Homework 4 Solution

Math 445, Spring 2017

Exercise 6

See handwritten solution.

Exercise 11

See handwritten solution.

Exercise 16

(a) See handwritten solution for derivation.

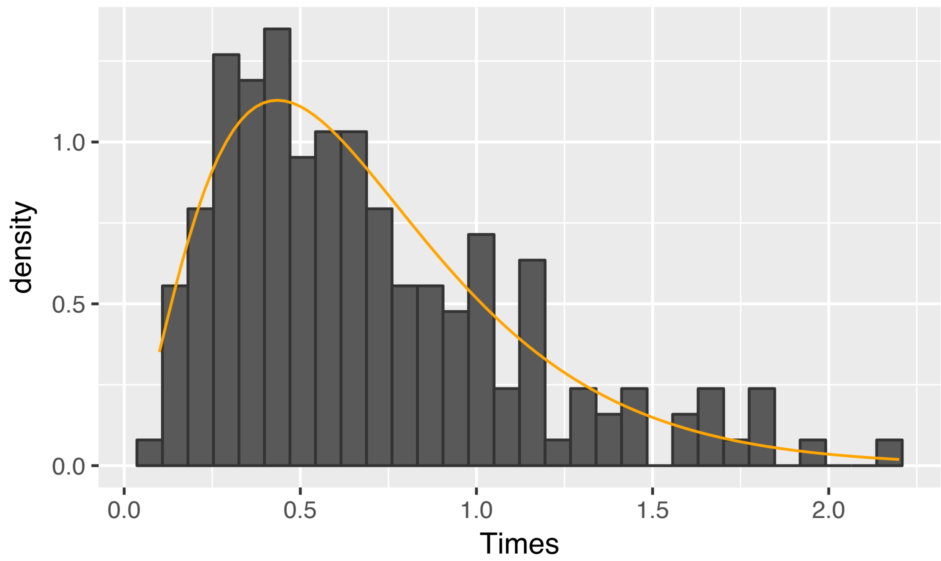
```
service <- read.csv("https://raw.githubusercontent.com/math445-LU/2016/master/data/Service.csv")

# Calculating MOM estimates using "by hand" formulas
xbar <- mean(service$Times)
sumxi2 <- sum(service$Times^2)
n <- nrow(service)

r_mom <- xbar^2 / ((1/n) * sumxi2 - xbar^2); r_mom

## [1] 2.670167

lambda_mom <- xbar / ((1/n) * sumxi2 - xbar^2); lambda_mom

## [1] 3.84239
```

(b) Based on a χ^2 test statistic of 1.977 and associaated p-value of 0.9611, there is no evidence that the distribution of service times differs from Gamma(2.67, 3.84).

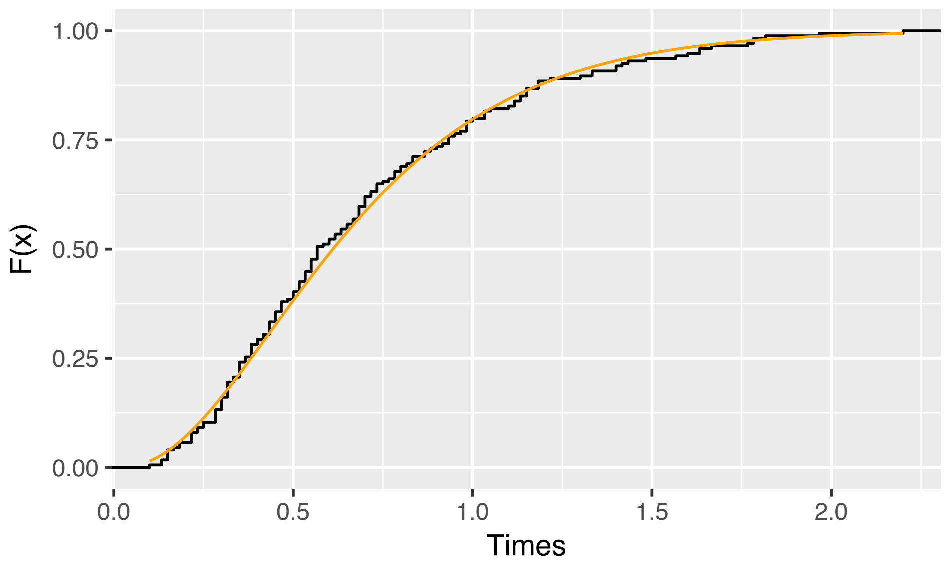
```
# Define the bins
q <- qgamma(p = seq(0 , 1, by = .1), shape = r_mom, rate = lambda_mom)

# Get the counts in each sub-interval
count <- hist(service$Times, breaks = q, plot = FALSE)$counts
expected <- length(service$Times) * .1

# Compute chi-square test statistic and associated p-value
stat <- sum ((count-expected)^2/expected); stat

## [1] 1.977011

1 - pchisq(stat, df = 10 - 2- 1)

## [1] 0.9611018
```

