Math 3070 § 1. Treibergs | **Maximum Estimator Example: Simulation of Confidence Intervals** | Name: Example July 9, 2011

The maximum estimator is introduced in Problem 7.59 of Devore, *Probability and Statistics for Engineering and the Sciences, 8th ed.,* Brooks Cole, 2012. Devore applies it to fake data that is supposed to be waiting times for a bus.

Assume that the pdf is uniform depending on the parameter $\theta > 0$,

$$f(x) = \begin{cases} \dfrac{1}{\theta}, & \text{if } 0 \le x \le \theta; \\ 0, & \text{otherwise.} \end{cases}$$

Integrating gives the cdf

$$F(x) = \begin{cases} 0, & \text{if } x \le 0; \\ \dfrac{x}{\theta}, & \text{if } 0 < x \le \theta; \\ 1, & \text{if } \theta < x. \end{cases}$$

An estimator for θ is the maximum of the sample $X_1, X_2, \ldots, X_n \sim \mathrm{Unif}(0, \theta)$, but as we have seen, it is biased. An unbiased estimator is given by

$$\hat{\theta} = \frac{n+1}{n} \max(X_1, X_2, \ldots, X_n).$$

To see this, let us recall the derivation of its pdf. The cdf, by independence

$$\begin{aligned} F(x) &= \mathrm{P}\left(\hat{\theta} \le x\right) \\ &= \mathrm{P}\left(\frac{n+1}{n} \max(X_1, X_2, \ldots, X_n) \le x\right) \\ &= \mathrm{P}\left(\left\{X_1 \le \frac{nx}{n+1}\right\} \cap \left\{X_2 \le \frac{nx}{n+1}\right\} \cap \cdots \cap \left\{X_n \le \frac{nx}{n+1}\right\}\right) \\ &= \mathrm{P}\left(X_1 \le \frac{nx}{n+1}\right) \cdot \mathrm{P}\left(X_2 \le \frac{nx}{n+1}\right) \cdots \mathrm{P}\left(X_n \le \frac{nx}{n+1}\right) \\ &= F\left(\frac{nx}{n+1}\right)^n \\ &= \begin{cases} 0, & \text{if } x \le 0; \\ \left(\dfrac{nx}{(n+1)\theta}\right)^n, & \text{if } 0 < x \le \dfrac{n+1}{n}\theta; \\ 1, & \text{if } \dfrac{n+1}{n}\theta < x. \end{cases} \end{aligned}$$

The pdf for $\hat{\theta}$ is thus

$$f_{\hat{\theta}}(x) = \frac{d}{dx} F(x) = \begin{cases} \dfrac{n^{n+1} x^{n-1}}{(n+1)^n \theta^n}, & \text{if } 0 \le x \le \dfrac{n+1}{n}\theta; \\ 0, & \text{otherwise.} \end{cases}$$

Let $0 < \alpha < 1$. Let us compute the probability

$$\begin{aligned}
\mathrm{P}\left(\theta\left(\frac{\alpha}{2}\right)^{1/n} \le \max(X_1, X_2, \ldots, X_n) \le \theta\left(1-\frac{\alpha}{2}\right)^{1/n}\right) \\
&= \mathrm{P}\left(\theta\left(\frac{\alpha}{2}\right)^{1/n} \le \frac{n\hat\theta}{n+1} \le \theta\left(1-\frac{\alpha}{2}\right)^{1/n}\right) \\
&= F\left(\theta\left(\frac{\alpha}{2}\right)^{1/n}\right) - F\left(\theta\left(1-\frac{\alpha}{2}\right)^{1/n}\right) \\
&= \left(\frac{\alpha}{2}\right) - \left(1-\frac{\alpha}{2}\right) \\
&= 1-\alpha.
\end{aligned}$$

