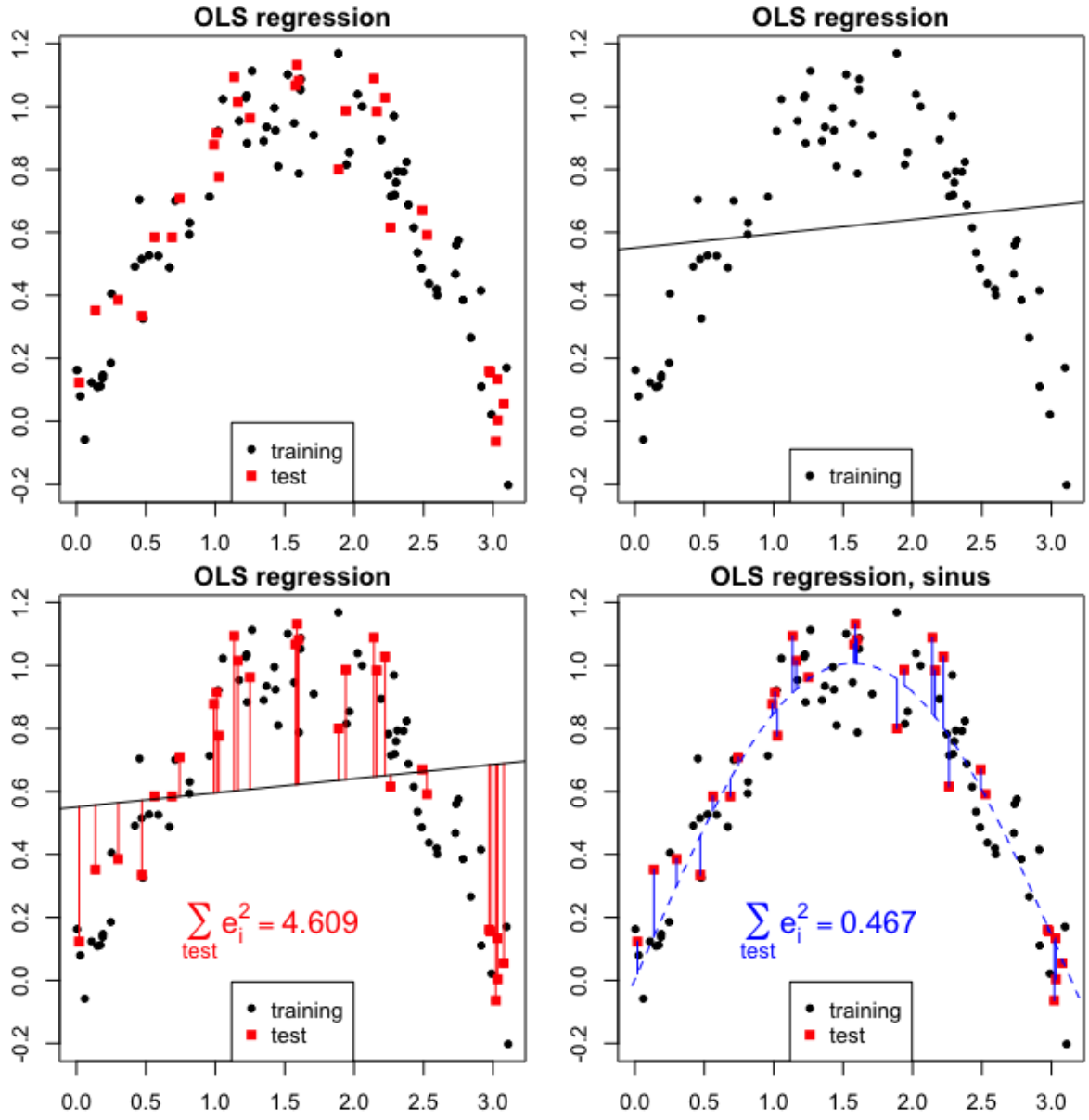

non-parametric
parametric
Frequency
0
40
80
0.55
0.70
0.85
1000 bootstrap replicates
1000 bootstap replicates

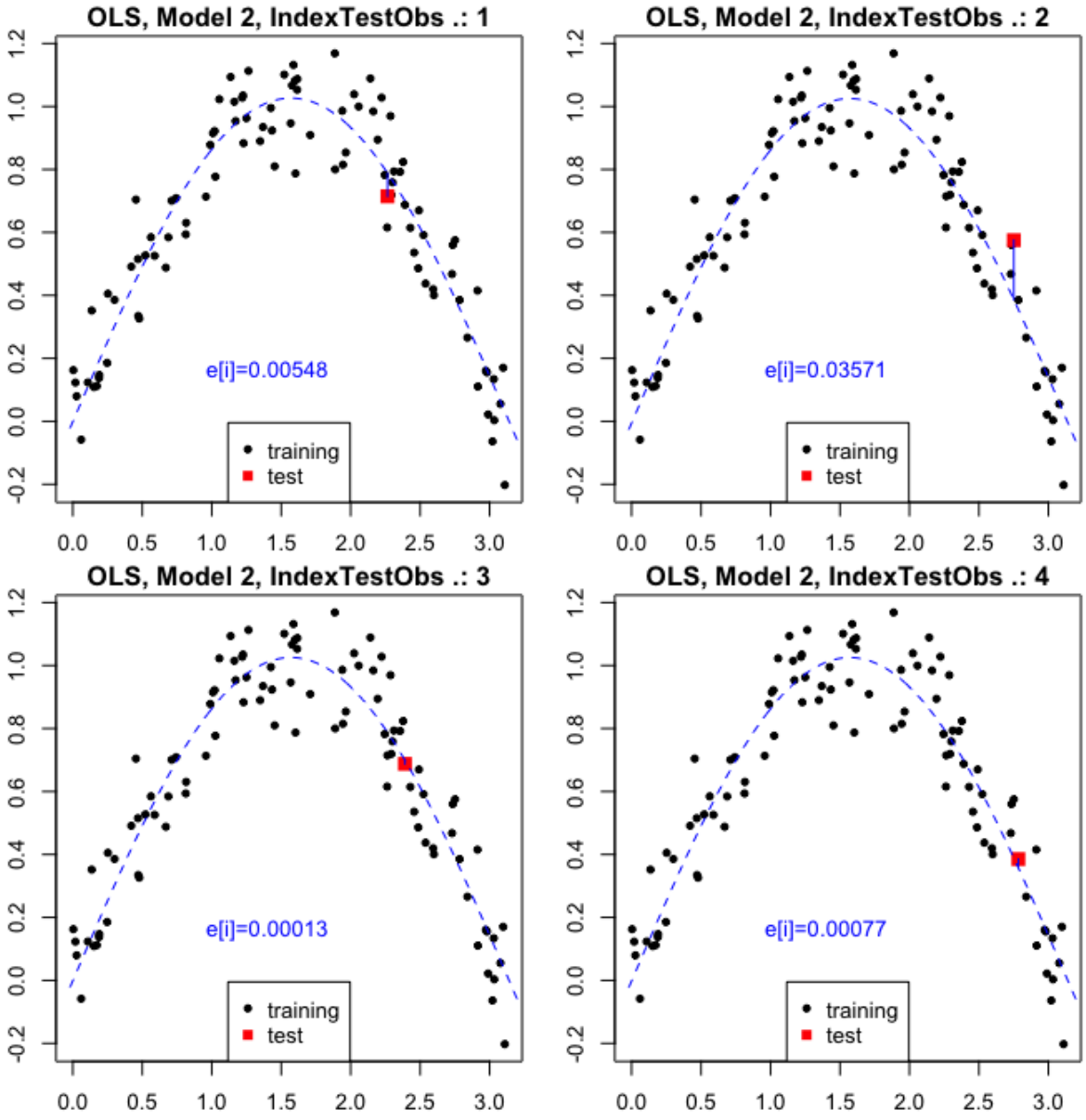

Daten
Frequency
0
10
20
30
40
50
0
5
10
x
percentile
BCa
classical
Hall
bias corrected
arithm. mean

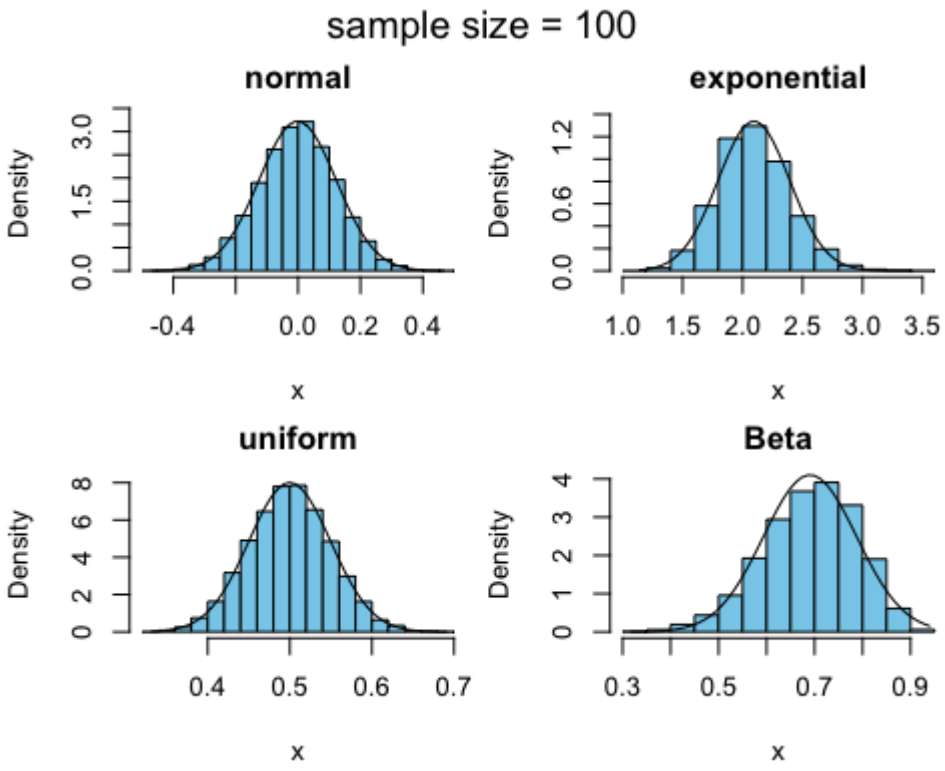

0
2
4
6
8
10
percentile
BCa
classical
Hall
bias corrected
arithm. mean

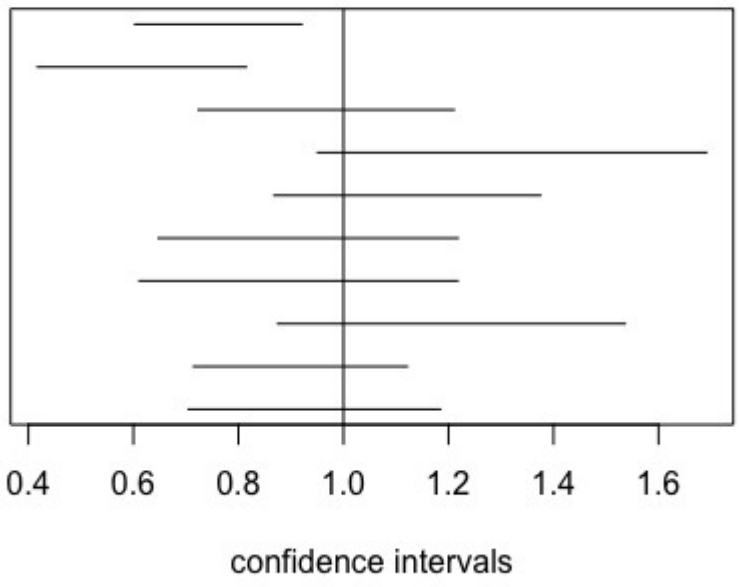

5, 10, 16, 50, 84, 90, 95 %-iles of (T*-t)
0.10
0.00
-0.10
-0.20
0
2
4
6
standardized jackknife value

