Stat 641 Fall 2021

Solutions for Assignment 3

P1. (10 points) Let Y have a double exponentiall distribution.

(a.) The quantile function

$$Q(u) = \begin{cases} \theta + \beta\ log(2u) & \text{for} \quad u \le .5 \\ \theta - \beta\ log(2(1-u)) & \text{for} \quad u \ge .5 \end{cases}$$

(b.) The survival function is given by

$$S(y) = P(Y > y) = 1 - F(y) \ \Rightarrow \qquad S(y) = \begin{cases} 1 - \frac{1}{2} e^{-\left(\frac{\theta - y}{\beta}\right)} & \text{for} \quad y < \theta \\ \frac{1}{2} e^{-\left(\frac{y-\theta}{\beta}\right)} & \text{for} \quad y \ge \theta \end{cases}$$

(c.) The hazard function is given by

$$h(y) = \frac{f(y)}{S(y)} \ \Rightarrow \qquad h(y) = \begin{cases} \dfrac{\frac{1}{\beta} e^{\left(\frac{y-\theta}{\beta}\right)}}{2 - e^{\left(\frac{y-\theta}{\beta}\right)}} & \text{for} \quad y \le \theta \\ \frac{1}{\beta} & \text{for} \quad y > \theta \end{cases}$$

P2. (10 points) $n = 44 \quad \Rightarrow \quad \widehat{Q}(u) = Y_{(43u+1)} \ \Rightarrow$

- $\widehat{Q}(.25) = Y_{(11.75)} = .25Y_{(11)} + .75Y_{(12)} = .25(3.24) + .75(3.39) = 3.35$
- $\widehat{Q}(.5) = Y_{(22.5)} = .5Y_{(22)} + .5Y_{(23)} = .5(7.89) + .5(7.97) = 7.93$
- $\widehat{Q}(.75) = Y_{(33.25)} = .75Y_{(33)} + .25Y_{(34)} = .75(16.00) + .25(18.61) = 16.65$

P3. (24 points) Using the R code:

```
y = c(0.94,    1.26,    1.44,    1.49,    1.63,    1.80,    2.00,    2.00,    2.56,
    2.58,    3.24,    3.39,    3.53,    3.77,    4.36,    4.41,    4.60,    4.67,
    5.39,    6.25,    7.02,    7.89,    7.97,    8.00,    8.28,    8.83,    8.91,
    8.96,    9.92,   11.36,   12.15 ,  14.40,   16.00,   18.61,   18.75,   19.05,
   21.00,   21.41,   23.27,   24.71,   25.00,   28.75,   30.23,   35.45   )
h=3
n=length(y)
deni <- function(x){
  (1/sqrt(2*pi))*exp(-((x-y)/h)^2/2)/(n*h)
}
f3 = sum(sapply(3,deni))
f16 = sum(sapply(16,deni))
f16i = sapply(16,deni)
min = min(f16i)
imin = which(f16i==min)
ymin=y[imin]
max = max(f16i)
imax=which(f16i==max)
ymax=y[imax]
```

(a.) The value for $\widehat{f}(3)$ is f3 = 0.059703 and for $\widehat{f}(16)$ is f16 = 0.01669353

(b.) Using a relative frequency histogram with a bin width of 5, with $n_j = \#Y_i$'s in $[0.94 + 5(j-1), 0.94 + 5j)$, we have $n_1 = 19,\ n_2 = 10,\ n_3 = 3,\ n_4 = 4,\ n_5 = 5,\ n_6 = 2,\ n_7 = 1$.

Therefore, the estimates are $\widehat{f}(3) = 19/(44 \times 5) = 0.08636364$ and for $\widehat{f}(16) = 4/(44 \times 5) = 0.01818182$. A fairly close agreement between the estimates obtained by the two methods.

