Homework 9: Bootstrapping (Solutions)

Oliver d'Pug

Create a b.var function

I'm lazy and will use the **boot** package to run my bootstrap. I need a function that computes the variance that can be passed to the **boot** function.

```
b.var = function(d, i){
                        var(d[i])
                        }
```

Create a dataset of 100 normal observations

```
args(rnorm)
```

```
function (n, mean = 0, sd = 1)
NULL
```

```
n = 100
mu = 2
sig = 7
normal_data = rnorm(n, mu, sig)
write.csv(normal_data, "hw9_normal_data.csv")
c(summary(normal_data), var_x=var(normal_data), sd_x=sd(normal_data))
```

```
     Min.    1st Qu.     Median       Mean    3rd Qu.       Max.      var_x
-11.825385  -3.140909   1.204382   2.123063   7.292964  19.047259  48.100915
     sd_x
  6.935482
```

Bootstrap the variance using b.var

```
### Use the boot function to run the bootstrap
normal_b.var = boot(normal_data, b.var, R=9999)
normal_b.var
```

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = normal_data, statistic = b.var, R = 9999)


Bootstrap Statistics :
    original      bias    std. error
t1* 48.10092 -0.4462118    5.802214
```

How much bias does the estimate have? Is the bootstrap distribution normal or chisquare?

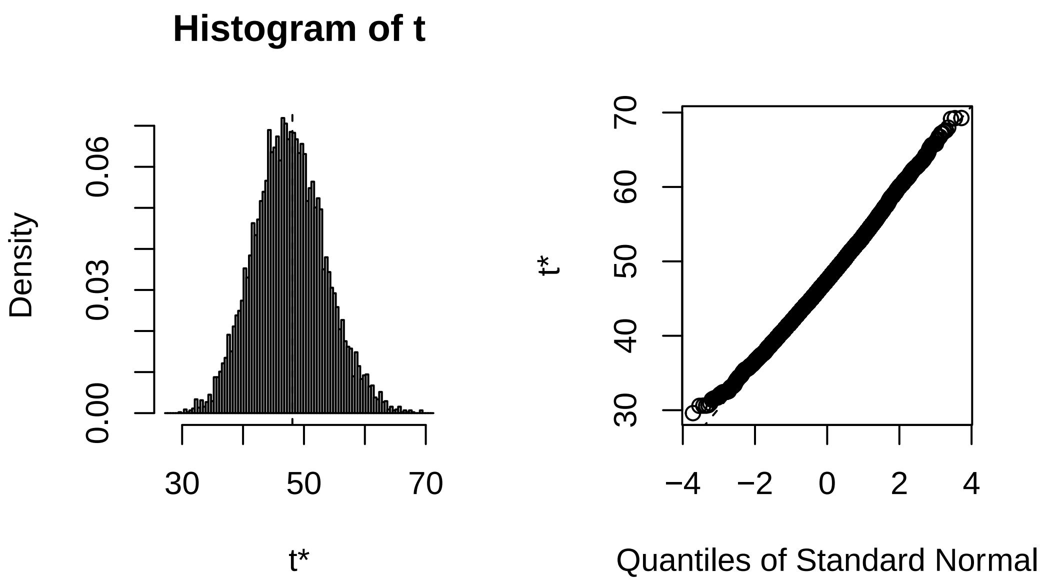
```
attributes(summary(normal_b.var))
```

```
$names
[1] "R"        "original" "bootBias" "bootSE"   "bootMed"

