Stat 641

Solutions for Assignment 3

I. (15 points) Let Y have a 3-parameter Weibull distribution. (a.) The survival function is given by

$$S(y) = P(Y > y) = 1 - F(y) \;\Rightarrow \qquad S(y) = e^{-\left(\frac{y-\theta}{\alpha}\right)^{\gamma}} I(y \geq \theta) + I(y < \theta)$$

(b.) The hazard function is given by

$$h(y) = \frac{f(y)}{S(y)} \;\Rightarrow \qquad h(y) = \begin{cases} \frac{\gamma}{\alpha^{\gamma}}(y-\theta)^{\gamma-1} & \text{for} \quad y \geq \theta \\ 0 & \text{for} \quad y < \theta \end{cases}$$

II. (15 points) $n = 51 \quad \Rightarrow \quad \widehat{Q}(u) = Y_{(50u+1)} \;\Rightarrow$

- $\widehat{Q}(.25) = Y_{(13.5)} = .5Y_{(13)} + .5Y_{(14)} = .5(2.48) + .5(2.74) = 2.61$
- $\widehat{Q}(.5) = Y_{(26)} = 5.00$
- $\widehat{Q}(.75) = Y_{(38.5)} = .5Y_{(38)} + .5Y_{(39)} = .5(10.41) + .5(10.48) = 10.445$