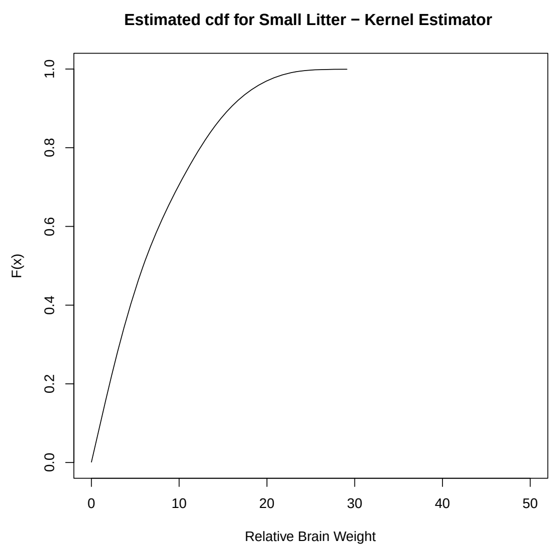
(c.) The data value provides the smallest contribution to the estimator at y=16, $\widehat{f}(16)$ is the data value furthest from 16, which is y = 35.45 with a contribution of 2.253479e-12 to $\widehat{f}(16)$=0.01669353. This is obtained by computing by hand:

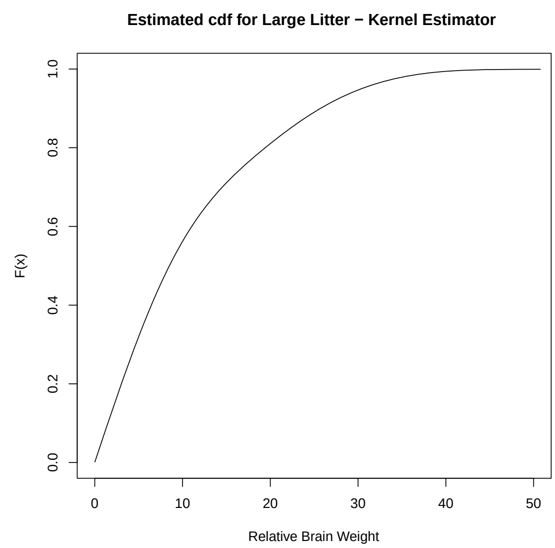
$$\frac{1}{nh}K\left(\frac{y-Y_i}{h}\right) = \frac{1}{44*3}K\left(\frac{16-35.45}{3}\right) = \left(\frac{1}{132}\right)\left(\frac{1}{2*\pi}\right)e^{-\left(\frac{16-35.45}{3}\right)^2/2} = 2.253479e^{-12}$$

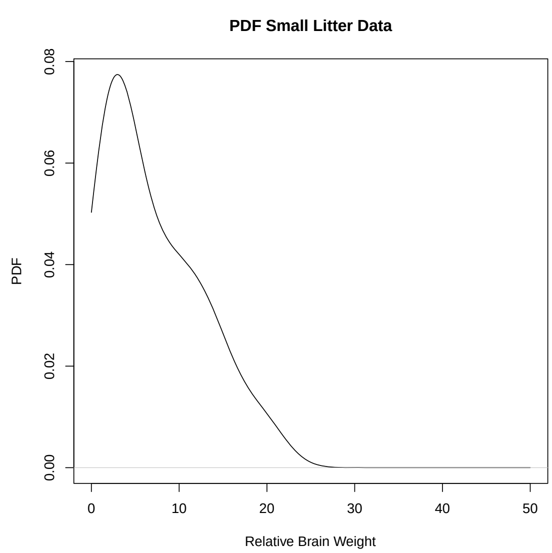
(d.) The data value provides the largest contribution to the estimator at y=16, $\widehat{f}(16)$ is the data value closest to 16, which is y = 16 with a contribution of 0.00302229 to $\widehat{f}(16)$=0.01669353. This is obtained by computing by hand:

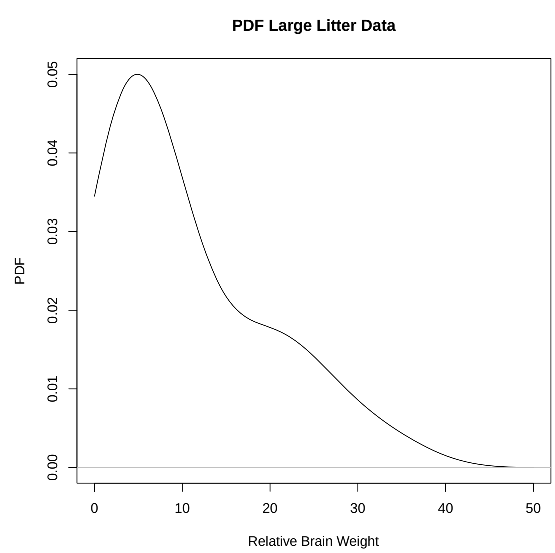
$$\frac{1}{nh}K\left(\frac{y-Y_i}{h}\right) = \frac{1}{44*3}K\left(\frac{16-16}{3}\right) = \left(\frac{1}{132}\right)\left(\frac{1}{2*\pi}\right)e^{-\left(\frac{16-16}{3}\right)^2/2} = .00302229$$

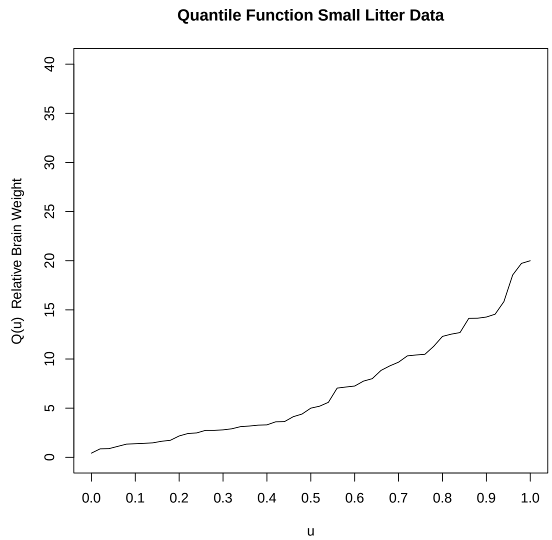
P4. (28 points) (a.) Plots of pdfs (kernel density estimator), edf (smoothed), and quantile (smoothed):

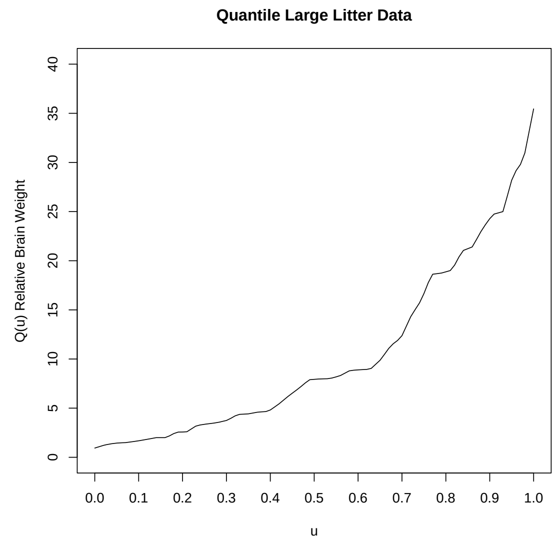












(a)

See code at end of document.

(b.) Small Litter: Relative brain weights are somewhat right skewed which indicates that a few species of mammals with small average litters have large brains relative to their body weights.

Large Litter: Relative brain weights are highly right skewed which indicates that sizeable proportion of the species of mammals with large average litters have large brains relative to their body weights.

(c.) Based on the graphs, I would conclude that there is a positive relation ship between average litter size and relative brain weights. However, it would be more informative to have the actual litter sizes associated with each species to draw a more concrete conclusion.

