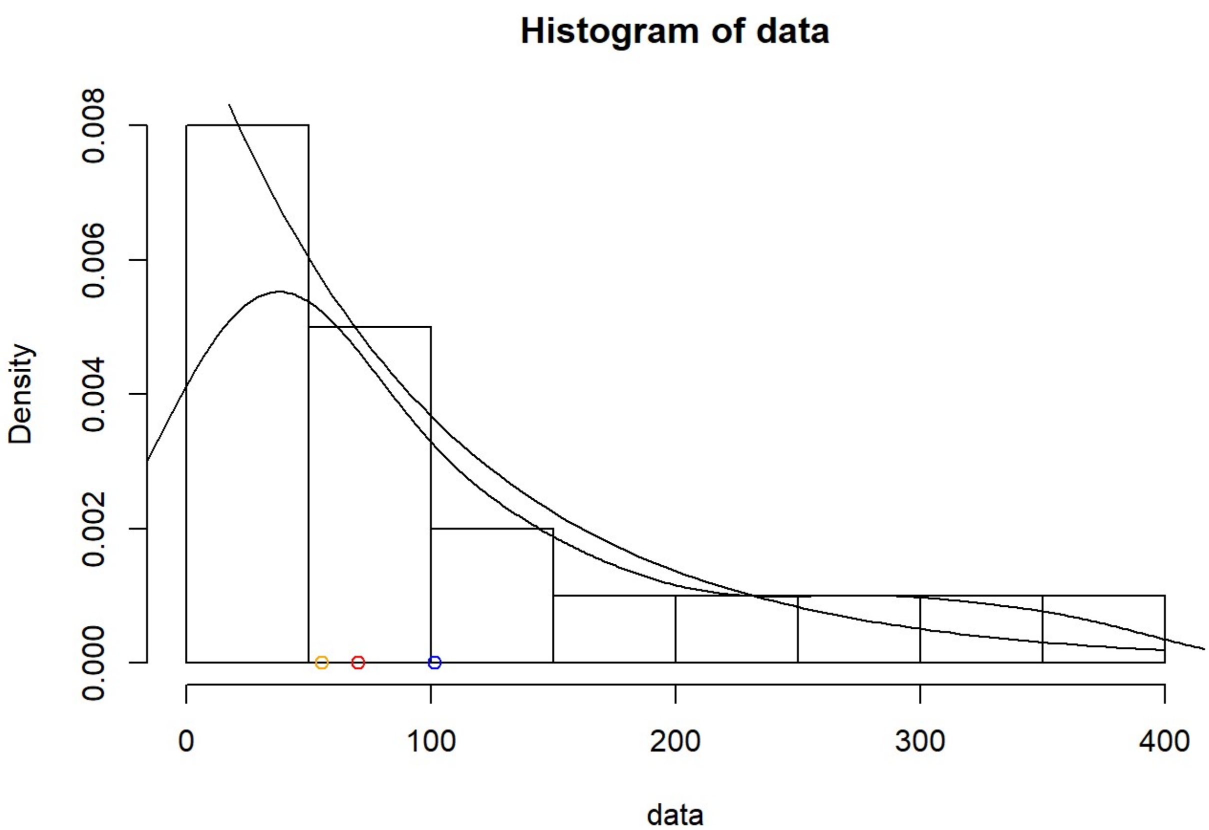
HW2Solutions

Dr. Syring

February 2, 2020

4.1.1

```
data <- c(1,4,5,21,22,28,40,42,51,53,58,67,95,124,124,160,202,260,303,363)
hist(data,pr=TRUE)
lines(density(data))
gam<-function(x) dgamma(x,shape = 1, scale = mean(data))
curve(gam, add = TRUE)
points(x=c(101.5,70.35, 55.5),y=c(0,0,0),col=c("blue", "red", "orange"))
```



Write the likelihood of the $\Gamma(1, \theta)$ model (in shape-scale parametrization as opposed to shape-rate parametrization) as

$$L(\theta; x) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x/\theta}.$$

You maye recognize $\Gamma(1, \theta)$ as an exponential distribution. Then, the loglikelihood is

$$\ell(\theta; x) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^{n} x_i.$$