III. Using the R code:

```
y = c(0.42,    0.86,    0.88,    1.11,    1.34,    1.38 ,   1.42,    1.47,    1.63,
      1.73,    2.17,    2.42,    2.48,    2.74,    2.74,    2.79,    2.90,    3.12,
      3.18,    3.27,    3.30,    3.61 ,   3.63,    4.13 ,   4.40,    5.00,    5.20,
      5.59,    7.04,    7.15,    7.25,    7.75,    8.00,    8.84,    9.30 ,   9.68,
     10.32,   10.41,   10.48,   11.29,   12.30,   12.53,   12.69,   14.14,   14.15,
     14.27,   14.56,   15.84,   18.55,   19.73,   20.00)
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.) (5 points) The value for $\widehat{f}(3)$ is f3 $= 0.07176$ and for $\widehat{f}(16)$ is f16 $= 0.02255$

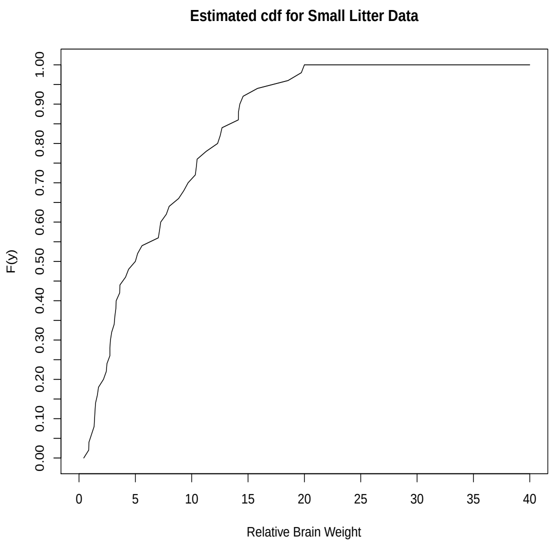
(b.) (5 points) Using a relative frequency histogram with a bin width of 5, with $n_j = \#Y_i$'s in $[0.42 + 5(j-1), 0.42 + 5j)$, we have $n_1 = 27,\; n_2 = 11,\; n_3 = 9,\; n_4 = 4.$

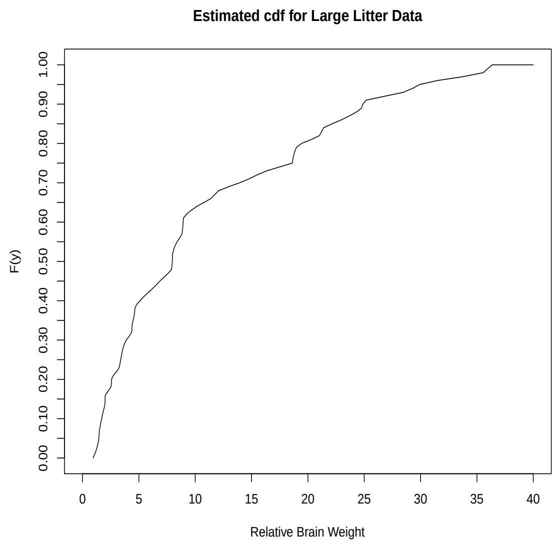
Therefore, the estimates are $\widehat{f}(3) = 27/(51)(5) = 0.10588$ and for $\widehat{f}(16) = 4/(51)(5) = 0.01569$. A very large discrepancy between the estimates obtained by the two methods.

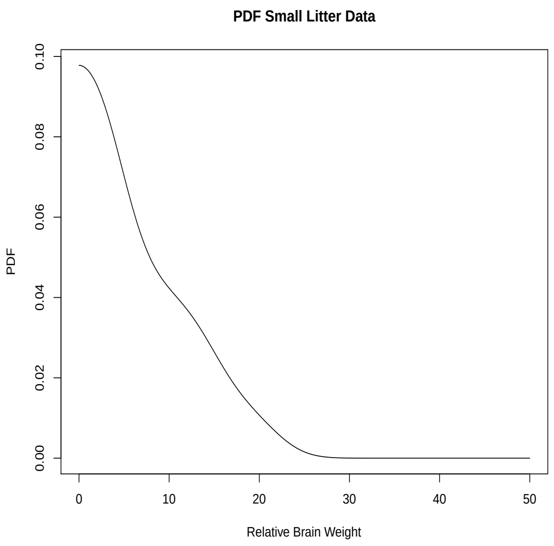
(c.) (5 points) The data value provides the smallest contribution to the estimator at y=16, $\widehat{f}(16)$ is the data value furthest from 16, which is y = 0.42 with a contribution of 3.627464e-09 to $\widehat{f}(16)$=0.02255. This is obtained from "ymin" and "min" in the R program given above.

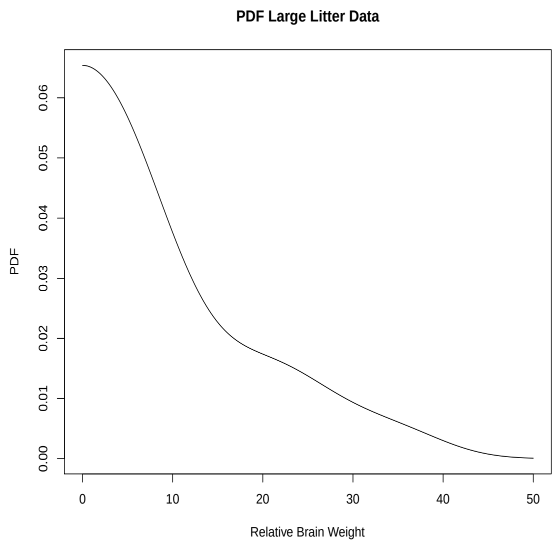
(d.) (5 points) The data value provides the largest contribution to the estimator at y=16, $\widehat{f}(16)$ is the data value closest to 16, which is y = 15.84 with a contribution of 0.00260376 to $\widehat{f}(16)$=0.02255. This is obtained from "ymax" and "max" in the R program given above.

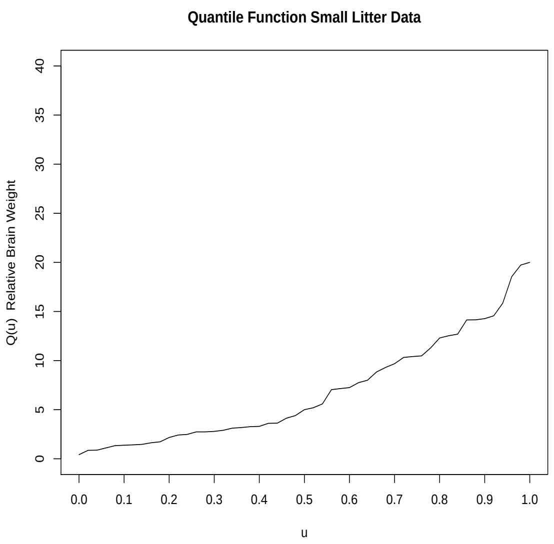
IV. (20 points) (a.) Plots of pdfs (kernel density estimator), edf (smoothed), and quantile (smoothed):

