Homework 6

Jason Maddox

5/2/2019

1. **Nonparametric Bootstrap CIs.**

(a) Bootstrap interval.

```
boot_pct_ci <- function(x, conf = .95, B = 1e3) {
  boot_means <- rerun(B, {
    sample(x, length(x), replace = TRUE) %>%
    mean()
  })
  boot_means %>%
    flatten_dbl() %>%
    quantile(c((1 - conf) / 2, (1 + conf) / 2), names = FALSE)
}
set.seed(1)
(data <- rnorm(20, mean = 5, sd = 3) %>% round(2))
```

```
##  [1]  3.12  5.55  2.49  9.79  5.99  2.54  6.46  7.21  6.73  4.08  9.54
## [12]  6.17  3.14 -1.64  8.37  4.87  4.95  7.83  7.46  6.78
```

```
data %>% boot_pct_ci() %>% round(2)
```

```
## [1] 4.36 6.71
```

(b) t-interval.

```
t_ci <- function(x, conf = .95) {
  c(mean(x) + qt((1 - conf)/2, length(x) - 1)*sd(x)/sqrt(length(x)),
    mean(x) - qt((1 - conf)/2, length(x) - 1)*sd(x)/sqrt(length(x)))
}
data %>% t_ci() %>% round(2)
```

```
## [1] 4.29 6.85
```

```
data %>% t.test() %>% pluck("conf.int") %>% round(2)
```

```
## [1] 4.29 6.85
## attr(,"conf.level")
## [1] 0.95
```

(c) Bias corrected and adjusted bootstrap interval.

```
boot_bca_ci <- function(x, conf = .95, B = 1e4) {
  bootstrap(x, mean, R = B) %>%
    CI.bca(probs = c((1 - conf)/2, (1 + conf)/2), names = FALSE) %>% as.vector()
}
data %>% boot_bca_ci() %>% round(2)
```

```
## [1] 4.11 6.71
```

(d) Simulation study.

```
cover <- function(method) {
  ci <- rweibull(250, .2, 5) %>% method()
  ci[1] < 600 & ci[2] > 600
```

```
}
covs <- rerun(10000, {
  list(boot_pct_ci, t_ci, boot_bca_ci) %>%
    map(cover) %>%
    flatten_lgl()
}) %>% ldply()
covs %>% summarize(nonpar_btstrp = mean(V1), t = mean(V2), BCa = mean(V3))
```

```
##   nonpar_btstrp      t    BCa
## 1         0.693 0.6759 0.7822
```