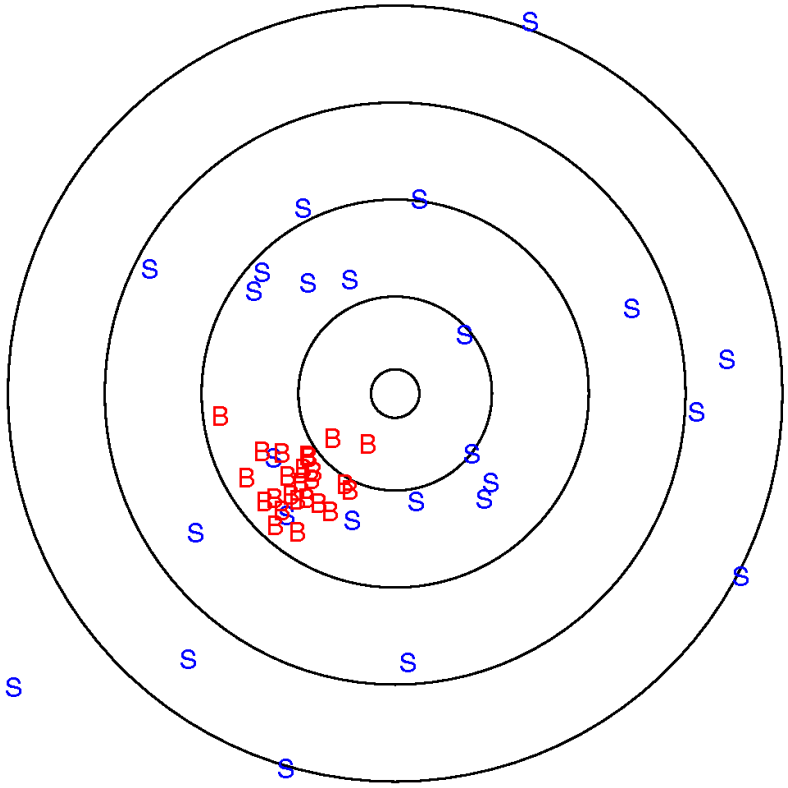

y
1.0
0.6
0.2
-0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x
$\beta_0 + \beta_1 x$
$\beta_0 + \sin(x)$
loess, 0.1


OLS regression
training
test
OLS regression
training
OLS regression
$\sum_{test} e_i^2 = 4.609$
training
test
OLS regression, sinus
$\sum_{test} e_i^2 = 0.467$
training
test

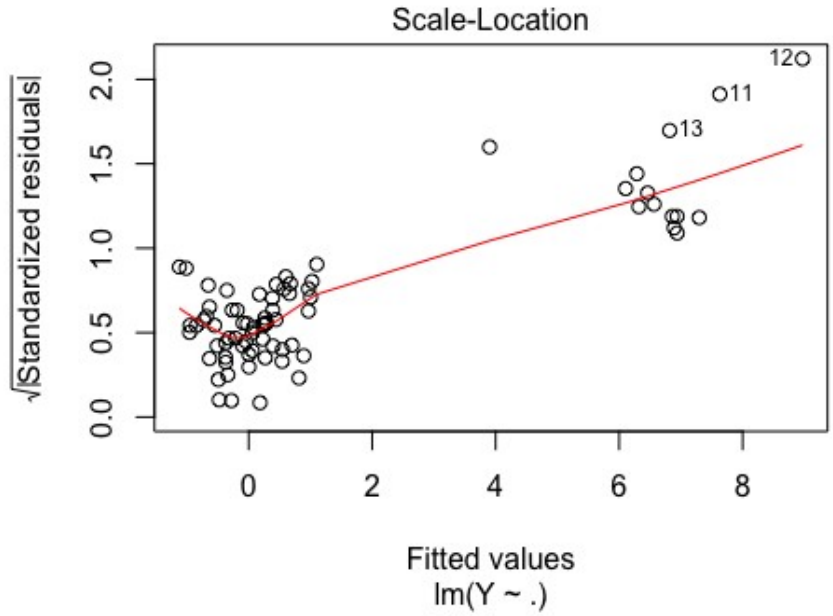

OLS, Model 2, IndexTestObs .: 1
e[i]=0.00548
training
test
OLS, Model 2, IndexTestObs .: 2
e[i]=0.03571
training
test
OLS, Model 2, IndexTestObs .: 3
e[i]=0.00013
training
test
OLS, Model 2, IndexTestObs .: 4
e[i]=0.00077
training
test

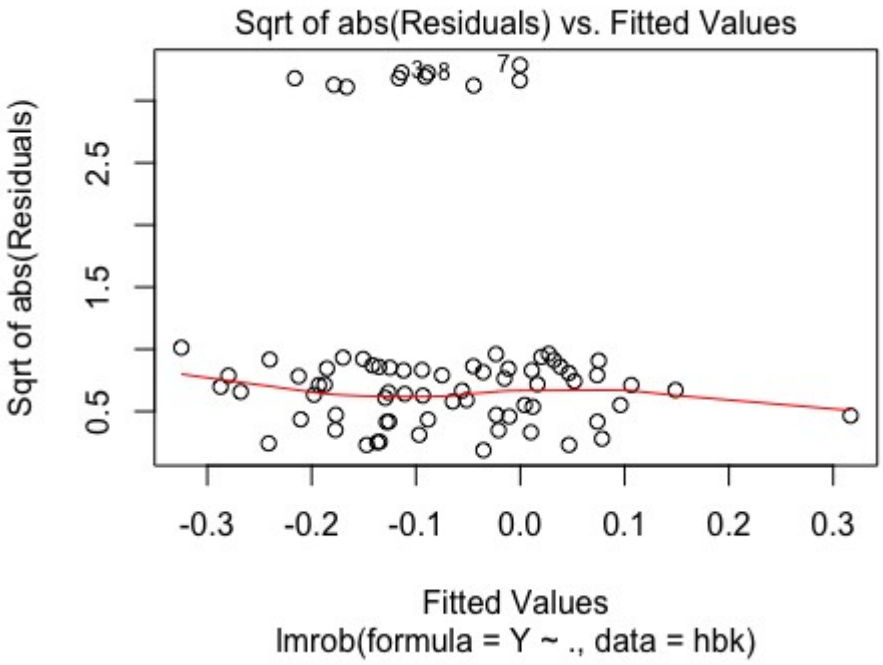

sample size = 100
normal
exponential
uniform
Beta
Density
x
0.0
1.5
3.0
0.6
1.2
0
2
4
6
8
1
3
-0.4
0.2
0.4
1.0
2.0
2.5
3.5
0.3
0.9

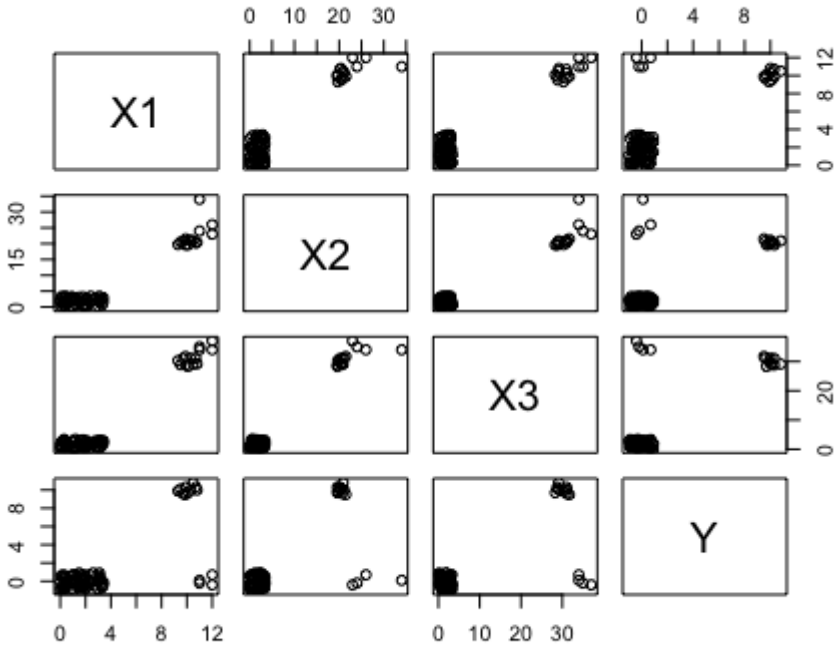

0.4
0.6
0.8
1.0
1.2
1.4
1.6
confidence intervals

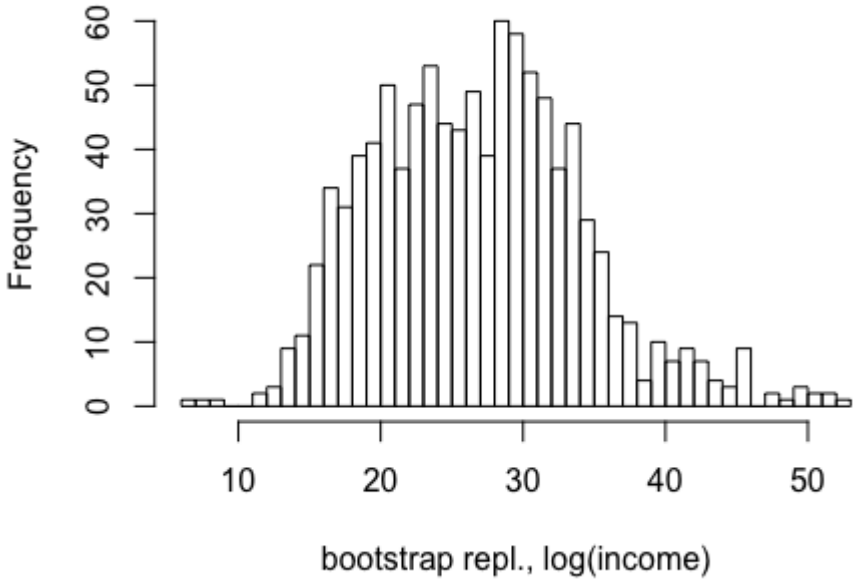


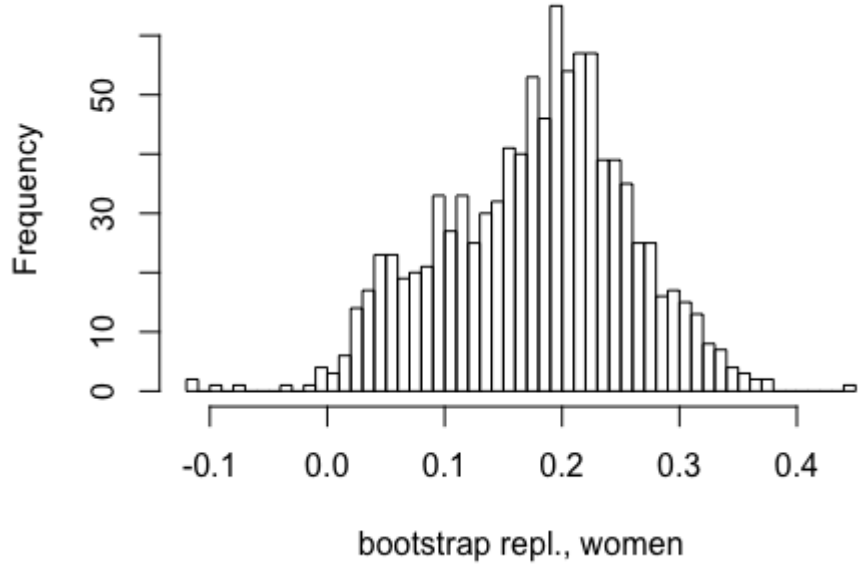
Chapter 8: Applications of Resampling Methods and Monte Carlo Tests