The estimating equation is $\frac{\partial \ell}{\partial \theta} = 0$, given by

$$0 = -n/\theta + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i$$

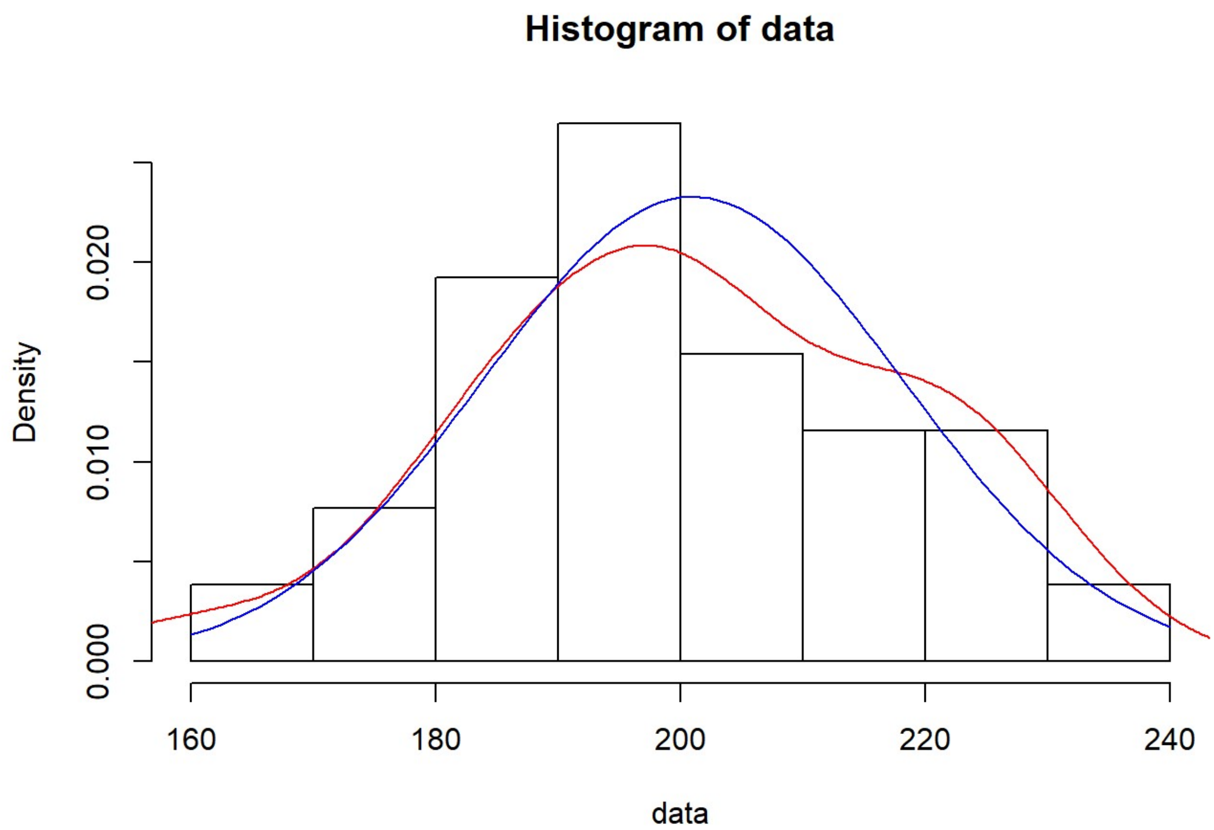
which provides $\hat{\theta}_{MLE} = \bar{x}$. For the given data, $\bar{x} = 101.15$. If you used the shape-rate parametrization then you got $\hat{\theta}_{MLE} = 1/\bar{x}$ which will result in the same Gamma density using that parametrization.

The median of an exponential density is given by $0.5 = 1 - e^{-x/\theta}$ where the right-hand side is simply the exponential CDF. Solving, we find the median occurs at $x = \theta \log 2$. Following the discussion on page 227 (after eq 4.1.3) note that the MLE of a function $g(\theta)$ is given by $g(\hat{\theta}_{MLE})$, a fact which may be surprising, but there is good reason for it (we will discuss in chapter 6…delta theorem). Therefore, the MLE estimate of the median is $\hat{\theta}_{MLE} \log 2 = 70.35444$ for the given data.

Whether or not the Gamma model with shape equals 1 fits the data is a somewhat subjective question. I would say the fit is fair, but perhaps not excellent. A more flexible model may fit better, for example, allow the shape parameter to vary.