Similarly

$$\begin{aligned}
\mathrm{P}\left(\theta\alpha^{1/n} \le \frac{n\hat\theta}{n+1} = \max(X_1, X_2, \ldots, X_n) \le \theta\right) &= F(\theta) - F\left(\theta\alpha^{1/n}\right) \\
&= 1-\alpha.
\end{aligned}$$

These two probability statements become confidence intervals by solving the inequalities for θ. With $1-\alpha$ confidence, we obtain

$$\frac{\max(X_1, X_2, \ldots, X_n)}{\left(1-\frac{\alpha}{2}\right)^{1/n}} \le \theta \le \left(\frac{2}{\alpha}\right)^{1/n} \max(X_1, X_2, \ldots, X_n) \tag{1}$$

$$\max(X_1, X_2, \ldots, X_n) \le \theta \le \frac{\max(X_1, X_2, \ldots, X_n)}{\alpha^{1/n}} \tag{2}$$

The gap in inequality (1) exceeds that in (2) because $2^{1/n} > 1$ and $2-\alpha > 1$ imply

$$\frac{1}{\alpha^{1/n}} - 1 < 2^{1/n}\left[\frac{1}{\alpha^{1/n}} - \frac{1}{(2-\alpha)^{1/n}}\right] = \left(\frac{2}{\alpha}\right)^{1/n} - \frac{1}{\left(1-\frac{\alpha}{2}\right)^{1/n}}$$

We plot both widths, the left vs. right sides of the inequality.

We check that the estimator $\hat\theta$ is unbiased. Note that for sample from $\mathrm{Unif}(0,\theta)$, the maximum of all X_i is θ so that the statistic ranges in $0 \le \hat\theta \le \frac{n+1}{n}\theta$. Thus

$$\begin{aligned}
\mathrm{E}(\hat\theta) &= \int_0^{\frac{n+1}{n}\theta} u\, f_{\hat\theta}(u)\, du \\
&= \int_0^{\frac{n+1}{n}\theta} \frac{n^{n+1}u^n}{(n+1)^n\theta^n}\, du \\
&= \left[\frac{n^{n+1}u^{n+1}}{(n+1)^{n+1}\theta^n}\right]_0^{\frac{n+1}{n}\theta} \\
&= \theta.
\end{aligned}$$