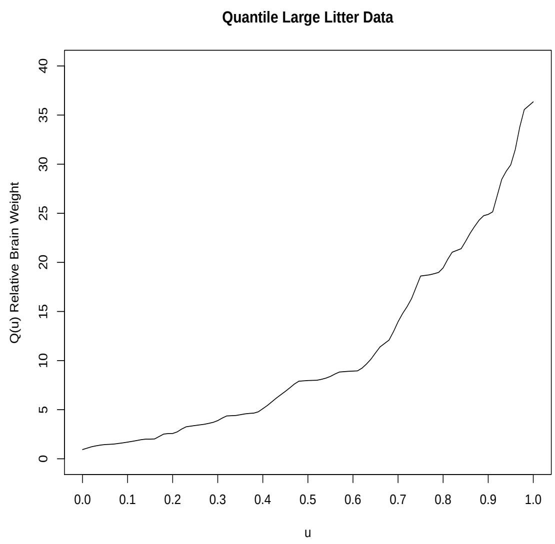












(b.) Small Litter: Relative brain weights are somewhat right skewed which indicates that a few species of mammals with small average litters have large brains relative to their body weights.

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.

V. (30 points) Multiple Choice Questions:

1. **E** Given any one of the four functions then you can derive the other three from the given function. See page 50 in Handout 3.
2. **B** See page 22 in Handout 4
3. **D** See pages 22 & 27 in Handout 4
4. **B** See pages 27 & 28 in Handout 4
5. **B** See pages 15 & 16 in Handout 5
6. **D** See pages 13 & 14 in Handout 5
7. **E** For exponential distribution, Q_1, the first quartile,

$$\int_0^{Q_1} \frac{1}{\beta}\exp(-u/\beta) = 0.25 \Rightarrow 1-\exp(-Q_1/\beta) = 0.25 \Rightarrow Q_1 = -\beta\log(0.75).$$

Similarly, $Q_2 = -\beta\log(0.5)$ and $Q_3 = -\beta\log(0.25)$. Now, let m =MAD= median$|Y - Q_2|$ where $Y \sim$ Exponential(β). Thus,

$$\begin{aligned}
& \mathrm{pr}(|Y-Q_2| \le m) = 0.5 \\
\Rightarrow\quad & \mathrm{pr}\{-m < (Y-Q_2) < m\} = 0.5 \\
\Rightarrow\quad & \mathrm{pr}\{Q_2 - m < Y < Q_2 + m\} = 0.5 \\
\Rightarrow\quad & 1-\exp(-\frac{Q_2+m}{\beta}) - \left\{1-\exp(-\frac{Q_2-m}{\beta})\right\} = 0.5 \\
\Rightarrow\quad & \exp(-\frac{Q_2-m}{\beta}) - \exp(-\frac{Q_2+m}{\beta}) = 0.5 \\
\Rightarrow\quad & \exp(-\frac{Q_2}{\beta})\left\{\exp(m/\beta)-\exp(-m/\beta)\right\} = 0.5 \\
\Rightarrow\quad & \exp\{\frac{\beta}{\beta}\log(0.5)\}\left\{\exp(m/\beta)-\exp(-m/\beta)\right\} = 0.5 \\
\Rightarrow\quad & (0.5)\left\{\exp(m/\beta)-\exp(-m/\beta)\right\} = 0.5 \\
\Rightarrow\quad & \left\{\exp(m/\beta)-\exp(-m/\beta)\right\} = 1.
\end{aligned}$$

8. **B** See page 27 in Handout 5
9. **E**

   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

10. **D** See page 32 in Handout 5:

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

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