$row.names
[1] 1

$class
[1] "summary.boot" "data.frame"
```

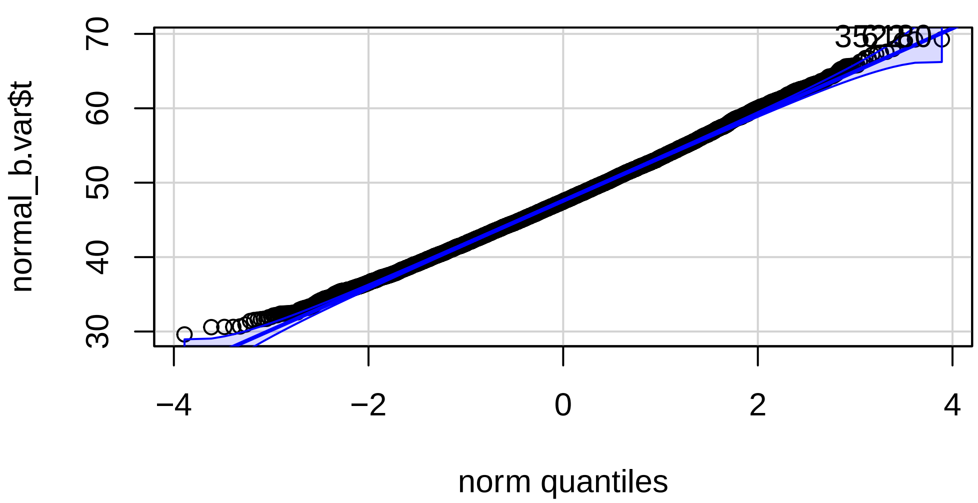
```
summary(normal_b.var)
```

```
     R original bootBias bootSE bootMed
1 9999   48.101 -0.44621 5.8022  47.494
```

```
plot(normal_b.var)
```

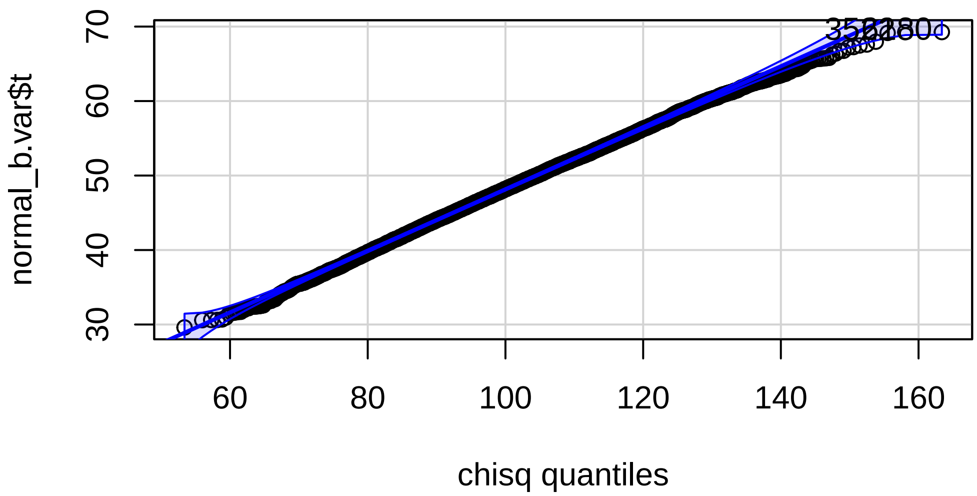



```
qqPlot(normal_b.var$t, distribution="norm")
```




```
[1] 6180 3522
```

```
qqPlot(normal_b.var$t, distribution="chisq", df=n-1)
```




```
[1] 6180 3522
```

The variance estimator appears to have bias=-0.4462118 and SE=5.8022137 when the data consist of 100 observations randomly drawn from a normal with mean $\mu = 2$ and variance $\sigma^2 = 49$. The variance appears to be neither normally nor chisquare distributed, although the chisquare may fit a bit better.

Create a dataset of 100 standard normal observations

```
  args(rnorm)
```

```
function (n, mean = 0, sd = 1)
NULL
```

```
  n = 100
  mu = 0
  sig = 1
  std_normal_data = rnorm(n, mu, sig)
  write.csv(std_normal_data, "hw9_std_normal_data.csv")
  c(summary(std_normal_data), var_x=var(std_normal_data), sd_x=sd(std_normal_data))
```