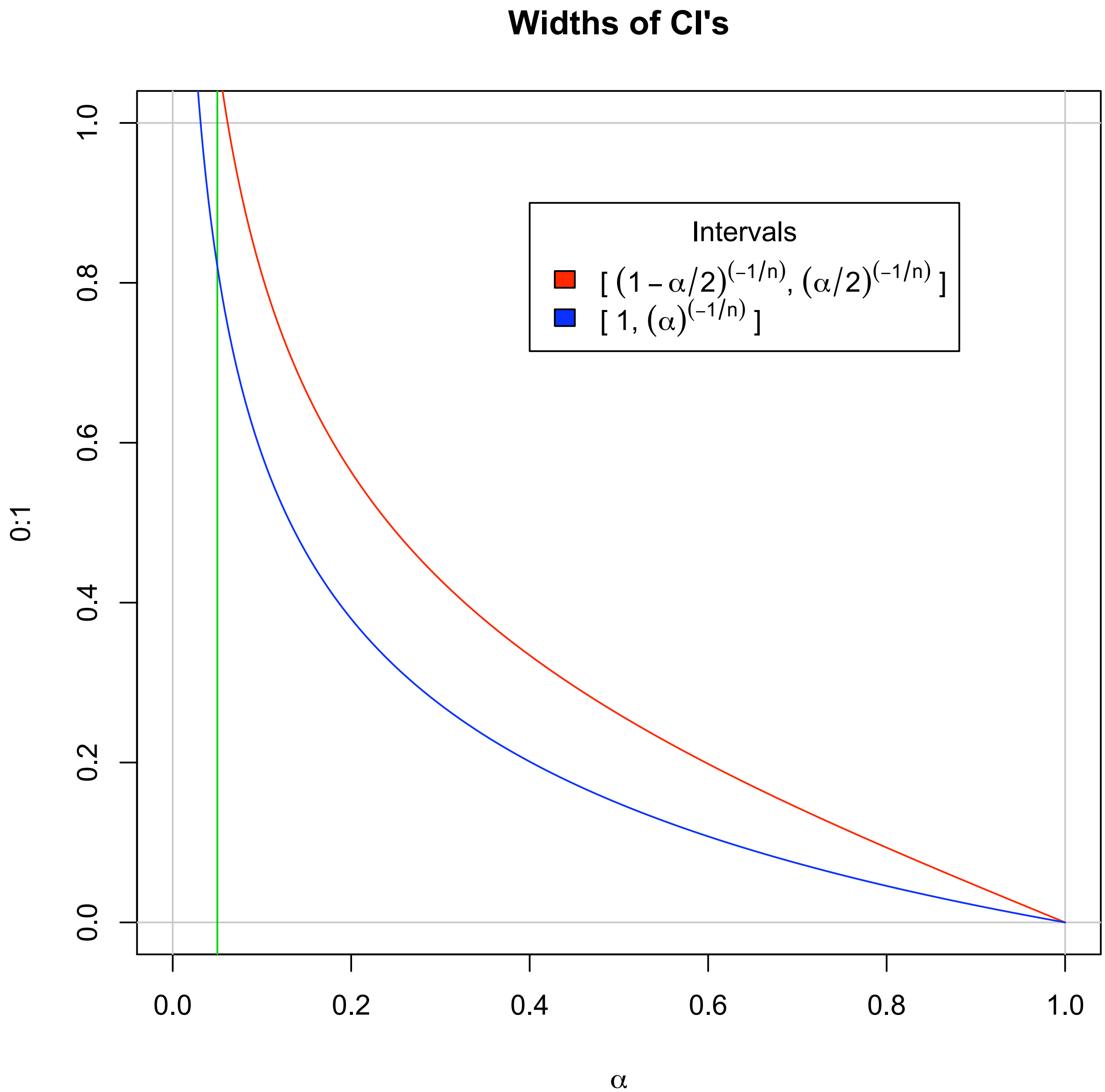
We simulate $B = 10,000$ samples of size $n = 5$ taken from $\mathrm{Unif}(0,1)$. We plot a histogram of the resulting $\hat\theta$'s and superimpose the densith $f_{\hat\theta}$.