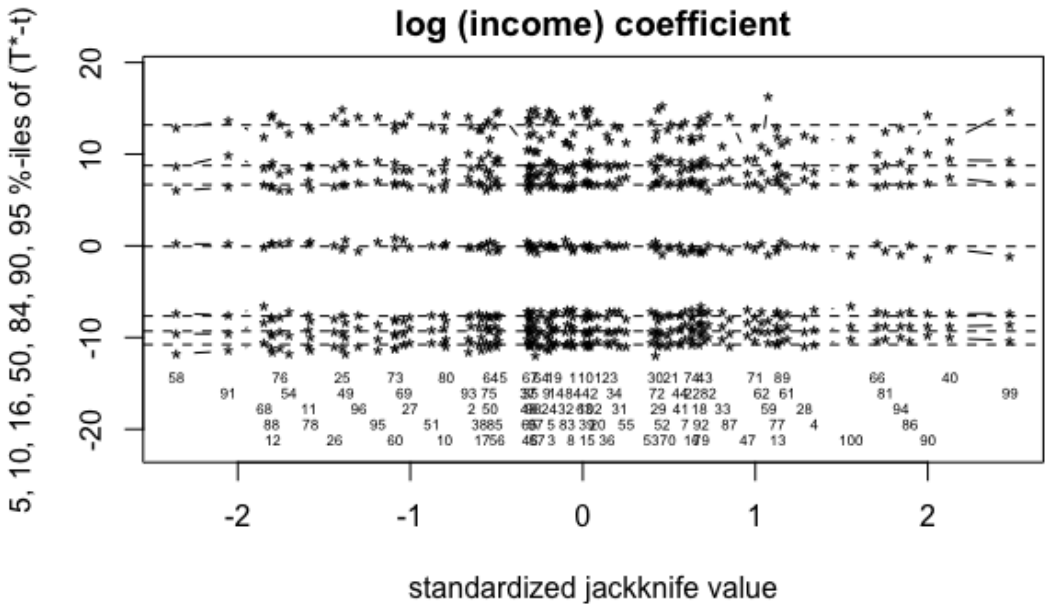


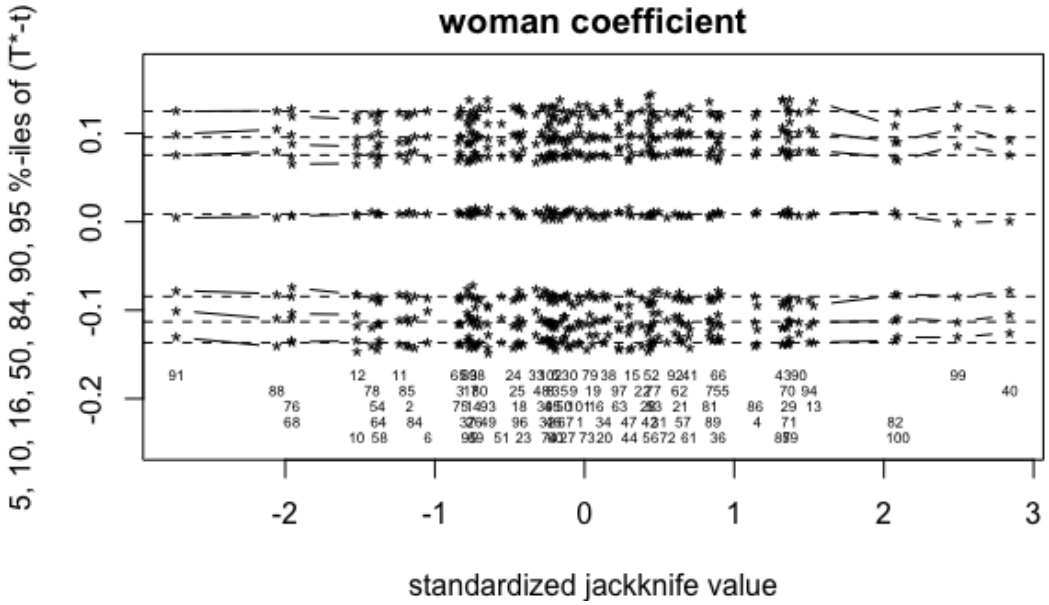

Sqrt of abs(Residuals) vs. Fitted Values
Sqrt of abs(Residuals)
0.5
1.5
2.5
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
Fitted Values
lmrob(formula = Y ~ ., data = hbk)

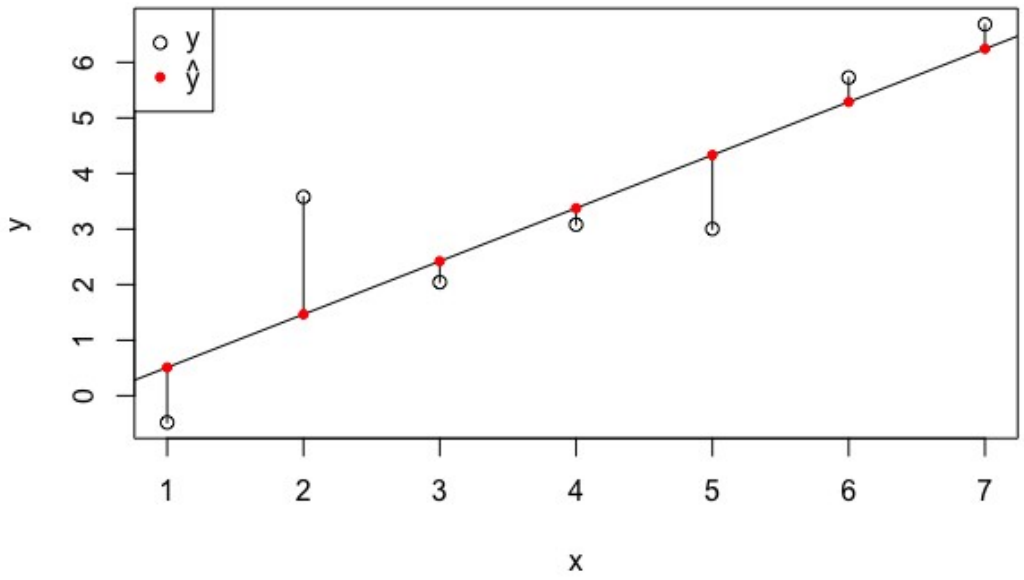

X1
X2
X3
Y
0 10 20 30
0 4 8
0 4 8 12
0 15 30
0 20

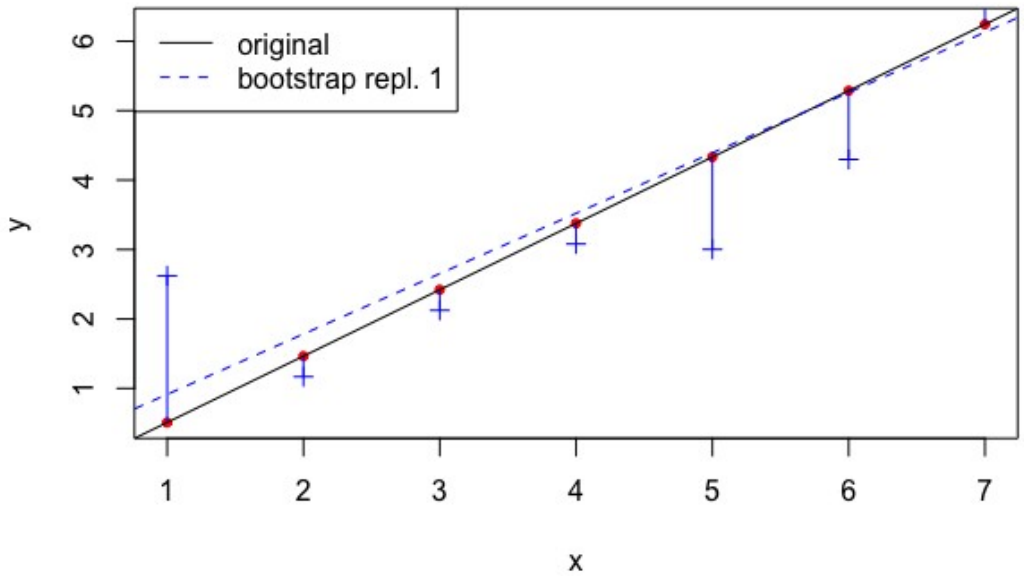

Frequency
0
10
20
30
40
50
60
10
20
30
40
50
bootstrap repl., log(income)

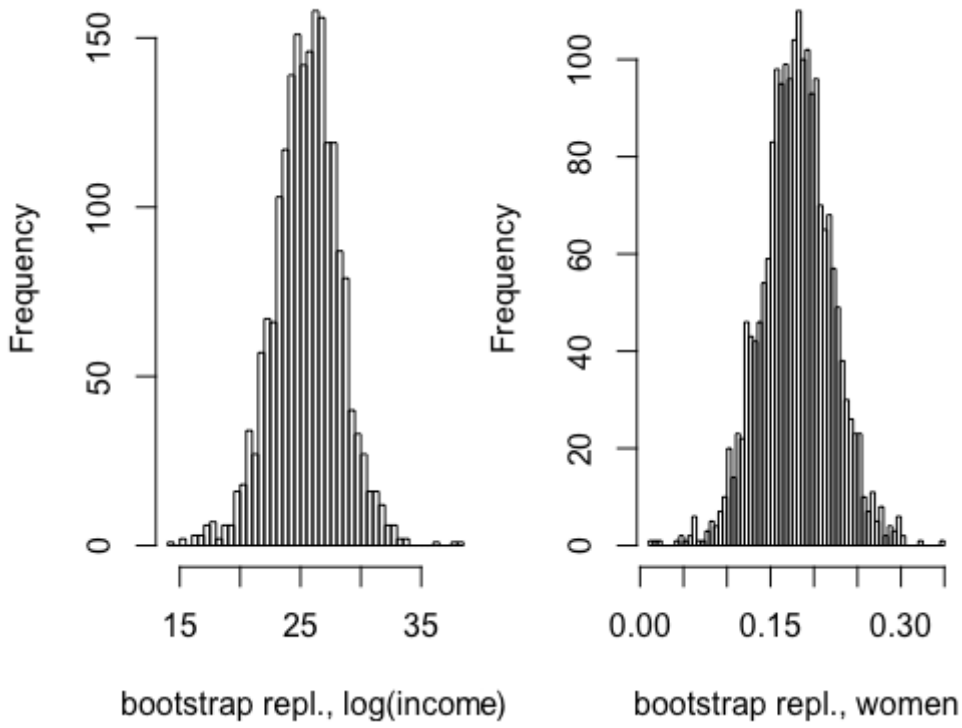

Frequency
0
10
30
50
-0.1
0.0
0.1
0.2
0.3
0.4
bootstrap repl., women

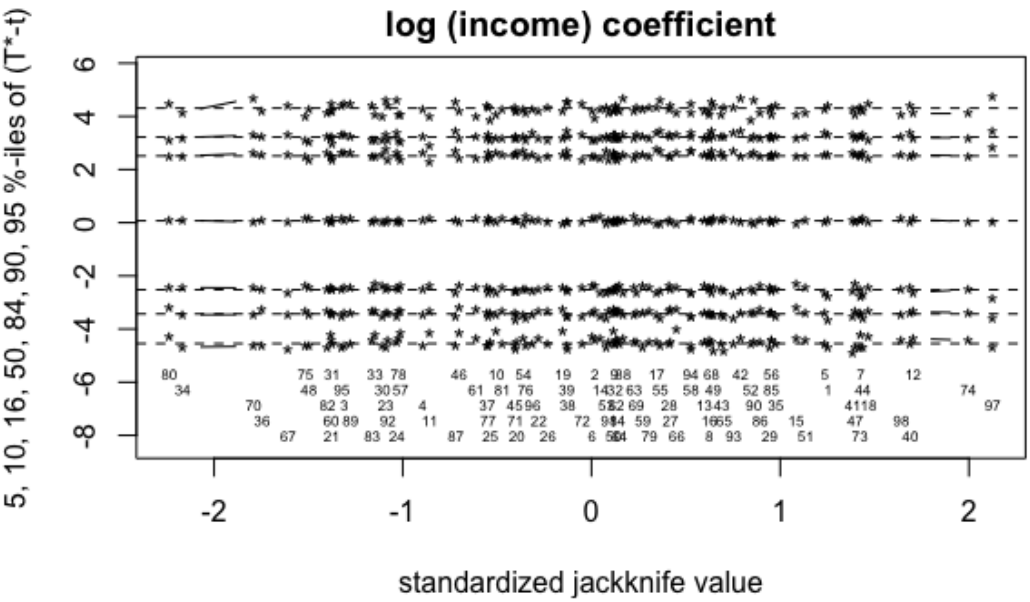

log (income) coefficient
5, 10, 16, 50, 84, 90, 95 %-iles of (T*-t)
20
10
0
-10
-20
-2
-1
0
1
2
standardized jackknife value