```
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.      var_x 
-2.3062419 -0.5809928  0.1401183  0.1373739  0.7372962  3.1574952  1.0463991 
      sd_x 
 1.0229365 
```

Bootstrap the variance using b.var

```
### Use the boot function to run the bootstrap
std_normal_b.var = boot(std_normal_data, b.var, R=9999)
std_normal_b.var
```

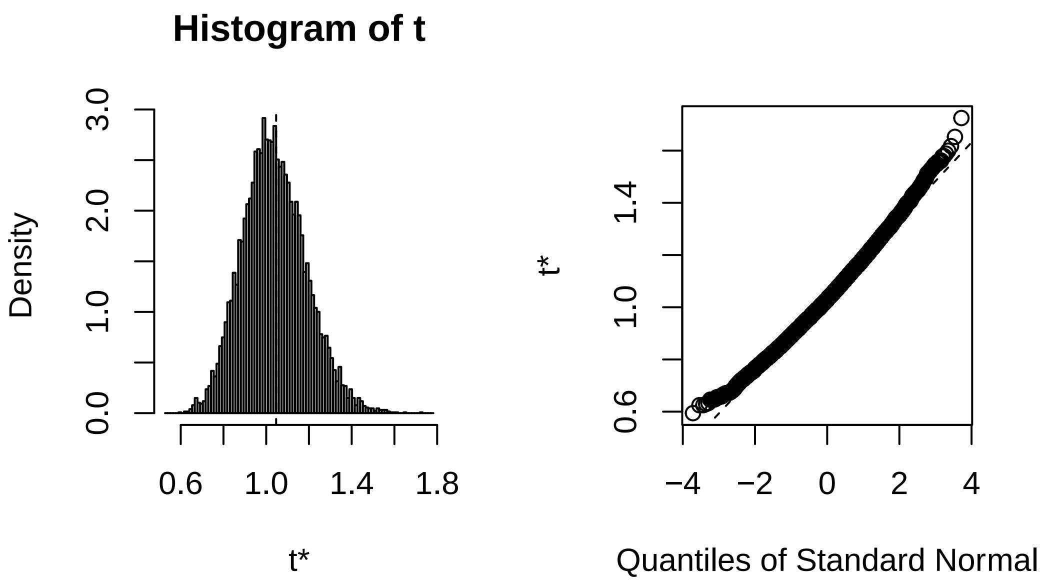
```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = std_normal_data, statistic = b.var, R = 9999)


Bootstrap Statistics :
    original        bias    std. error
t1* 1.046399 -0.008188473    0.148309
```

How much bias does the estimate have? Is the bootstrap distribution normal or chisquare?

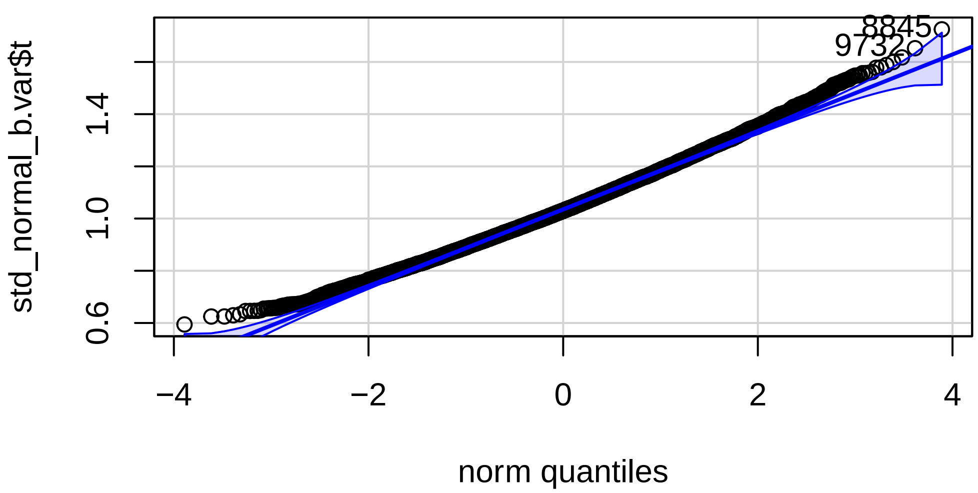
```
summary(std_normal_b.var)
```

