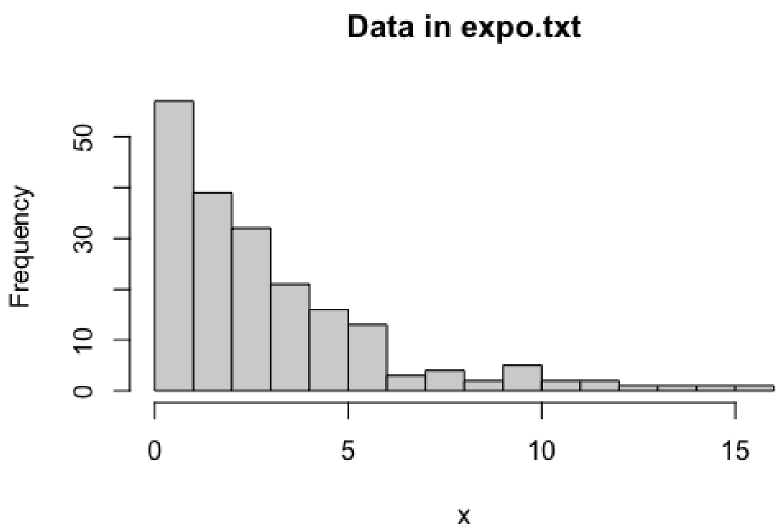
Problem Set 5 (answers)

S352

1. (a) The histogram shows the "ski slope" shape typical of an Exponential distribution, so it's quite plausible that's the right probability model. (Note: For this question, either a histogram or a density plot is fine. A QQ plot is not acceptable, as we are not checking normality.)

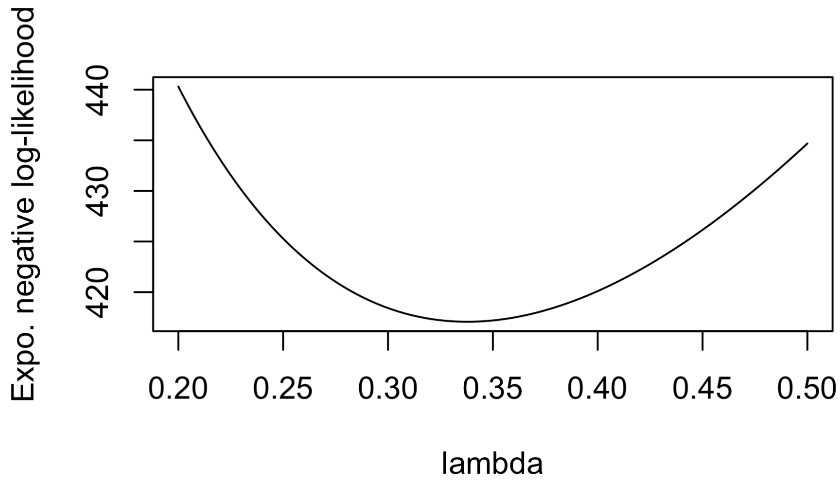


   (b) This NLL function should be for an exponential distribution!

```
expo.nll <- function(lambda, data){
    lik <- dexp(data, lambda)
    freturn(-sum(log(lik)))
}
```

   (c) Basic `plot` function would work fine. The curve's lowest point corresponds to a λ value around 0.34 and 0.35. So the MLE of λ is about 0.34.

```
lambda <- seq(0.2, 0.5, 0.001)
NLL <- rep(NA, length(lambda))
for(J in 1:length(lambda)){
    NLL[J] <- expo.nll(lambda[J])
}
plot(lambda, NLL, type = "l", ylab = "Expo. negative log-likelihood")
```



If you want to use the R function `curve`, you may need to vectorize your NLL function first, using `Vectorize()`. For example:

```
h <- Vectorize(expo.nll)
curve(h, 0.2, 0.5)
```

(d) The MLE of the rate parameter in the Exponential distribution, λ, is $1/\bar{X}$. In R, `1/mean(expo) = 0.338.`

Feed your NLL function into `optim()` to find the MLE numerically, 0.338.

```
expo.mle <- optim(par = 1, expo.nll, data = expo, method = "Brent",
                  lower = 0, upper = 50, hessian = TRUE)
expo.mle$par
[1] 0.3377749

expo.mle$hessian
[1,] 1753.004
```

(e) The MOM estimate of λ is the same as the MLE, $1/\bar{x} = 0.338$.

(f) The Hessian is 1753, so the estimated standard error is $1/\sqrt{1753} \approx 0.0239$. For a 95% confidence interval, just take $\hat{\lambda}$ and add/subtract 1.96 standard errors:

```
expo.mle$par - qnorm(.975) * 1 / sqrt(expo.mle$hessian)
expo.mle$par + qnorm(.975) * 1 / sqrt(expo.mle$hessian)
```

The 95% confidence interval for λ is $[0.291, 0.385]$ or so.

(g) Using `plkhci` to obtain the interval: 0.293 to 0.387. Profile likelihood CI is pretty close to the Wald interval of 0.291 to 0.385, because the sample size is somewhat big.

```
library(Bhat)
control.list = list(label = "lambda", est = 1/mean(expo), low = 0, upp = 999)
plkhci(control.list, expo.nll, "lambda")
```

2.