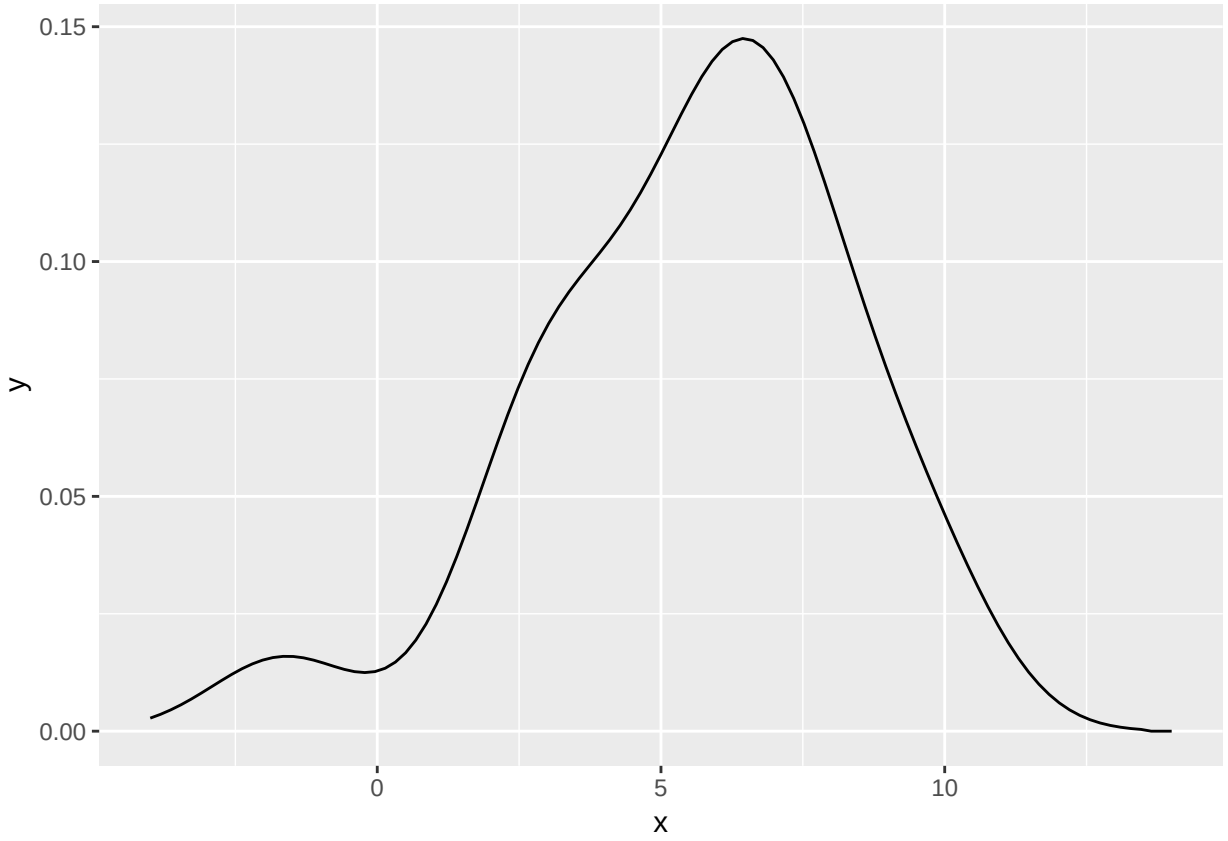
2. **Parametric and Smooth Bootstrap.**

(a) Kernel density estimate.

```
kdefun <- function(x) {
  kde <- (x %>% density())[c('x', 'y')] %>% transpose() %>% ldply(as_tibble)
  approxfun(kde$x, kde$y, yleft = 0, yright = 0)
}
```

(b) Plot kernel density estimate.

```
ggplot(tibble(x = c(-4, 14)), aes(x)) +
  stat_function(fun = kdefun(data))
```



(c) Integrate KDE.

```
integrate(kdefun(data), -Inf, Inf)
```

```
## 1.000872 with absolute error < 6.9e-05
```

(d) Empirical CDF.

```
ecdf2 <- function(x) {
  function(vals) {
    vals %>% map_dbl(~ mean(.x >= x))
  }
}
s <- seq(.5, 2.5, .5)
ecdf(1:2)(s)
```

```
## [1] 0.0 0.5 0.5 1.0 1.0
```

```
ecdf2(1:2)(s)
```

```
## [1] 0.0 0.5 0.5 1.0 1.0
```

(e) Smoothed empirical CDF.

```
secdf <- function(x) {
  function(vals) {
    vals %>% map_dbl(~ integrate(kdefun(x), lower = -Inf, upper = .x)$value)
  }
}
pnorm(-3:3) %>% round(4)
```