```
     R original    bootBias  bootSE bootMed
1 9999   1.0464 -0.0081885 0.14831  1.0299
```

```
plot(std_normal_b.var)
```

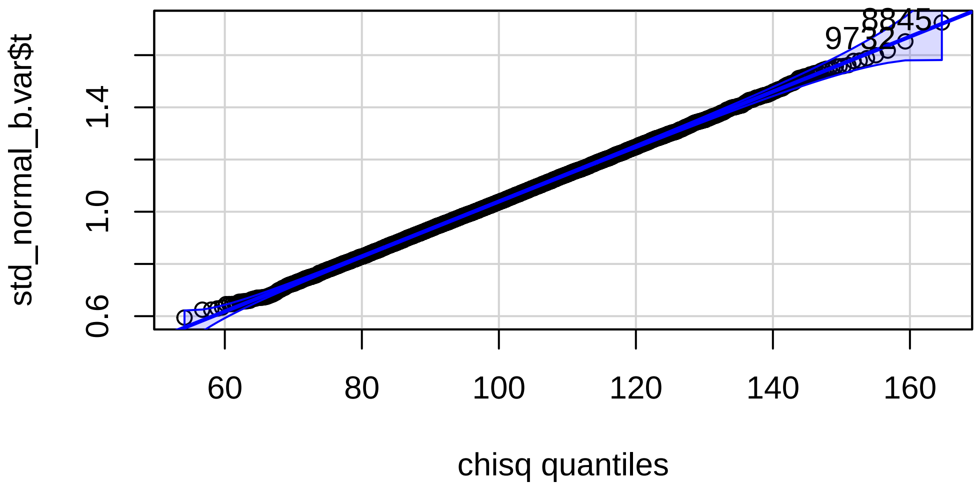



```
qqPlot(std_normal_b.var$t, distribution="norm")
```




```
[1] 8845 9732
```

```
qqPlot(std_normal_b.var$t, distribution="chisq", df=n)
```




```
[1] 8845 9732
```

The variance estimator appears to have bias=-0.0081885 and SE=0.148309 when the data consist of 100 observations randomly drawn from a normal with mean $\mu = 0$ and variance $\sigma^2 = 1$. The variance appears not to be normally distributed. However, it appears to have a chisquare distribution. The MGF of the sum of squared standard normals supports this contention.

Create a dataset of 100 U(0,1)

```
args(rnorm)
```

```
function (n, mean = 0, sd = 1)
NULL
```

```
n = 100
a = 0
b = 1
unif_data = runif(n, a, b)
write.csv(unif_data, "hw9_unif_data.csv")
c(summary(unif_data), var_x=var(unif_data), sd_x=sd(unif_data))
```

```
        Min.       1st Qu.        Median          Mean       3rd Qu.          Max. 
0.0007055753 0.1863097847 0.4563563421 0.4718261137 0.7602737419 0.9714556092 
       var_x         sd_x 
0.0976187830 0.3124400470 
```

Bootstrap the variance using b.var