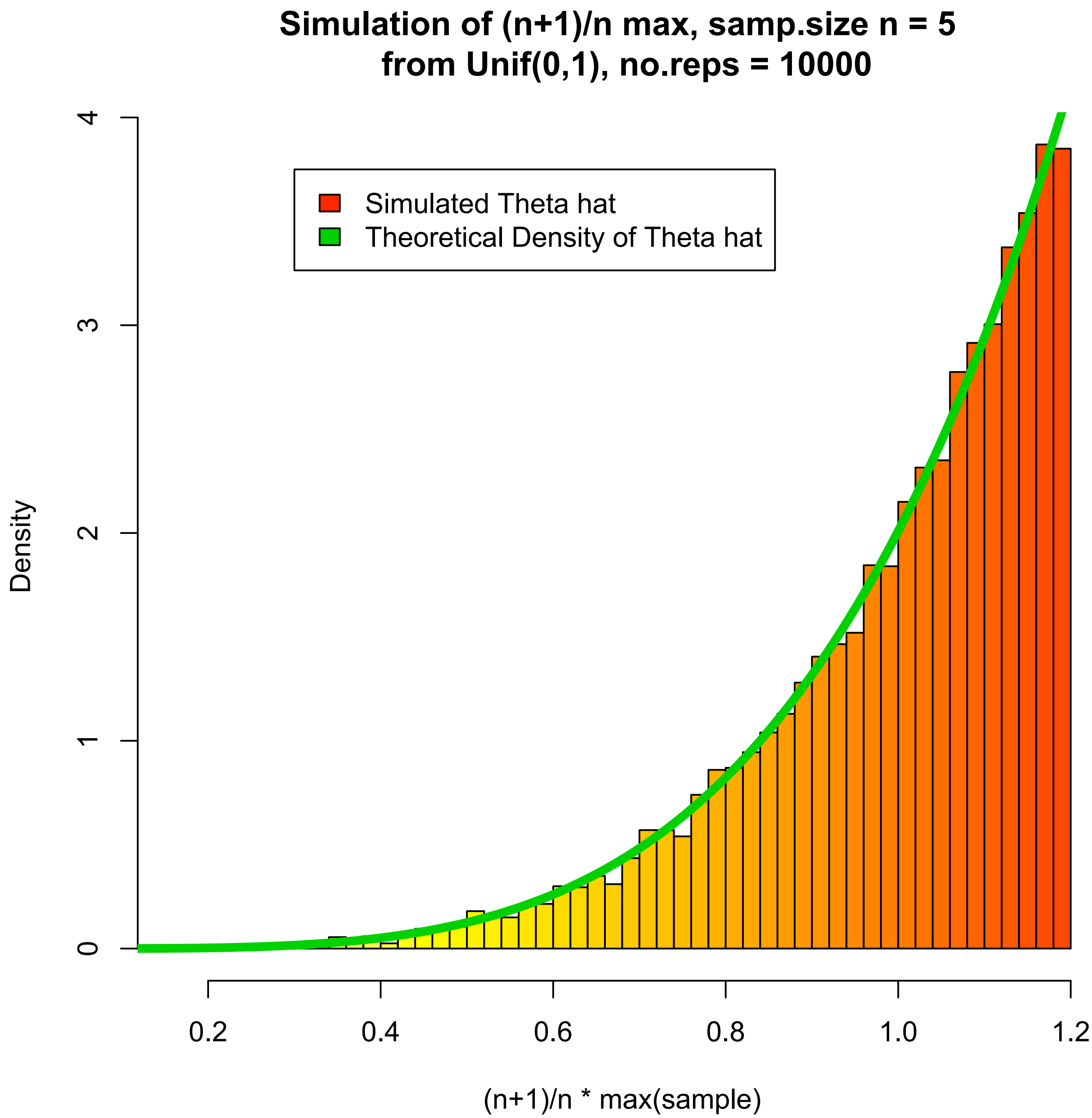
We also simulate $B = 100$ samples of size $n = 5$ taken from $\mathrm{Unif}(0,1)$ and plot the corresponding $\alpha = .05$ confidence intervals using (2). It turned out that six of these did not capture the $\theta = 1$, which is about the expected 5%.

R Session:

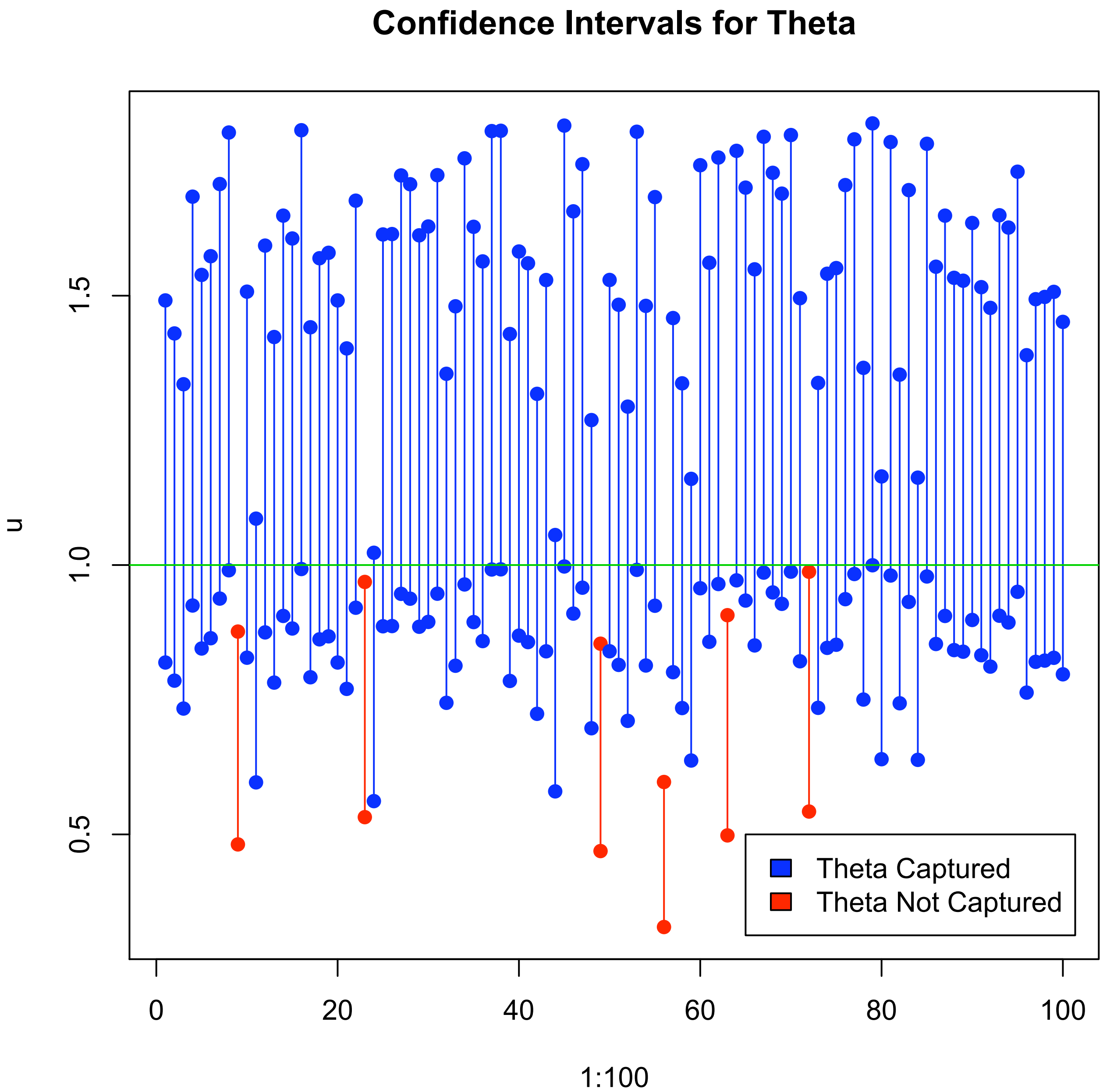
```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]

[Workspace restored from /Users/andrejstreibergs/.RData]

> ################# COMPUTE THE TWO CI'S FOR THETA ############################
>
> # two ci for th of dist unif(0,th)
> n <- 5
> alpha <- .05
> l1 <- function(y){max(y)*(1-alpha/2)^(-1/n)}
> u1 <- function(y){max(y)*(alpha/2)^(-1/n)}
> l2 <- function(y){max(y)}
> u2 <- function(y){max(y)*(alpha)^(-1/n)}
>
> ################# ENDTER DEVORE'S FAKE TIMES FROM EX. 7.59 ##################
> x <- scan()
1: 4.2  3.5 1.7 1.2 2.4
6:
Read 5 items
> # First CI and its width.
> c(l1(x),u1(x));u1(x)-l1(x)
[1] 4.221321 8.783372
[1] 4.562051
> # Second CI and its width.
> c(l2(x),u2(x));u2(x)-l2(x)
[1] 4.20000 7.64637
[1] 3.446370
```

```
> ################## PLOT TWO WIDTHS TO COMPARE #############################
> # Interval width as fn of alpha
> w1 <- function(a){(a/2)^(-1/n)-(1-a/2)^(-1/n)}
> w2 <- function(a){(a)^(-1/n)-1}
> xs <- seq(0,1,1/377)
> plot(0:1, xlim = 0:1, type = "n", ylim = 0:1, main = "Widths of CI's",
+ xlab = expression(alpha))
> abline(h = 0:1, col=8); abline(v = 0:1, col=8); abline(v = alpha, col = 3)
> lines(xs,w1(xs),col=2)
> lines(xs,w2(xs),col=4)
> legend(.4, .9, legend = c(expression(paste("[ ", (1-alpha/2)^(-1/n), ", ",
+ (alpha/2)^(-1/n), " ]")), expression(paste("[ 1, ", (alpha)^(-1/n) ," ]" ))),
+ fill = c(2,4), bg="white", title = "Intervals")
> # M3074MaxUnifCI1.pdf

> ################## SIMULATE SAMPLING DISTRIBUTION OF THETA HAT ############
>
> B <- 10000
> n <- 5
> c <- (n+1)/n
> v <- replicate(B, c*max(runif(n, 0, 1)))
> hist(v,breaks=40, col=rev(heat.colors(50)))
> hist(v,breaks=40, col=rev(heat.colors(60)))
> summary(v); sd(v)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1720  0.9111  1.0460  1.0010  1.1320  1.2000
[1] 0.1685048

> # Plot the histogram.
> hist(v, breaks = 40, col = rev(heat.colors(60)), xlab="(n+1)/n * max(sample)",
+ main = paste("Simulation of (n+1)/n max, samp.size n =", n,
+ "\n from Unif(0,1), no.reps =", B), freq = F)
> lines(xs, n/c*(xs/c)^(n-1), col = 3, lwd = 5)
> legend(.3, 3.75, legend = c("Simulated Theta hat",
+ "Theoretical Density of Theta hat"), fill = c(2, 3))
> # M3074MaxUnifCI2.pdf
```