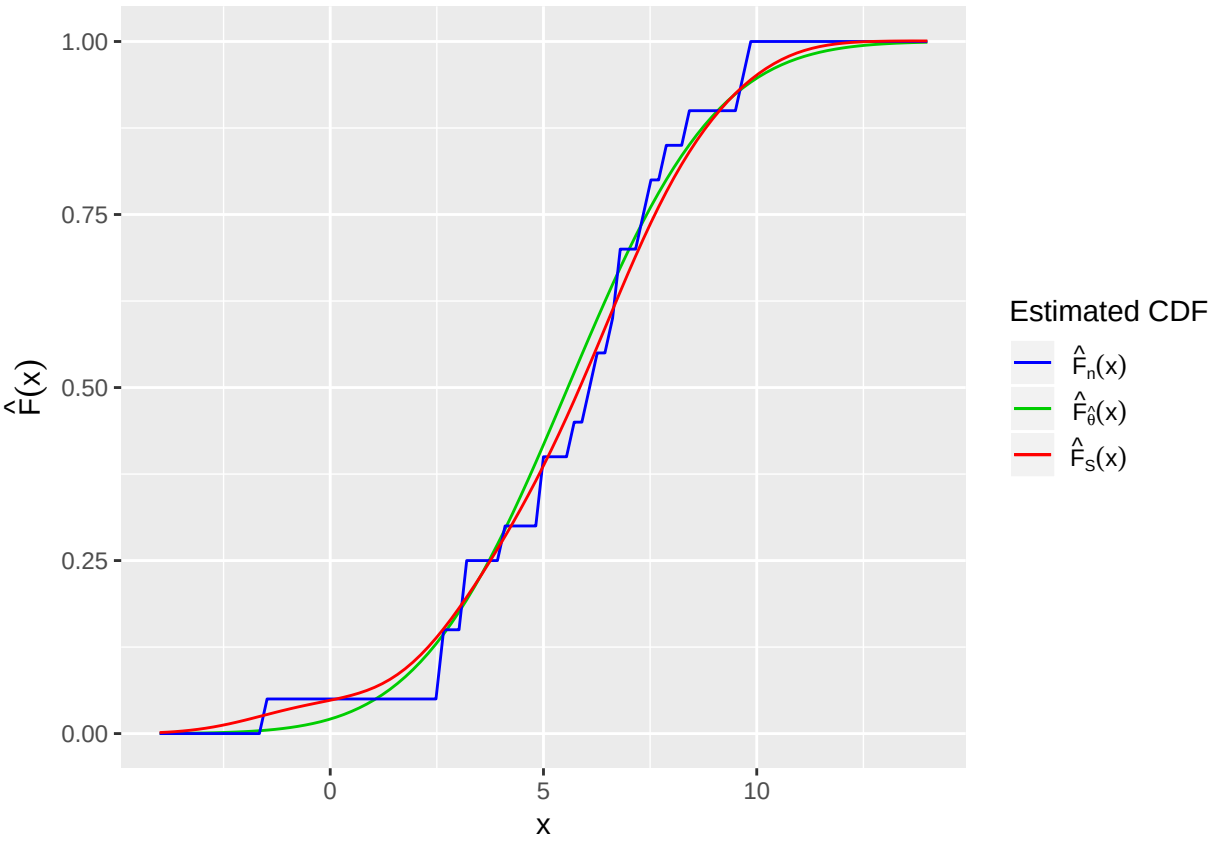
```
## [1] 0.0013 0.0228 0.1587 0.5000 0.8413 0.9772 0.9987
```

```
secdf(rnorm(1e6))(-3:3) %>% round(4)
```

```
## [1] 0.0014 0.0229 0.1596 0.5008 0.8419 0.9781 0.9996
```

(f) Graph of CDFs.

```
ggplot(tibble(x = c(-4, 14)), aes(x)) +
  stat_function(fun = pnorm,
                args = list(mean = mean(data), sd = sd(data)),
                aes(color = "green3")) +
  stat_function(fun = ecdf2(data), aes(color = "blue")) +
  stat_function(fun = secdf(data), aes(color = "red")) +
  scale_color_manual(name = "Estimated CDF",
                     values = c("red" = "red",
                                "blue" = "blue",
                                "green3" = "green3"),
                     labels = c(expression(hat(F)[n](x)),
                                expression(hat(F)[hat(theta)](x)),
                                expression(hat(F)[S](x)))) +
  ylab(expression(hat(F)(x)))
```



(g) Parametric bootstrap.

```
boot_param_ci <- function(x, conf = .95, B = 1e3) {
  boot_means <- rerun(B, {
    rnorm(length(x), mean(x), sd(x)) %>%
    mean()
  })
  boot_means %>%
    flatten_dbl() %>%
    quantile(c((1 - conf) / 2, (1 + conf) / 2), names = FALSE)
}
data %>% boot_param_ci() %>% round(2)
```

```
## [1] 4.36 6.80
```

(h) Smooth bootstrap.

```
boot_smooth_ci <- function(x, conf = .95, B = 1e3) {
  boot_means <- rerun(B, {
    (sample(x, length(x), replace = TRUE) + rnorm(length(x), 0, bw.nrd0(x))) %>%
    mean()
  })
  boot_means %>%
    flatten_dbl() %>%
    quantile(c((1 - conf) / 2, (1 + conf) / 2), names = FALSE)
}
data %>% boot_smooth_ci() %>% round(2)
```

```
## [1] 4.28 6.82
```

(i) Simulation study.

```
covs <- rerun(10000, {
  list(boot_pct_ci,
       t_ci,
       boot_bca_ci,
       boot_param_ci,
       boot_smooth_ci) %>%
    map(cover) %>%
    flatten_lgl()
}) %>% ldply()
covs %>% summarize(nonpar_btstrp = mean(V1),
                   t = mean(V2),
                   BCa = mean(V3),
                   par_btstrp = mean(V4),
                   smooth_btstrp = mean(V5))
```

```
##   nonpar_btstrp      t    BCa par_btstrp smooth_btstrp
## 1        0.7068 0.6733 0.7763     0.6672        0.7005
```