```
### Use the boot function to run the bootstrap
unif_b.var = boot(unif_data, b.var, R=9999)
unif_b.var
```

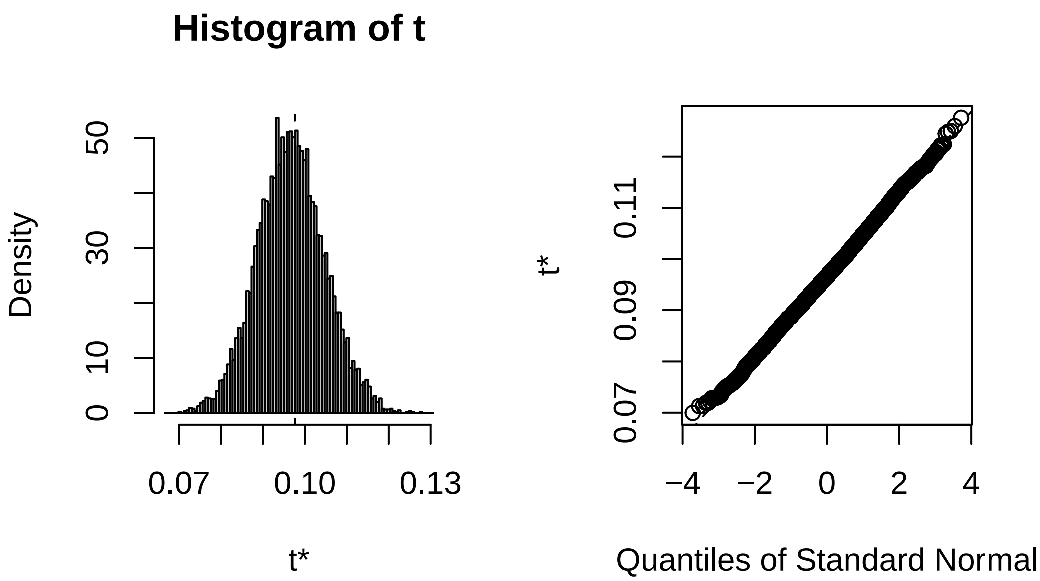
```


ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = unif_data, statistic = b.var, R = 9999)


Bootstrap Statistics :
      original          bias    std. error
t1* 0.09761878 -0.0008691127 0.008003808
```

How much bias does the estimate have? Is the bootstrap distribution normal or chisquare?

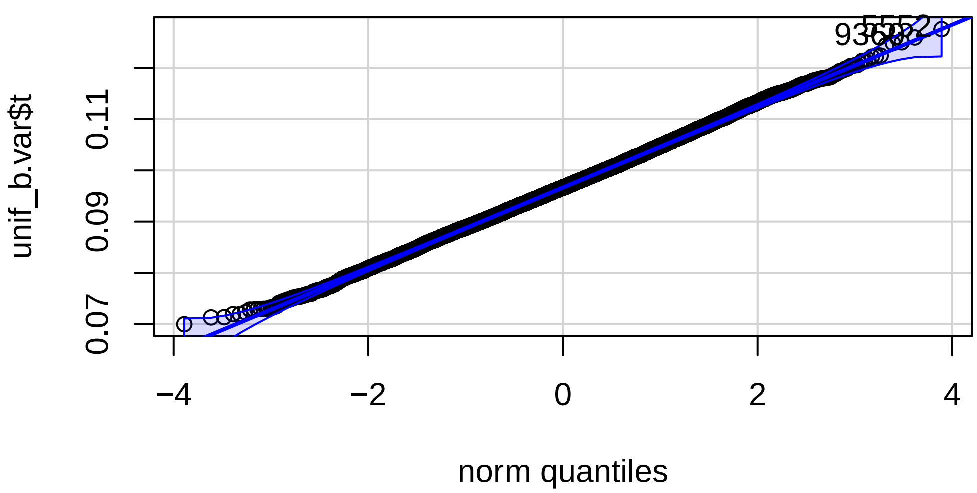
```
summary(unif_b.var)
```

```
     R original     bootBias    bootSE  bootMed
1 9999 0.097619 -0.00086911 0.0080038 0.096643
```