P5. (28 points) Multiple Choice Questions:

1. **E** Given any one of the four functions then you can derive the other three from the given function
2. **D** See page 24 in Handout 4
3. **B** See page 24 in Handout 4
4. **D** See pages 30 & 32 in Handout 4
5. **B** See page 50 in Handout 4
6. **B** See page 37 in Handout 4
7. **D** See pages 16 & 17 in Handout 5
8. **C or D** See page 13 & 14 in Handout 5
9. **C** See page 23 in Handout 5
10. **A or D** See page 27 in Handout 5
11. **E** See page 14 in Handout 5
12. **E** See pages 25-27 in Handout 5

A. is false because σ does not exist for Cauchy which is symmetric whereas both SIQR and MAD exist and are equal

B. is false because MAD is nearly always preferred to SIQR

C. is false because for a normal distribution MAD=SIQR

13. **B** See page 33 in Handout 5:

$\theta = 22.3,\ \rho = .6,\ \sigma_e^2 = 2.8\ \Rightarrow \mu_X = \frac{\theta}{1-\rho} = \frac{22.3}{1-.6} = 55.75$

$\sigma_X^2 = \frac{\sigma_e^2}{1-\rho^2} = \frac{2.8}{1-.36} = 4.375$

```
##
## (2)
##

dta <- read.csv("Assign3_BrainSize.csv")

y <- dta[, 2]
y <- y[!is.na(y)]
n <- length(y)

y_s <- sort(y)

## 0.25: (n - 1) * 0.25 + 1 = 11.75
(n - 1) * 0.25 + 1
y_s[11] + 0.75 * (y_s[12] - y_s[11])

## 0.5: (n - 1) * 0.5 + 1 = 22.5
(n - 1) * 0.5 + 1
y_s[22] + 0.5 * (y_s[23] - y_s[22])

## 0.75: (n - 1) * 0.5 + 1 = 33.25
(n - 1) * 0.75 + 1
y_s[33] + 0.25 * (y_s[34] - y_s[33])

##
## (3)
##

dd <- density(y_s)
plot(dd, type = "l")

h <- 3

## (a)
K_u <- function(u) {
  return(dnorm(u))
}

## f(3)
sum(K_u((3 - y_s) / h)) / (n * h)

## f(16)
sum(K_u((16 - y_s) / h)) / (n * h)

## (b)
brks <- y_s[1] + 5 * (0:7)
n_j <- hist(y_s, prob = FALSE, breaks = brks)$counts
R_j <- n_j / n
f_hat_j <- R_j / 5

hist(y_s, prob = TRUE, breaks = brks)$density

## (c)
kk <- K_u((16 - y) / h) / (n * h)
y[which.min(kk)]

## (d)
```

```
y[which.max(kk)]

##
## (4)
##

y_small <- dta[, 1]
y_large <- dta[, 2]
y_large <- y_large[!is.na(y_large)]

## (a)

## PDFs
par(mfrow = c(1, 2))
plot(density(y_small), xlab = "y", ylab = "f", main = "Small Litters", xlim = c(-10, 50),
  cex.axis = 0.75)
plot(density(y_large), xlab = "y", ylab = "f", main = "Large Litters", xlim = c(-10, 50),
  cex.axis = 0.75)

## EDFs
qq_small <- quantile(y_small, probs <- seq(0, 1, by = 0.01))
qq_large <- quantile(y_large, probs)

plot(qq_small, probs, type = "s", xlab = "y", ylab = "F", main = "Small Litters",
  xlim = c(0, 35), cex.axis = 0.75)
plot(qq_large, probs, type = "s", xlab = "y", ylab = "F", main = "Large Litters",
  xlim = c(0, 35), cex.axis = 0.75)

## Quantile functions
plot(probs, qq_small, type = "s", xlab = "Q(u)", ylab = "u", main = "Small Litters",
  ylim = c(0, 35), cex.axis = 0.75)
plot(probs, qq_large, type = "s", xlab = "Q(u)", ylab = "u", main = "Large Litters",
  ylim = c(0, 35), cex.axis = 0.75)
```