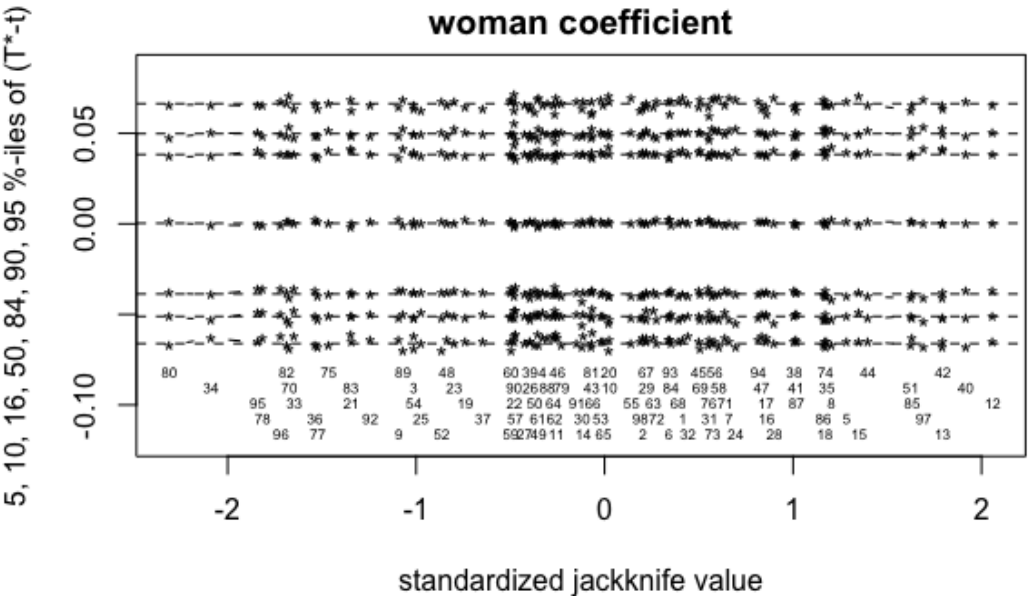

woman coefficient
5, 10, 16, 50, 84, 90, 95 %-iles of (T*-t)
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
3
standardized jackknife value

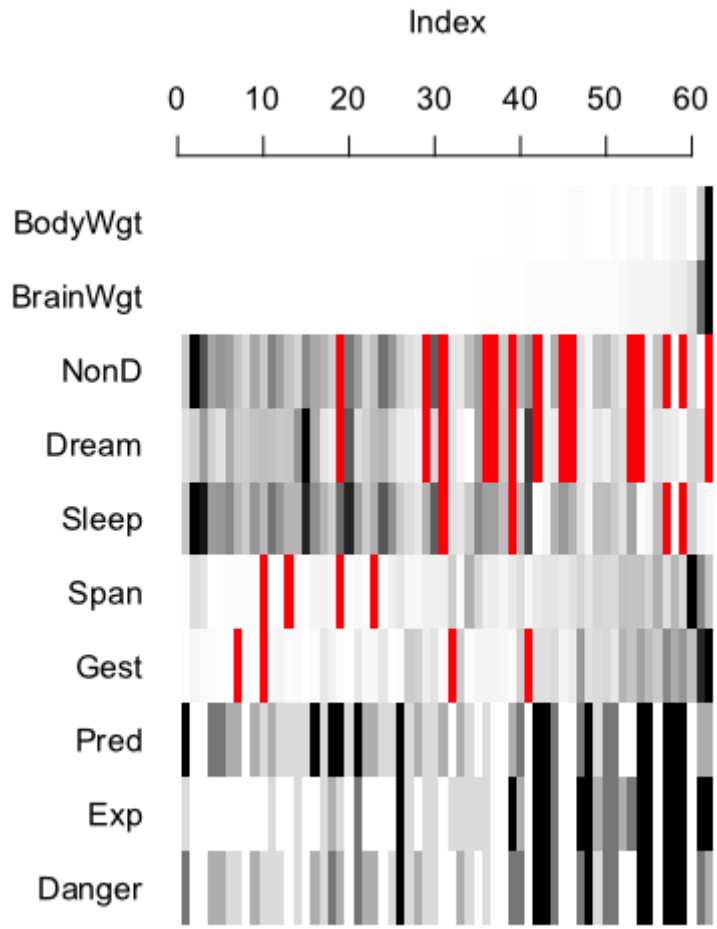

y
ŷ
y
x
0
1
2
3
4
5
6
1
2
3
4
5
6
7


original
bootstrap repl. 1
y
x
1
2
3
4
5
6
1
2
3
4
5
6
7

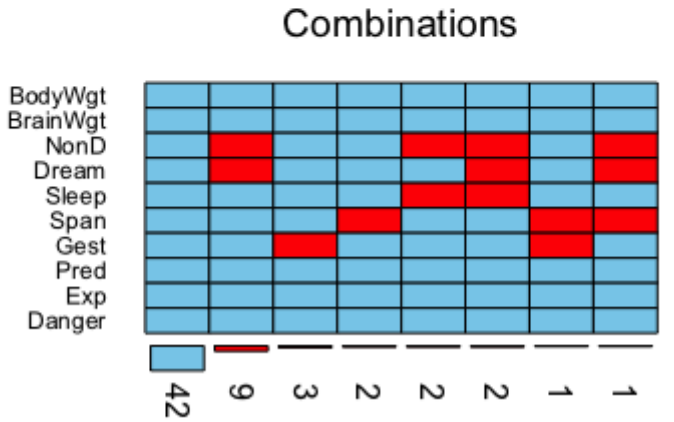

Frequency
0
50
100
150
15
25
35
bootstrap repl., log(income)
Frequency
0
20
40
60
80
100
0.00
0.15
0.30
bootstrap repl., women

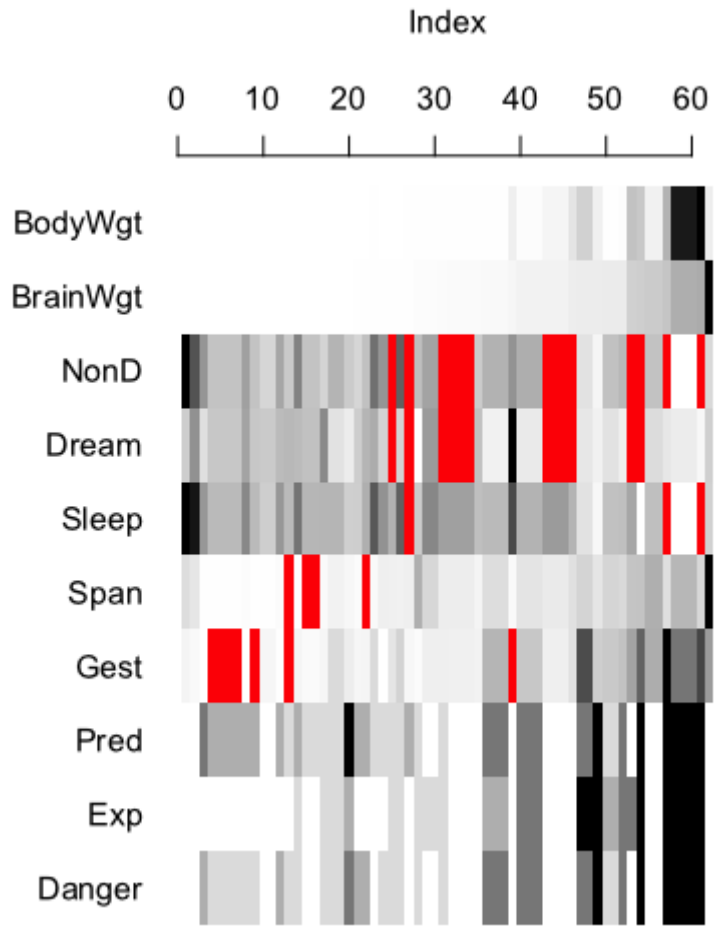

log (income) coefficient
5, 10, 16, 50, 84, 90, 95 %-iles of (T*-t)
standardized jackknife value

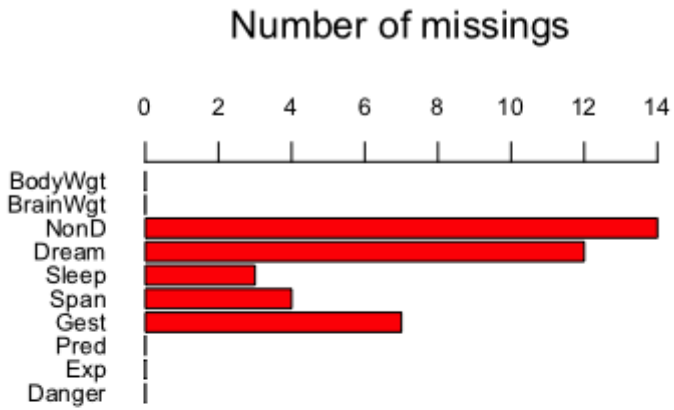

woman coefficient
5, 10, 16, 50, 84, 90, 95 %-iles of (T*-t)
standardized jackknife value

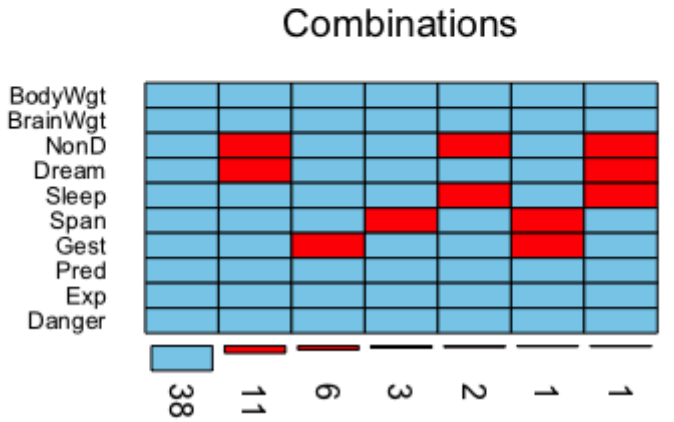

Index
0
10
20
30
40
50
60
BodyWgt
BrainWgt
NonD
Dream
Sleep
Span
Gest
Pred
Exp
Danger

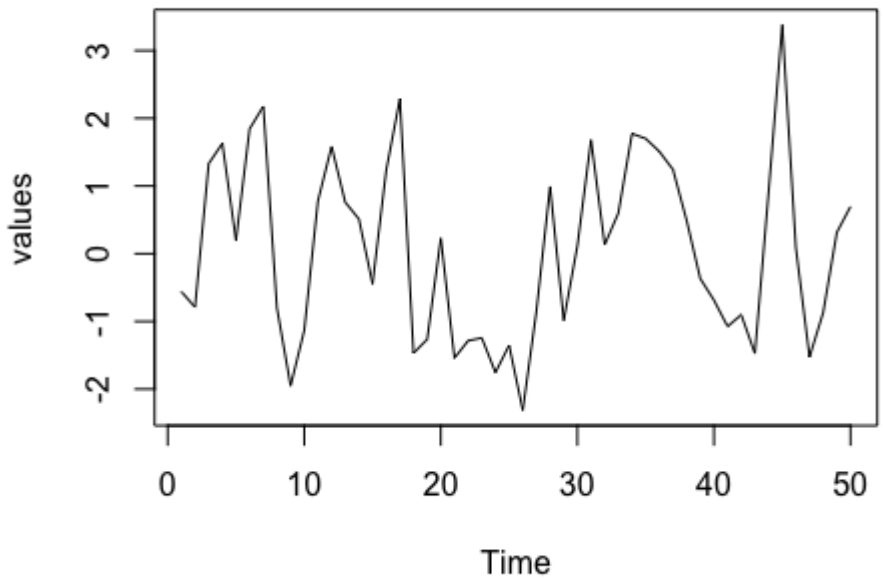

Number of missings
0
2
4
6
8
10
12
14
BodyWgt
BrainWgt
NonD
Dream
Sleep
Span
Gest
Pred
Exp
Danger

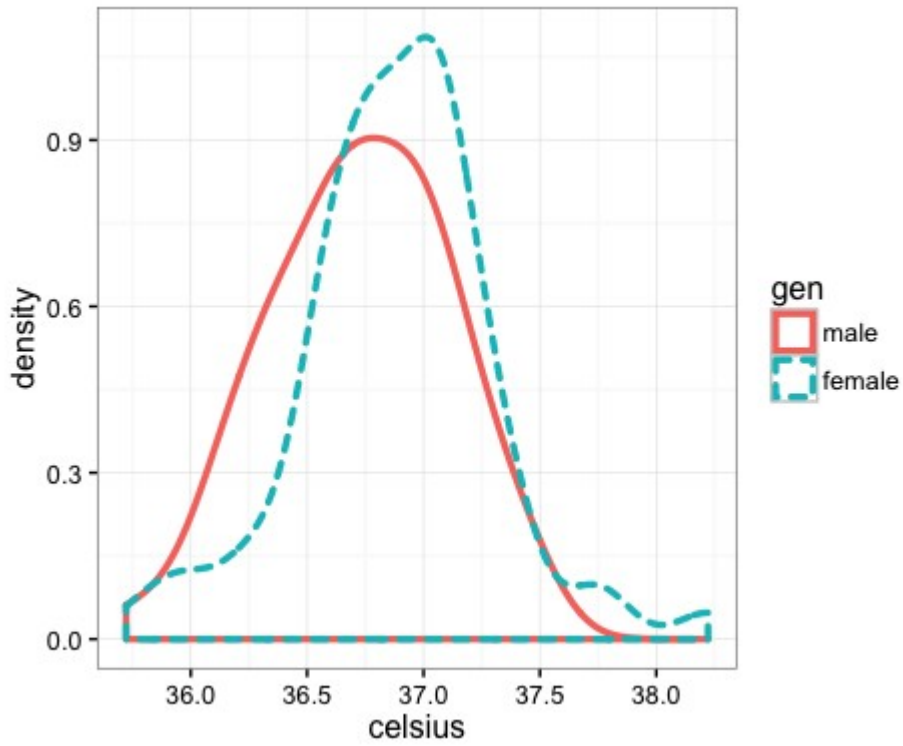

Combinations
BodyWgt
BrainWgt
NonD
Dream
Sleep
Span
Gest
Pred
Exp
Danger
42
9
3
2
2
2
1
1

