STAT4520 HW4

Anton Yang

2024-10-18

Problem 1

Suppose data is generated from the exponential distribution with density

$$f(y) = \lambda e^{-\lambda y}$$

. We can write this as exponential family:

$$f(y|\theta, \phi) = e^{-\lambda y + \log(\lambda)}$$

So, $\theta = -\lambda$, $\phi = 1$, $a(\phi) = \phi = 1$, $b(\theta) = -\log(\lambda)$, and $c(y, \phi) = 0$.

We can make the exponential distribution into Canonical Form, which we let $\eta = -\lambda$.

Therefore, $\eta = -\lambda$, $\phi = 1$, $a(\phi) = \phi = 1$, $b(\eta) = -\log(-\eta)$, and $c(y, \phi) = 0$.

Thus, we can solve $E[X]$ and $Var[X]$ with $b(\theta)$.

$$E[X] = b'(\theta) = \frac{-1}{\eta} = \frac{1}{\lambda}$$

$$Var[X] = b''(\theta) = \frac{1}{\eta^2} = \frac{1}{\lambda^2}$$

Problem 2

```
library(faraway)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(mgcv)
```

```
## Loading required package: nlme
```

```
##
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
## This is mgcv 1.8-40. For overview type 'help("mgcv-package")'.
```

```
data<-chicago

tw_model<-gam(involact ~ age + theft + log(income), family = tw(link = "log"), data = data)
summary(tw_model)
```

```
##
## Family: Tweedie(p=1.152)
## Link function: log
##
## Formula:
## involact ~ age + theft + log(income)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.080266   5.189569   4.640 3.26e-05 ***
## age          0.022475   0.008638   2.602   0.0127 *
## theft       -0.004138   0.005237  -0.790   0.4338
## log(income) -2.834038   0.554128  -5.114 6.95e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.437   Deviance explained = 51.6%
## -REML = 44.163  Scale est. = 0.44757   n = 47
```

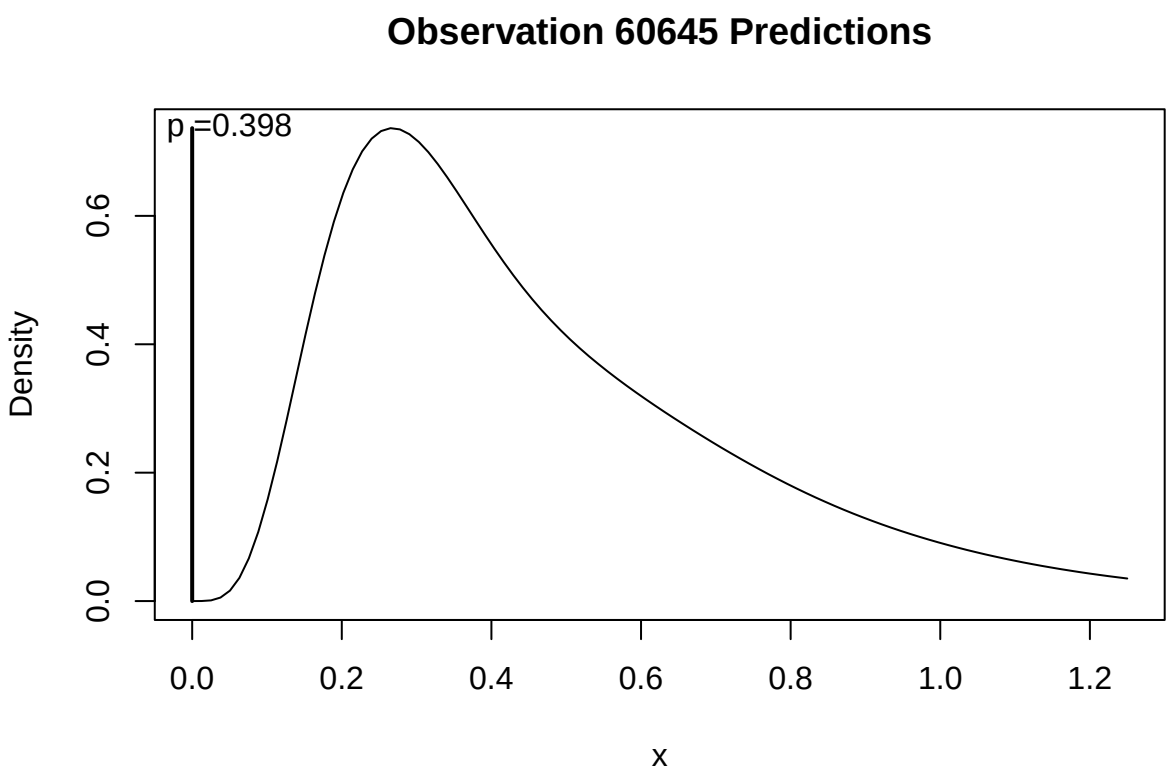
According to the summary, we have a p of 1.152 and $\phi = 0.44757$.

```
xgrid<-seq(1e-10, 1.25, len = 100)
p <- 1.152
phi <- 0.44757
mu <- tw_model$fit[1]
poismean<-mu^(2-p)/(phi * (2-p))
p0<-exp(-poismean)

twden<-exp(ldTweedie(xgrid, mu, p = p, phi = phi)[,1])
data$involact[nrow(data)]
```

```
## [1] 0
```

```r
plot(xgrid, twden*(1-p0), type = "l", xlab = "x", ylab = "Density", main = "Observation 60645 Predictio
dmax<-max(twden * (1-p0))
segments(0, 0, 0, dmax, lwd =  2)
text(0.05, dmax, paste0("p =", signif(p0, 3)))
```