(a)
```
mensweights <- scan("mensweights.txt")
summary(mensweights)
hist(mensweights, breaks = seq(100, 500, 10),
    xlab = "Weight (pounds)", main = "Histogram of men's weights")
```

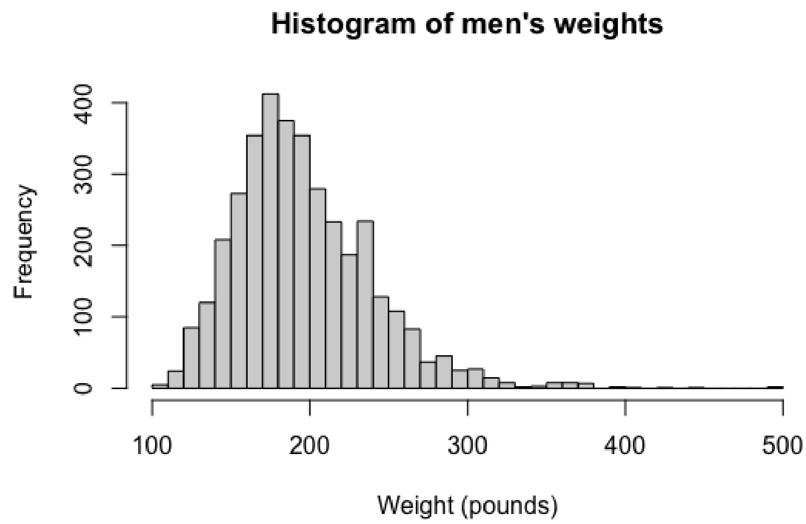


Figure 1: Histogram of the weights of 3,654 adult U.S. men, question 2(a).

(b) The two options that come to my mind are:

- Model the weights with a gamma distribution;
- Model the log weights with a Normal distribution.

Both are acceptable. Let's use the gamma because it is covered in class.
Let's find the MOMs first:

```
x.bar <- mean(mensweights)
sigma2.hat <- mean(mensweights^2) - x.bar^2
k.mom <- x.bar^2 / sigma2.hat
lambda.mom <- x.bar / sigma2.hat
```

I get $\hat{k}_{MOM} = 20.1$ and $\hat{\lambda} = 0.102$. Now find the MLEs:

```
weight.nll <- function(pars){
    k <- pars[1]
    lambda <- pars[2]
    lik <- dgamma(mensweights, shape = k, rate = lambda)
    loglik <- log(lik)
    return(-sum(loglik))
```

```
}
weight.mle <- optim(par = c(k.mom, lambda.mom), fn = weight.nll)
weight.mle$par
```

I get $\hat{k} = 21.7$ and $\hat{\lambda} = 0.111$.

(c)
```
set.seed(352)
weight.sim <- rgamma(3654, shape = weight.mle$par[1], rate = weight.mle$par[2])
summary(weight.sim)
hist(weight.sim, breaks = seq(80, 380, 10), xlab = "Weight (pounds)",
    main = "Simulated weights from a gamma distribution")
```

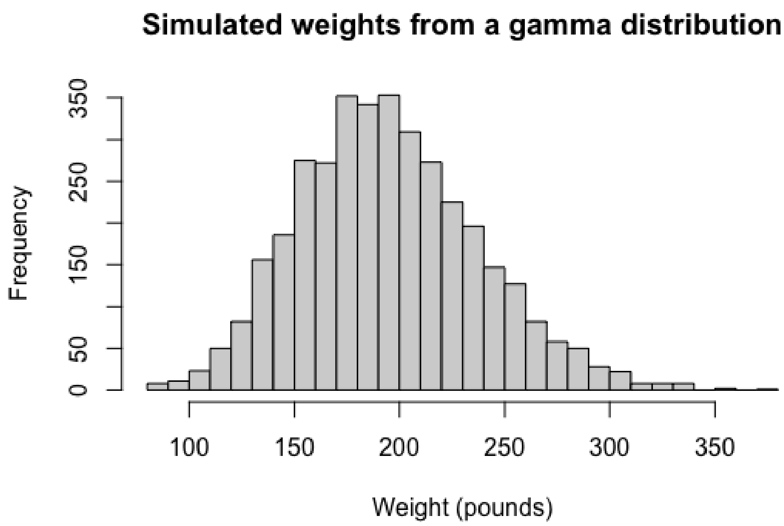


Figure 2: Histogram of simulated weights of adult U.S. men from a Gamma(21.7, 0.111) distribution, question 2(c).

The gamma fit isn't bad. There are a few differences between the raw and simulated data. The simulated data has a few too many low weights (below 150 pounds) and not enough mid-range weights from 150 to 200 pounds. The real data also has a few more extremely high weights.

(d) The mean of the real weights is 195.9 pounds. The mean of my simulated weights was 195.0 pounds (yours may vary), so pretty close.

```
> mean(mensweights)
[1] 195.8951
```

```
> mean(weight.sim)
[1] 194.9704
```

(e) The SD of the real weights (using the sample SD s) is 43.7 pounds, compared to 42.7 pounds in the simulated data. Again, not too bad.

```
> sd(mensweights)
[1] 43.72325
```

```
> sd(weight.sim)
[1] 42.67548
```