```
plot(unif_b.var)
```

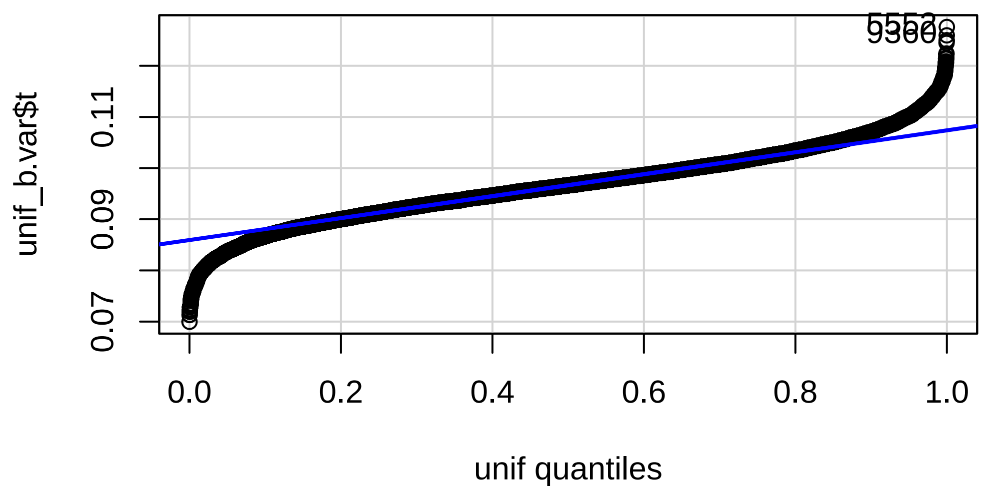



```
qqPlot(unif_b.var$t, distribution="norm",)
```



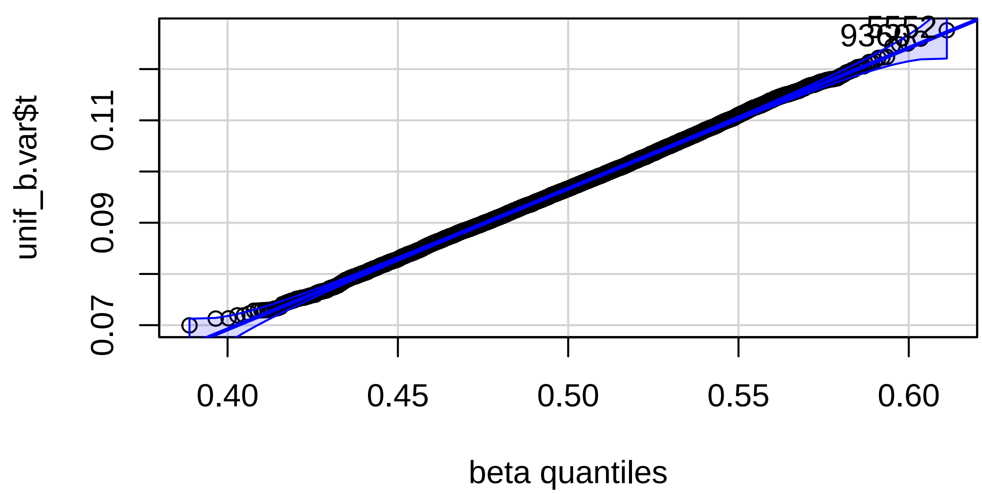

```
[1] 5552 9360
```

```
qqPlot(unif_b.var$t, distribution="unif", min=a, max=b)
```




```
[1] 5552 9360
```

```
qqPlot(unif_b.var$t, distribution="beta", shape1=(3*n-1)/2, shape2=(3*n-1)/2)
```




```
[1] 5552 9360
```

When the data consist of 100 observations randomly drawn from a U(0 , 1), the variance estimator appears to be unbiased as bias=$-8.6911266 \times 10^{-4}$. The standard error is SE=0.0080038. The variance appears to be somewhat normally distributed, but it is not U(0 , 1) distributed. For appropriately chosen shape parameters, α and β, the distribution of the variance appears to be almost beta.