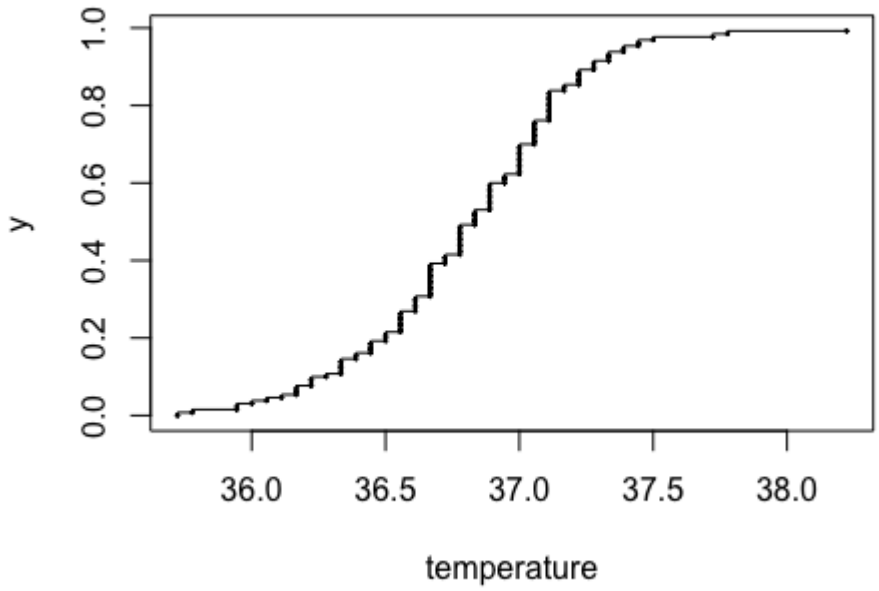

Index
0
10
20
30
40
50
60
BodyWgt
BrainWgt
NonD
Dream
Sleep
Span
Gest
Pred
Exp
Danger

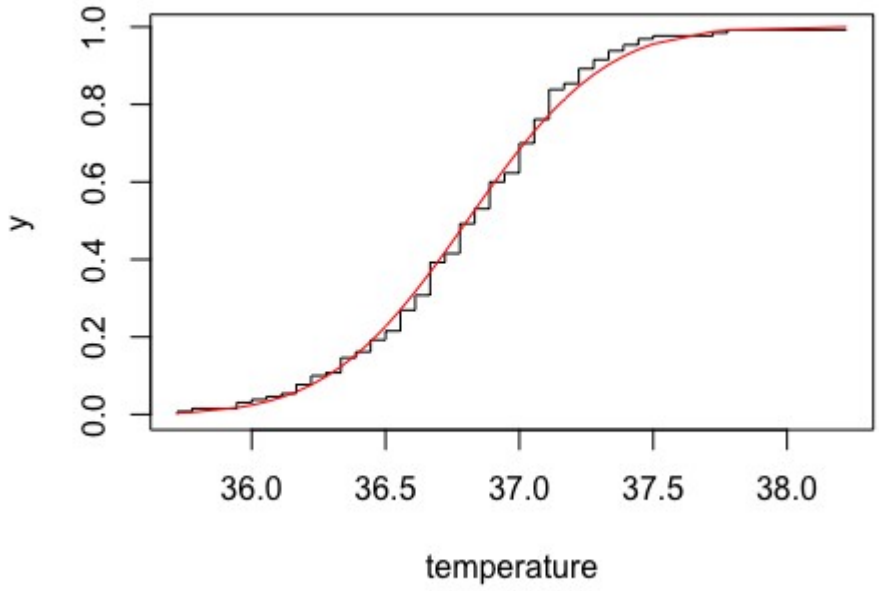

Number of missings
0
2
4
6
8
10
12
14
BodyWgt
BrainWgt
NonD
Dream
Sleep
Span
Gest
Pred
Exp
Danger

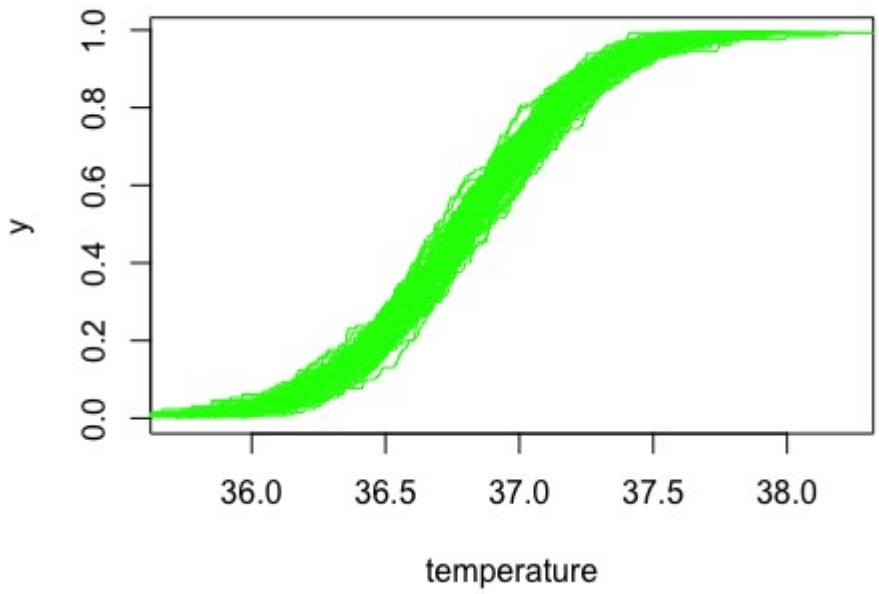

Combinations
BodyWgt
BrainWgt
NonD
Dream
Sleep
Span
Gest
Pred
Exp
Danger
38
11
6
3
2
1
1