Widths of CI's
Intervals
[$(1-\alpha/2)^{(-1/n)}$, $(\alpha/2)^{(-1/n)}$]
[1, $(\alpha)^{(-1/n)}$]
0:1
α
0.0
0.2
0.4
0.6
0.8
1.0


Simulation of (n+1)/n max, samp.size n = 5
from Unif(0,1), no.reps = 10000
Simulated Theta hat
Theoretical Density of Theta hat
Density
0
1
2
3
4
0.2
0.4
0.6
0.8
1.0
1.2
(n+1)/n * max(sample)

```
> ################## SIMULATE 100 CI'S FOR THETA ##############################
>
> u <- numeric(100);l <- numeric(100)
> # Fill color vector with blues.
> cl <- rep(4,100)
> # Generate CI's. Test if  theta  is captured.
> for(i in 1:100)
+               {
+                sam <- runif(n,0,1)
+                l[i] <- l2(sam)
+                u[i] <- u2(sam);
+                if(l[i] > 1 | u[i] < 1)
+                                      {
+                                       cl[i]<-2
+                                      }
+               }
> mx <- max(u)
> mi <- min(l)
> plot(1:100, u, ylim = c(mi, mx), type = "n",
+ main = "Confidence Intervals for Theta")
> for(i in 1:100)
+               {
+                points(c(i, i), c(l[i], u[i]), col = cl[i], pch = 19)
+                lines(c(i, i), c(l[i], u[i]), col = cl[i])
+               }
> abline(h = 1, col = 3)
> legend(65, .5, legend = c("Theta Captured", "Theta Not Captured"),
+ fill = c(4, 2), bg = "white")
> # M3074MaxUnifCI3.pdf
```


Confidence Intervals for Theta
u
1.5
1.0
0.5
0
20
40
60
80
100
1:100
Theta Captured
Theta Not Captured