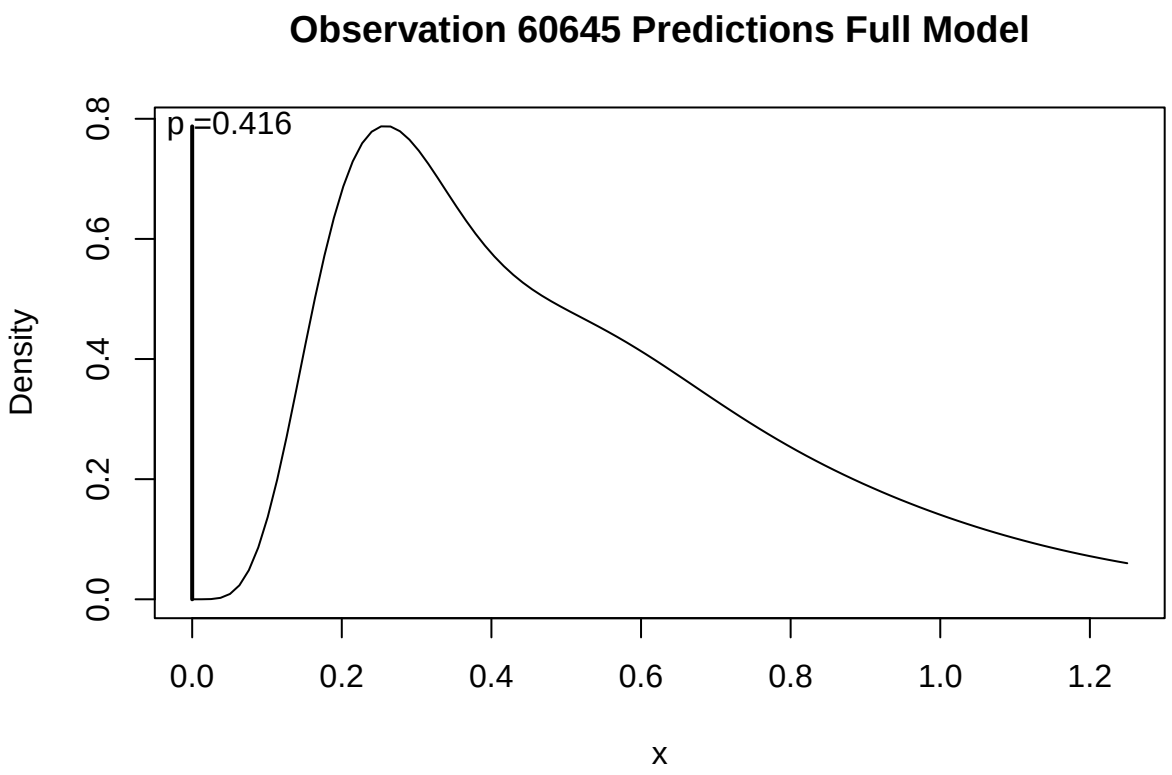



```r
tw_model2<-gam(involact ~ age + theft + log(income) + fire + volact, family = tw(link = "log"), data = c
summary(tw_model2)
```

```
##
## Family: Tweedie(p=1.132)
## Link function: log
##
## Formula:
## involact ~ age + theft + log(income) + fire + volact
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.141517   8.712058   0.016   0.9871
## age          0.013493   0.008863   1.522   0.1356
## theft       -0.005364   0.005837  -0.919   0.3635
## log(income) -0.111412   0.979311  -0.114   0.9100
## fire         0.031161   0.017089   1.823   0.0755 .
## volact      -0.169565   0.077180  -2.197   0.0337 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.649   Deviance explained = 61.8%
## -REML = 44.186  Scale est. = 0.3841    n = 47
```

```r
p2 <- 1.132
phi2 <- 0.3841
mu2 <- tw_model2$fit[1]
poismean2<-mu^(2-p2)/(phi * (2-p2))
p0_2<-exp(-poismean2)

twden2<-exp(ldTweedie(xgrid, mu2, p = p2, phi = phi2)[,1])
plot(xgrid, twden2*(1-p0_2), type = "l", xlab = "x", ylab = "Density", main = "Observation 60645 Predic
dmax2<-max(twden2 * (1-p0_2))
segments(0, 0, 0, dmax2, lwd =  2)
text(0.05, dmax2, paste0("p =", signif(p0_2, 3)))
```



The p value for the full model (0.416) is slightly higher than that of the smaller model (0.398). Additionally, the density plot shows that the full model has a higher maximum value, and its curve appears more irregular compared to the smoother curve of the smaller model