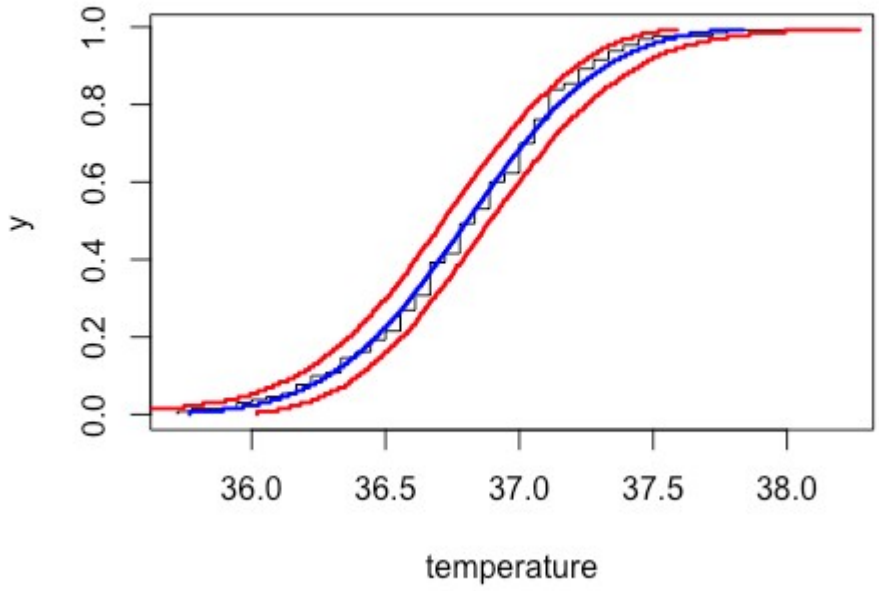

values
3
2
1
0
-1
-2
0
10
20
30
40
50
Time

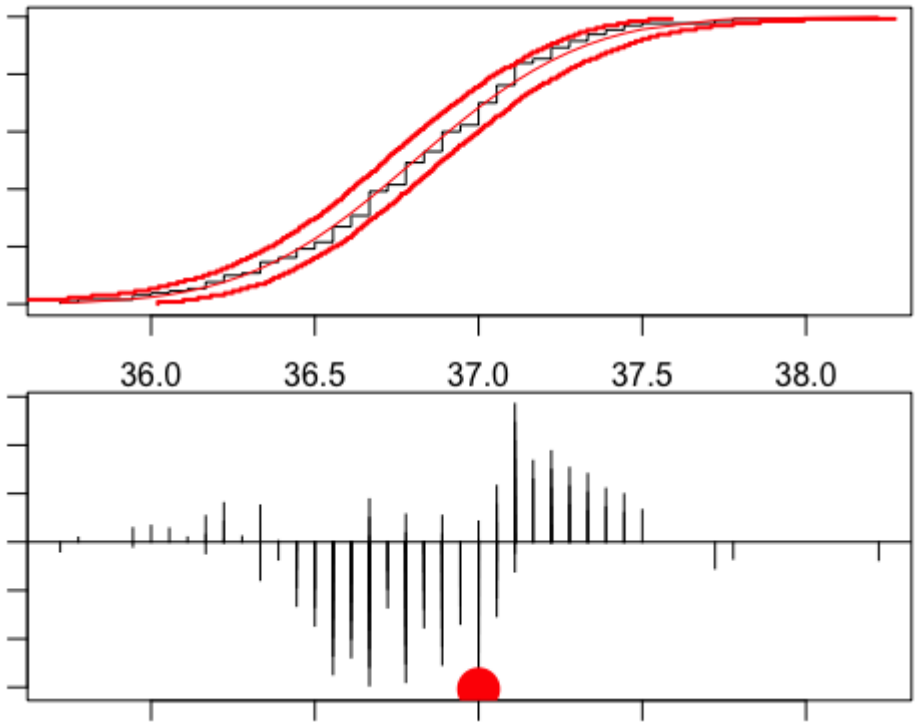

density
0.9
0.6
0.3
0.0
36.0
36.5
37.0
37.5
38.0
celsius
gen
male
female


y
0.0
0.2
0.4
0.6
0.8
1.0
36.0
36.5
37.0
37.5
38.0
temperature

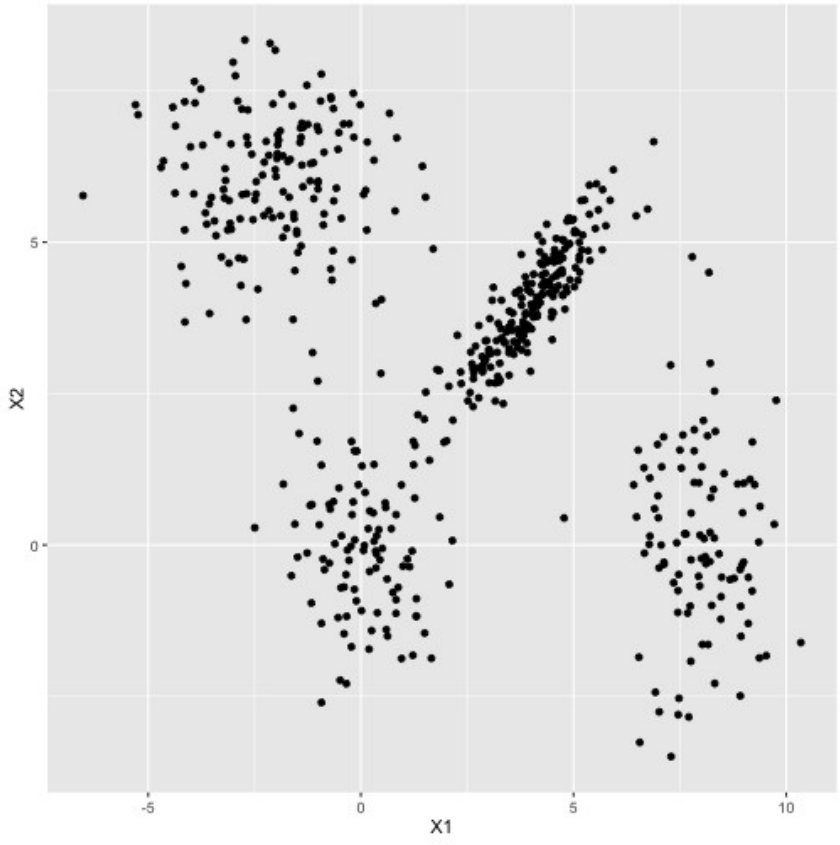

y
0.0
0.2
0.4
0.6
0.8
1.0
36.0
36.5
37.0
37.5
38.0
temperature

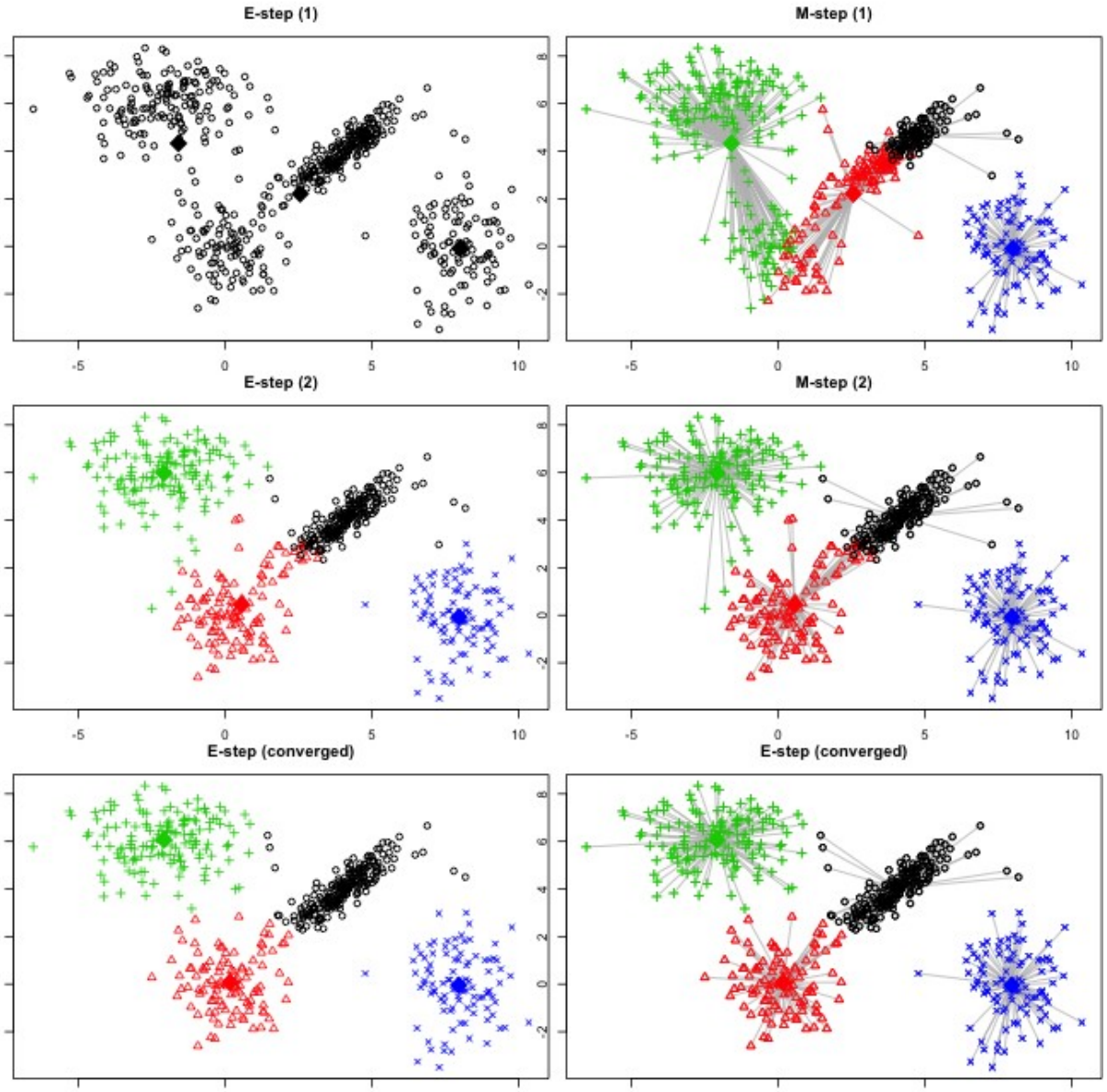

y
0.0
0.2
0.4
0.6
0.8
1.0
36.0
36.5
37.0
37.5
38.0
temperature


y
0.0
0.2
0.4
0.6
0.8
1.0
36.0
36.5
37.0
37.5
38.0
temperature

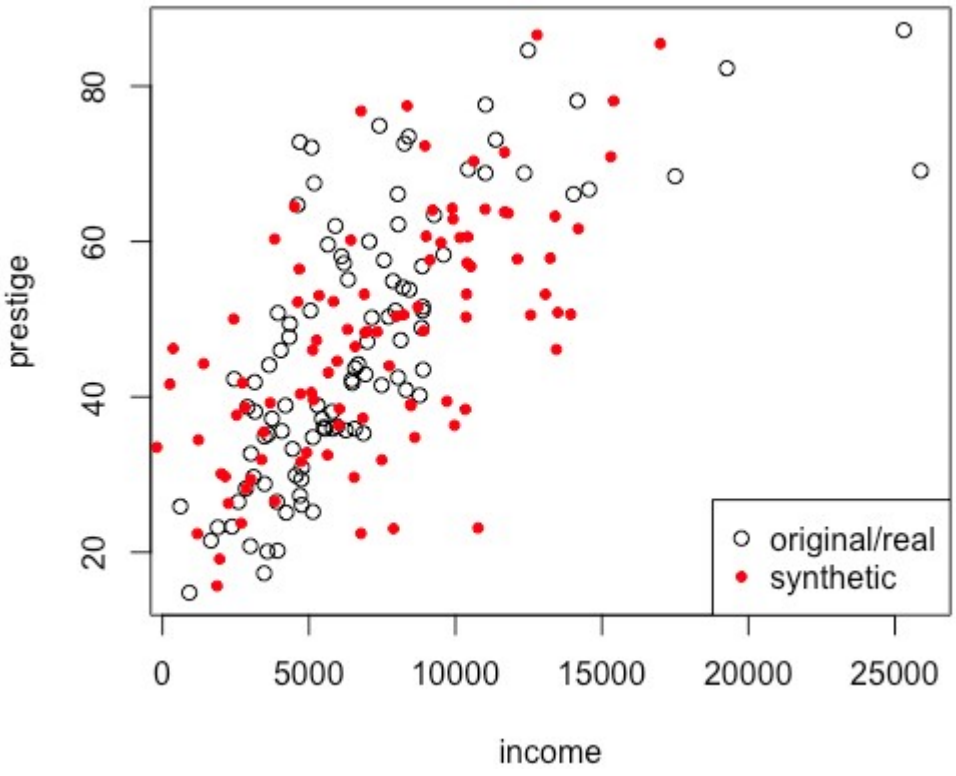


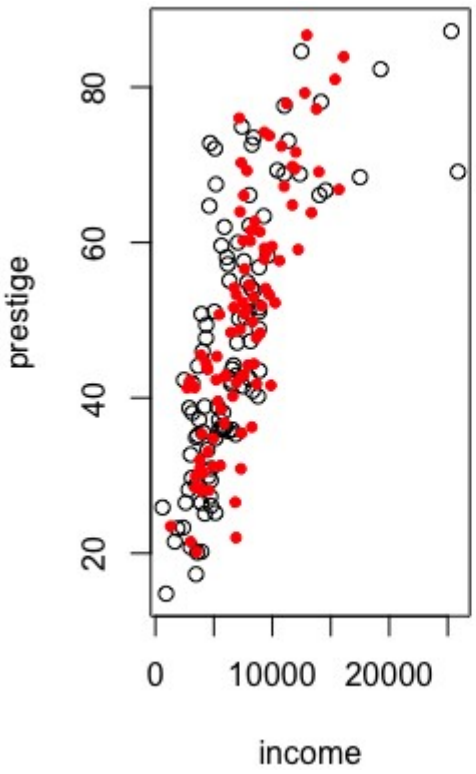
Chapter 9: The EM Algorithm

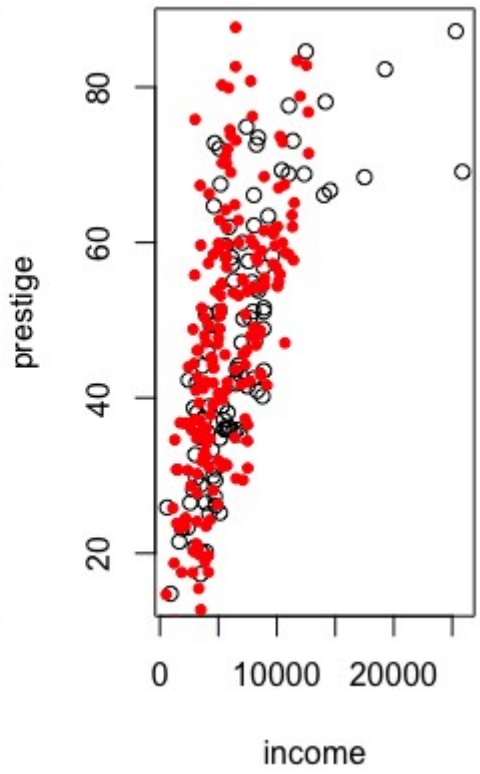


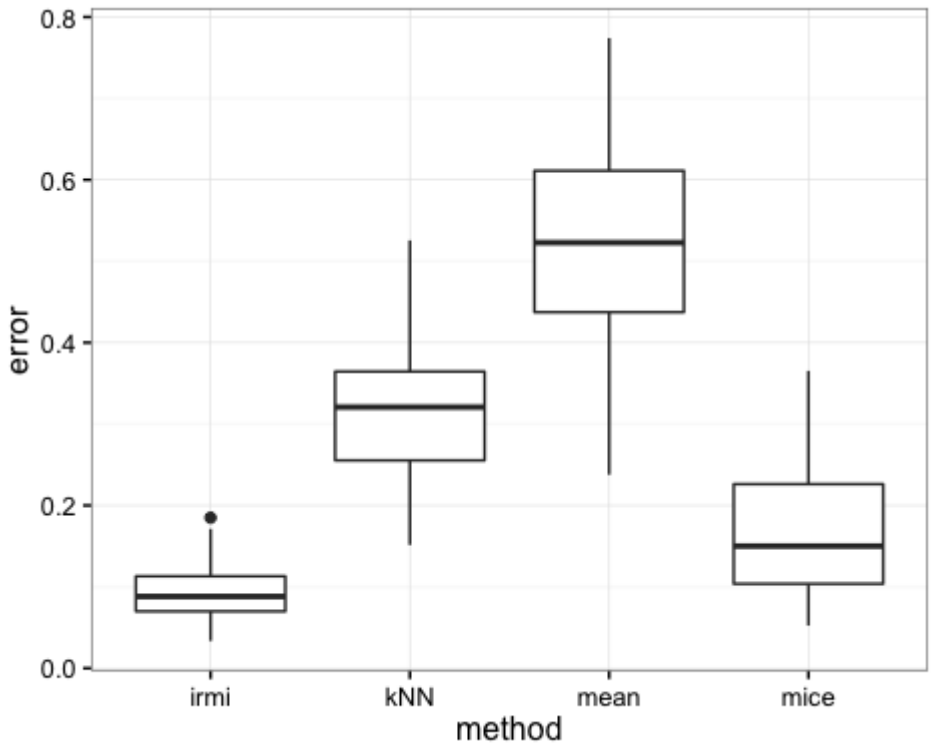

E-step (1)
M-step (1)
E-step (2)
M-step (2)
E-step (converged)
E-step (converged)