4.1.2

```
data <- c(160,175,180,185,185,185,190,190,195,195,195,200,200,200,200,205,205,210,210,
218,219,220,222,225,225,232)
hist(data,pr=TRUE)
lines(density(data),col='red')
nm<-function(x) dnorm(x,mean = mean(data), sd = sqrt(var(data)*(25/26)))
curve(nm, add = TRUE, col = 'blue')
```



The MLEs are simply the sample mean and (adjusted) sample variance. For functions of these parameters, the estimates are those functions of the MLEs:

$$\hat{\mu}_{MLE} = \bar{x} = 201$$

$$\hat{\sigma^2}_{MLE} = (n-1)s^2/n = 293.9231$$

$$\hat{\sigma}_{MLE} = \sqrt{\hat{\sigma^2}_{MLE}} = \sqrt{(n-1)s^2/n} = 17.1442$$

$$\hat{\mu/\sigma}_{MLE} = \hat{\mu}_{MLE}/\hat{\sigma}_{MLE} = 11.724$$

$$\hat{p} = 1 - \Phi(\frac{215-201}{17.1442}) = 0.2070778$$

The Binomial model simply means treating the observations as iid Bernoulli 1 or 0 whether or not weight is greater than 215. Mathematically, consider the data $y_i = 1(x_i > 215)$ where $1(\cdot)$ denotes the indicator function and the $x_i's$ are the above weights. Then, the MLE is the sample proportion $\hat{p} = 7/26$. Note the discrepancy between this value and the normal model.

The Normal model is an ok fit, but the data have slightly more extreme weight values than the normal model suggests.

4.1.3

```
data <- c(9,7,9,15,10,13,11,7,2,12)
mean(data)
```

```
## [1] 9.5
```

The loglikelihood of the Poisson is

$$\ell(\theta; x) = \sum_i \log\left(\lambda^{x_i} e^{-\lambda}/x_i!\right).$$

The estimating equation is then given by $\frac{1}{\lambda}\sum_i x_i - n$ so that $\hat{\lambda}_{MLE} = \bar{x}$ which is the value 9.5 for the given data. For any given day I would expect 9 or 10 customers to enter between 9am and 10am.

4.1.4

The normal density is given by $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. For iid data the joint pdf is

$$f(x; \mu, \sigma^2) = \prod_{i=1^n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}.$$

Taking the natural log we have

$$\ell(\mu, \sigma^2; x) = -(n/2)\log\sigma^2 - (n/2)\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2.$$

(The books takes a derivative w.r.t. σ but this is not a very good idea becuase we typically think of σ^2 as the parameter, not σ). Then, the estimating equations are

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu) = 0$$

and

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^n (x_i - \mu)^2 = 0$$

To solve the system, first solve the first equation to obtain $\hat{\mu}_{MLE} = \bar{x}$. Plug this into the second equation to obtain

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0$$

Multiply through by $(\sigma^2)^2$ and solve to get $\hat{\sigma^2}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

4.1.9

Consider the histogram estimate of the density f at a point x defined by

$$\hat{f(x)} = \frac{1}{2hn}\sum_{i=1}^{n} 1(x - h < X_i < x + h).$$

This estimator is a sample average, divided by the interval width. It can be viewed as a constant multiple of a sum of of iid Bernoulli r.v.', which is also a constant multiple of a Binomial r.v. Therefore, the mean can be computed

$$E(\hat{f(x)}) = \frac{1}{2hn}\sum_{i} E(1(x - h < X_i < x + h)).$$

The mean $E(1(x - h < X_i < x + h))$ is the mean of a Bernoulli r.v. with probability $p = F_X(x + h) - F_X(x - h)$. Therefore,

$$E(\hat{f(x)}) = \frac{F_X(x + h) - F_X(x - h)}{2h}.$$

If you imagine h is a small positive number, then you can recognize this expression as an approximation to the derivative of the cdf F at the point x , which is the pdf f at x , the target of our estimation. Then, the bias of $\hat{f(x)}$ is simply

$$E(\hat{f(x)}) - f(x) = \frac{F_X(x + h) - F_X(x - h)}{2h} - f(x).$$

Clearly, this is a function of h and is small for small h , vanishing for $h \to 0$.

Similarly, the variance is the variance of a Bernoulli, with the leading constant squared,

$$V(\hat{f(x)}) = \frac{1}{4h^2n^2} n(F_X(x + h) - F_X(x - h))(1 - F_X(x + h) + F_X(x - h))$$

$$= \frac{1}{2nh}\frac{F_X(x + h) - F_X(x - h)}{2h} - \frac{1}{n}\left[\frac{F_X(x + h) - F_X(x - h)}{2h}\right]^2$$

This provides some interesting clues…Suppose h is a vanishing sequence in n , that is, $h := n^{-p}$ for some $p > 0$. Then, the second term above vanishes as $n \to 0$. If $p > 1$ the first term diverges; if $p = 1$, then the first term has a non-zero proper limit; and, if $0 < p < 1$ the first term vanishes. This is a nice example of bias-variance trade-off. We want h to vanish quickly to minimize bias, but if it vanishes too quickly, it causes the variance to explode.