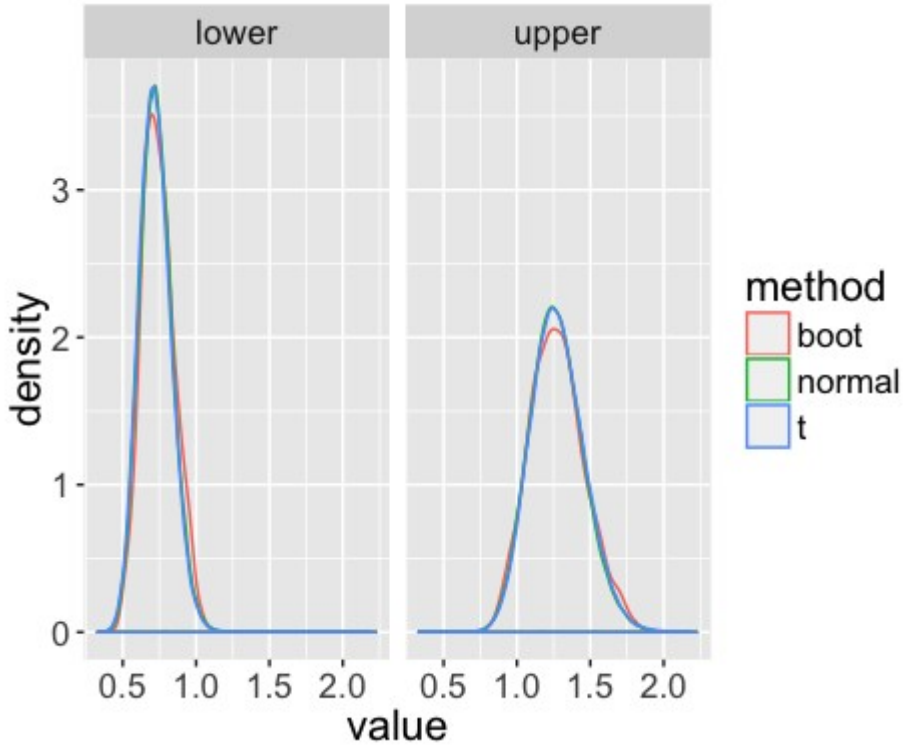
Chapter 9: The EM Algorithm

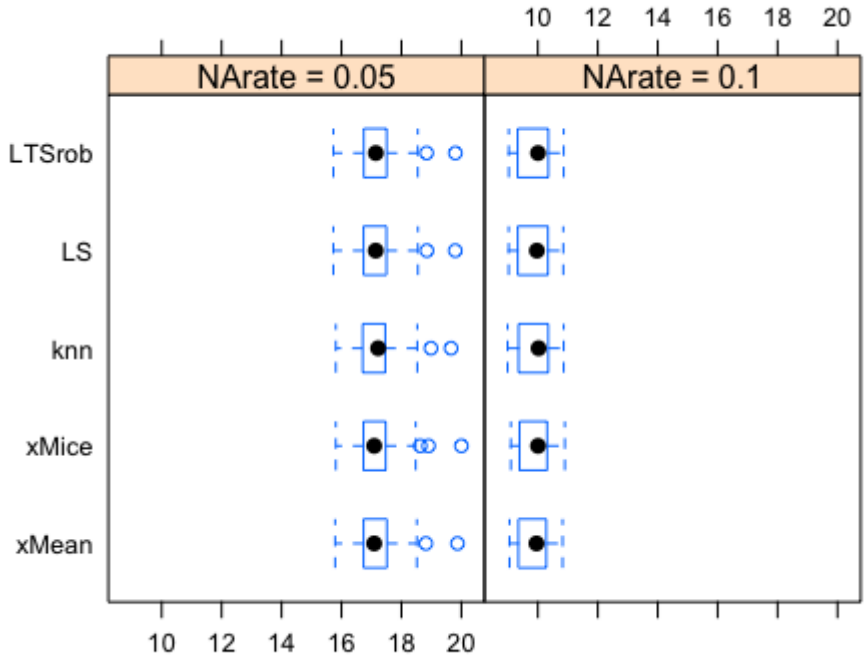


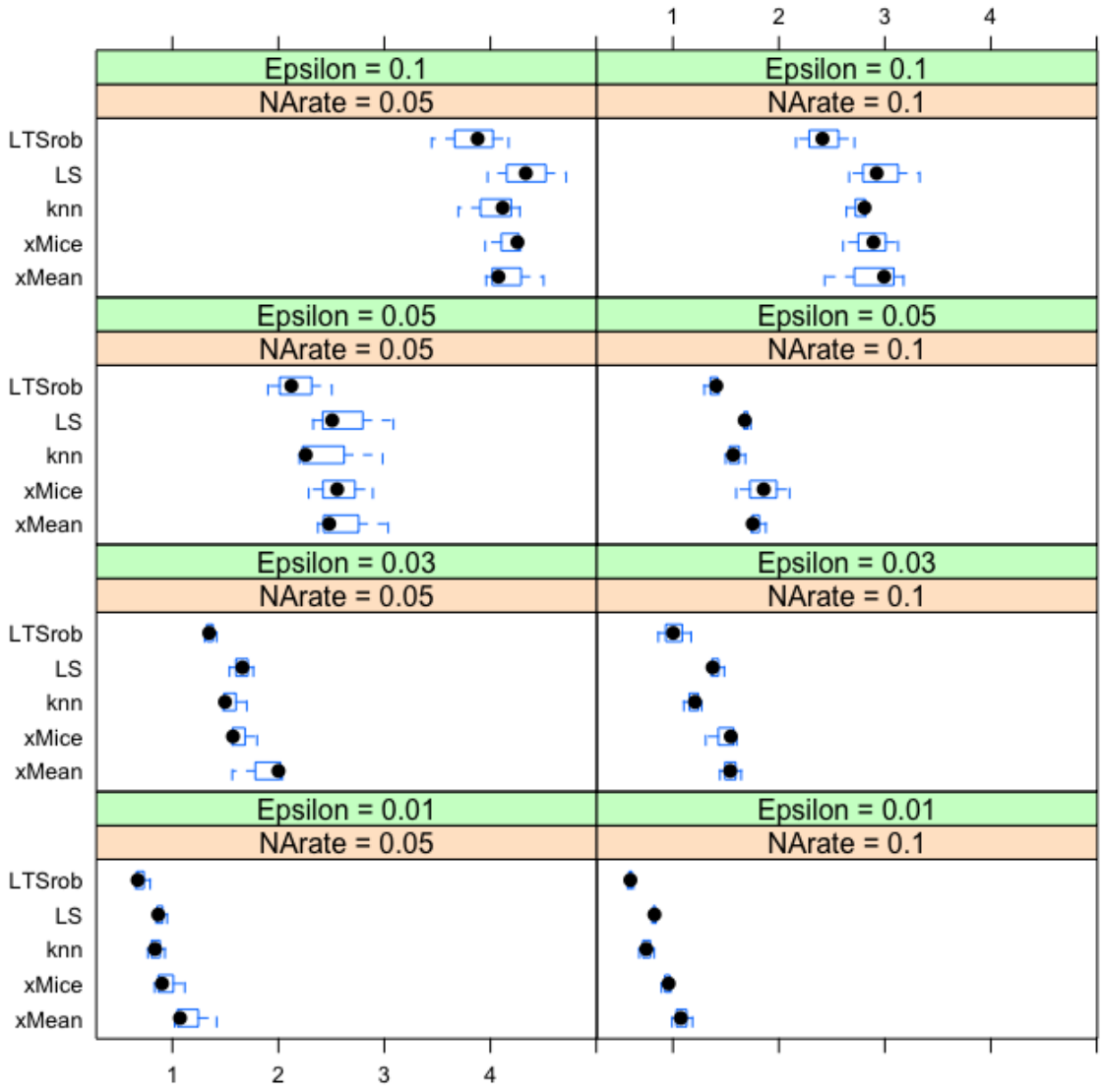


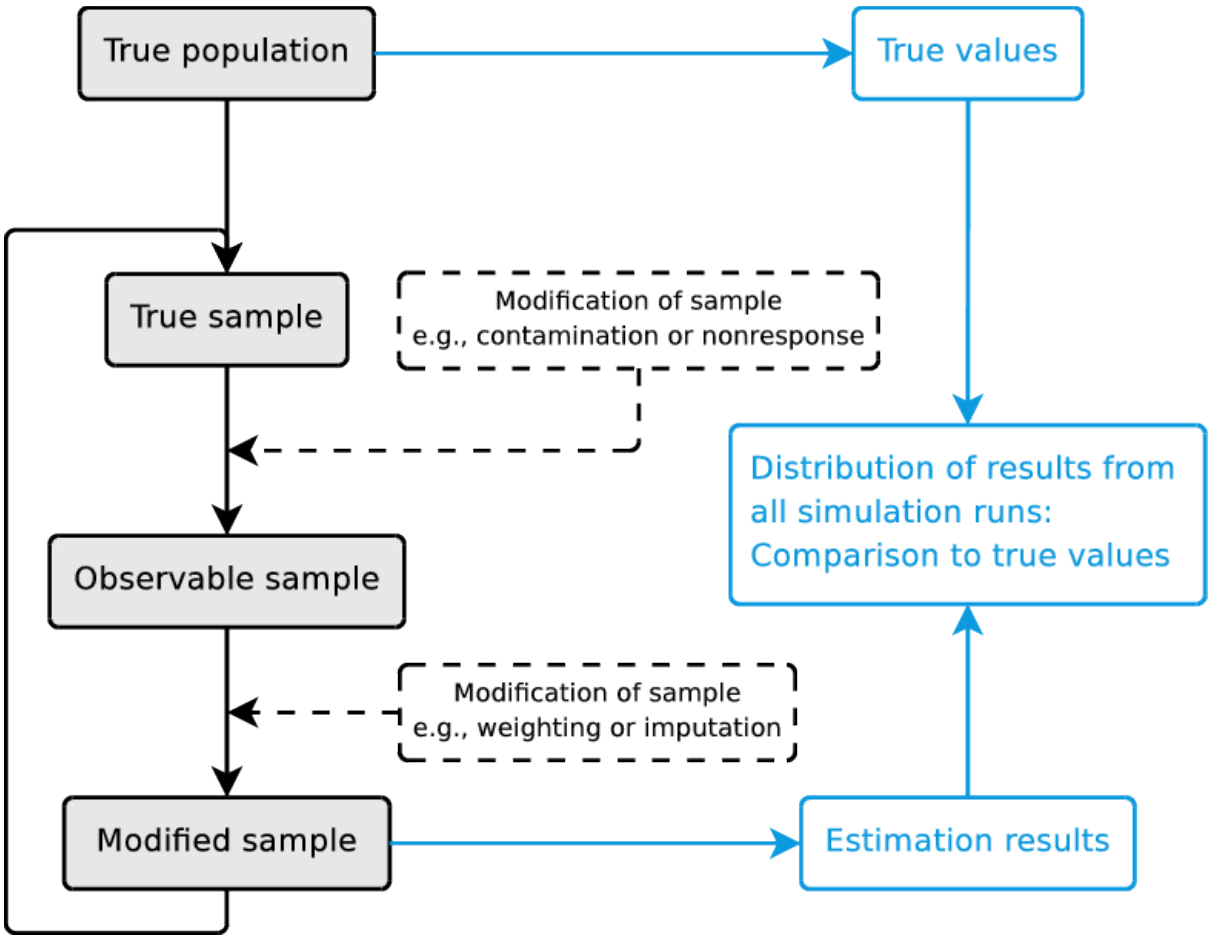


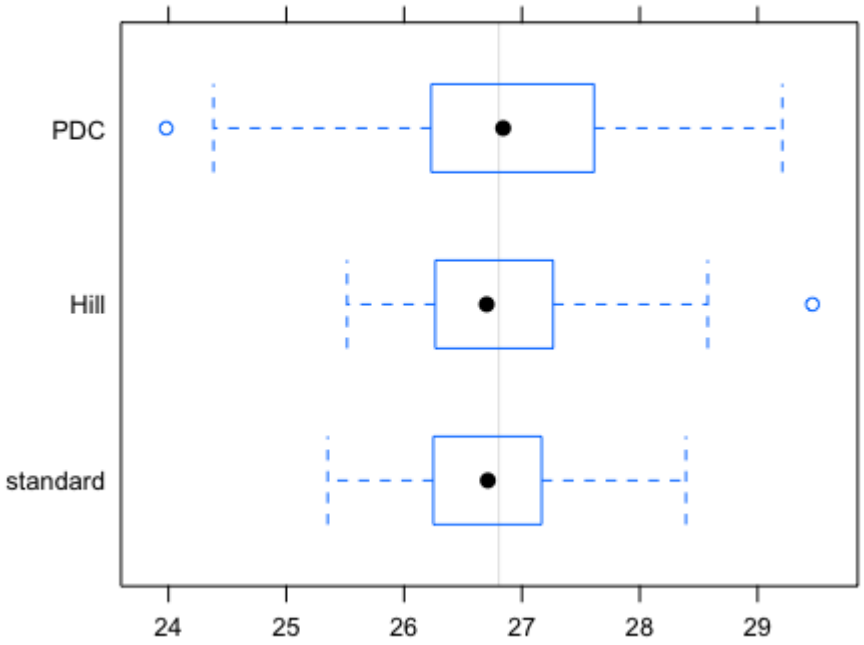

0.8
0.6
0.4
0.2
0.0
error
irmi
kNN
mean
mice
method

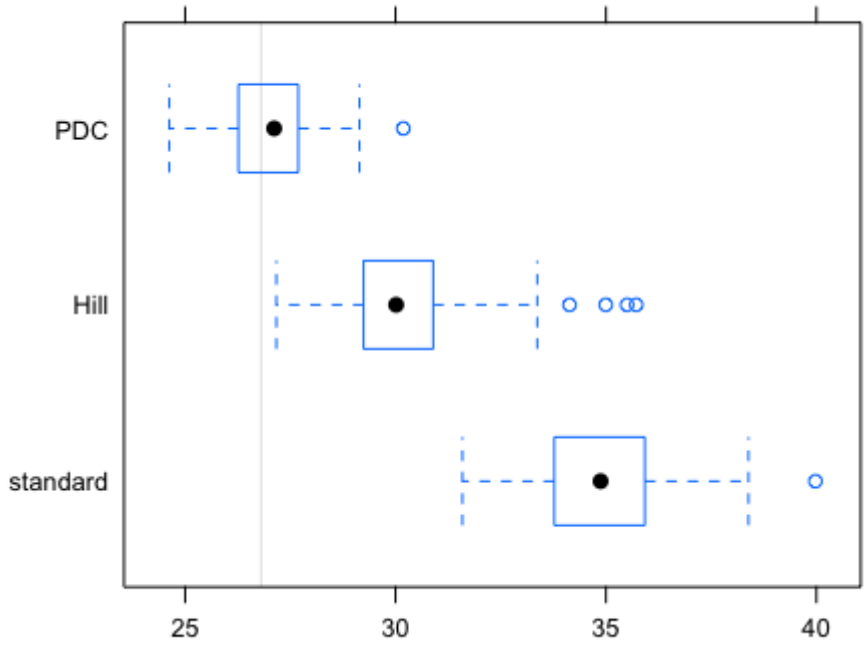

lower
upper
3
2
1
0
density
0.5
1.0
1.5
2.0
value
method
boot
normal
t

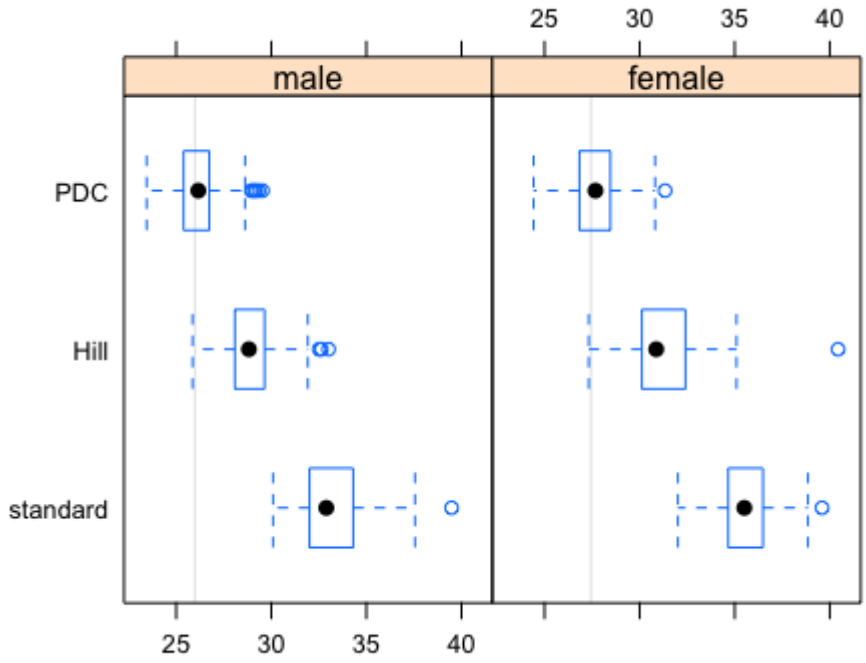

NArate = 0.05
NArate = 0.1
LTSrob
LS
knn
xMice
xMean
10
12
14
16
18
20


Epsilon = 0.1
NArate = 0.05
Epsilon = 0.1
NArate = 0.1
Epsilon = 0.05
NArate = 0.05
Epsilon = 0.05
NArate = 0.1
Epsilon = 0.03
NArate = 0.05
Epsilon = 0.03
NArate = 0.1
Epsilon = 0.01
NArate = 0.05
Epsilon = 0.01
NArate = 0.1
LTSrob
LS
knn
xMice
xMean
1
2
3
4

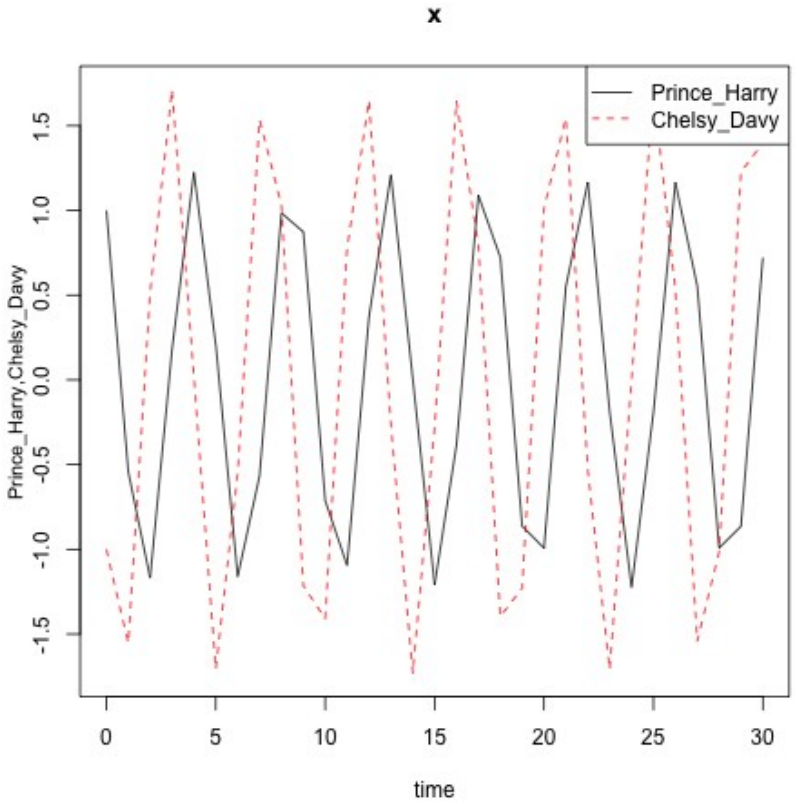

True population
True values
True sample
Modification of sample
e.g., contamination or nonresponse
Distribution of results from
all simulation runs:
Comparison to true values
Observable sample
Modification of sample
e.g., weighting or imputation
Modified sample
Estimation results

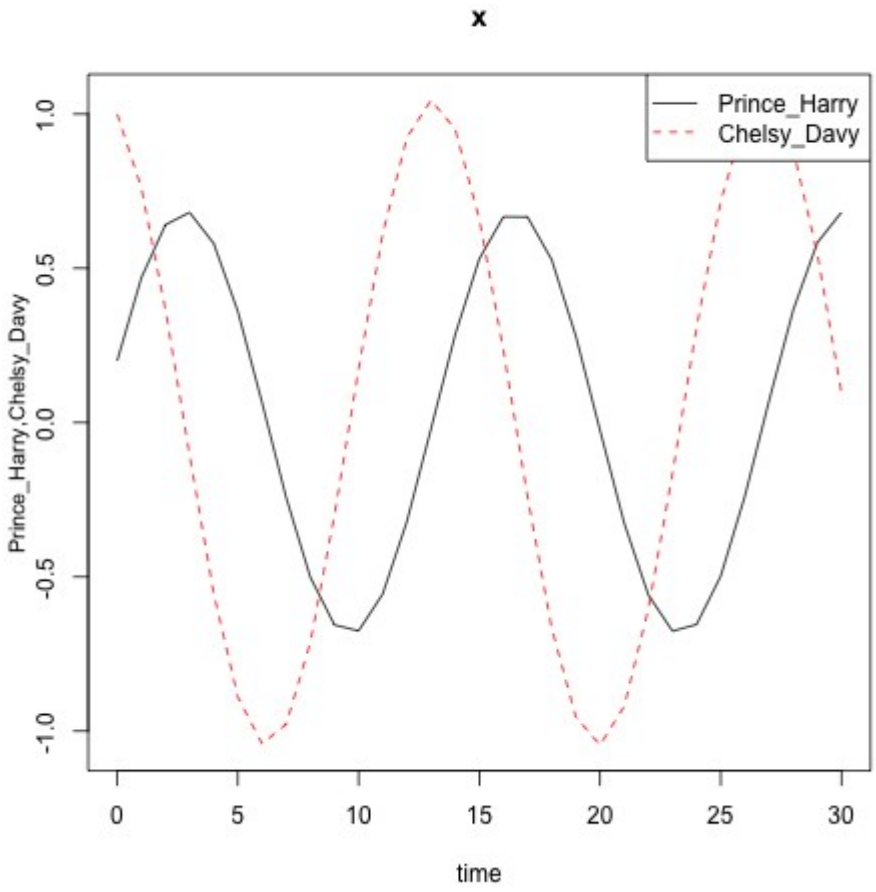

PDC
Hill
standard
24
25
26
27
28
29

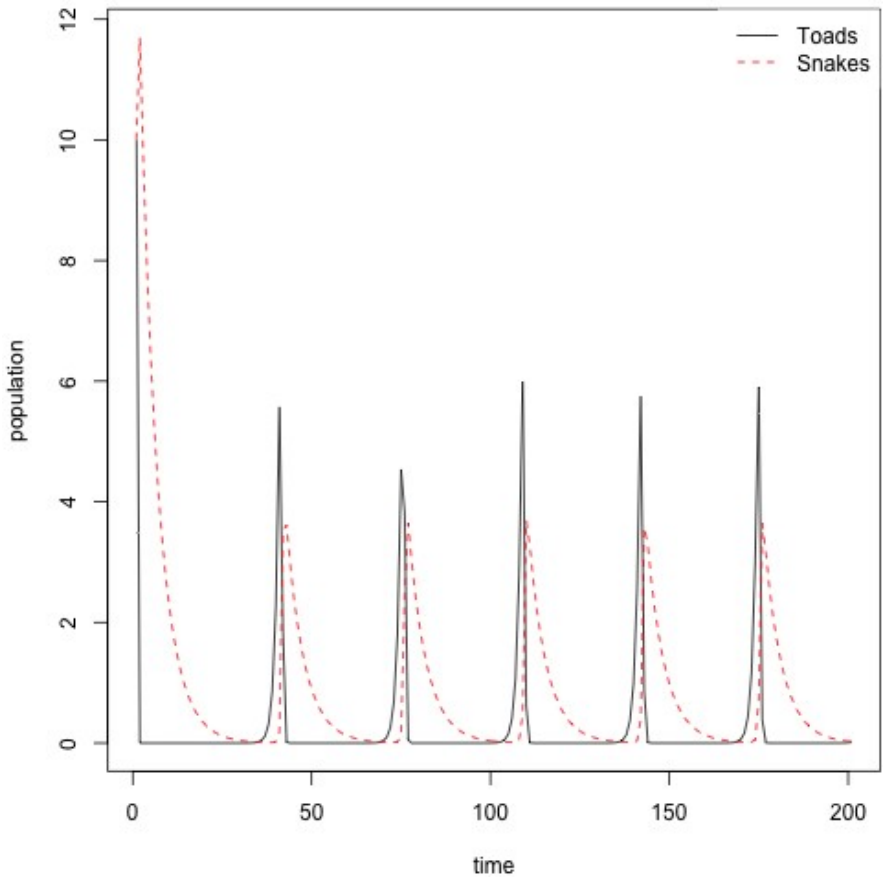

PDC
Hill
standard
25
30
35
40


25
30
35
40
male
female
PDC
Hill
standard
25
30
35
40

Chapter 11: System Dynamics and Agent-Based Models




x
Prince_Harry
Chelsy_Davy
Prince_Harry,Chelsy_Davy
1.0
0.5
0.0
-0.5
-1.0
0
5
10
15
20
25
30
time


Toads
Snakes
population
12
10
8
6
4
2
0
0
50
100
150
200
time