(c)

```
# Histogram
library(ggplot2)
ggplot(data = service) +
  geom_histogram(mapping = aes(x = Times, y = ..density..), colour = "gray20") +
  stat_function(fun = dgamma, geom = "line", args = list(shape = r_mom, rate = lambda_mom), colour = "orang

## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```




```
# ECDF
ggplot(data = service) +
  stat_ecdf(mapping = aes(x = Times), geom = "step") +
  stat_function(fun = pgamma, geom = "line", args = list(shape = r_mom, rate = lambda_mom), colour = "orang
  ylab("F(x)")
```



Exercise 19

See handwritten solution.

Exercise 23

See handwritten solution.

Problem 5: Text messages

The number of text messages sent per day by students at Lawrence is thought to follow a log-normal distribution. Let $X_1, X_2, \ldots, X_n$ be a random sample from LogNorm(μ, σ^2). The PDF of X_i is given by

$$f(x_i|\mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(\frac{-[\log(x) - \mu]^2}{2\sigma^2}\right), \quad x > 0, \quad \mu \in \mathbb{R}, \quad \sigma > 0$$

Note: $\exp x = e^x$.

(a) Let $Y \sim \mathcal{N}(\mu,\ \sigma^2)$. Define $X = e^Y$. Show that $X \sim \text{LogNorm}(\mu,\ \sigma^2)$.

See handwritten solution.

(b) Calculate the mean of X. (*Hint:* this can be found directly using the PDF, or by exploiting it's relationship with the normal distribution via the MGF.)

See handwritten solution.

(c) Calculate the variance of X. (*Hint:* this can be found directly using the PDF, or by exploiting it's relationship with the normal distribution via the MGF.)

See handwritten solution.

(d) Find the method of moments estimators for μ and σ^2 by hand.

See handwritten solution.

(e) What is the likelihood function for μ and σ^2?

See handwritten solution.

(f) What is the log-likelihood function for μ and σ^2?

See handwritten solution.

(g) Find the maximum likelihood estimators of μ and σ^2 by hand.

See handwritten solution.

(h) Write a function in R called `lognorm_loglik` that computes the log-likelihood for a lognormal distribution. Include the code in your solution.

```r
lognorm_loglik <- function(theta, x)
{
  sum(dlnorm(x, meanlog = theta[1], sdlog = theta[2], log = TRUE))
}
```

(i) The data set `stat111texts.csv` contains a variable, *texts*, which measures the typical number of text messages sent per day for a college student a certain prestigious university with a statistics department. In R, read in the data set, and calculate the mean, variance, and number of observations for the variable texts. Include both the code and output in your solution.

```r
stat111texts <- read.csv("https://raw.githubusercontent.com/math445-LU/2016/master/data/stat111texts.csv")

texts <- stat111texts$texts
```

```r
xbar <- mean(texts); xbar
```

```
## [1] 39.66667
```

```r
s2 <- var(texts); s2
```

```
## [1] 2318.315
```

```r
n <- length(texts); n
```

```
## [1] 90
```

(j) Define a new variable called logtexts = log texts. Calculate the mean and variance of logtexts. Include both the code and output in your solution.

```r
logtexts <- log(texts)

log_xbar <- mean(logtexts); log_xbar
```

```
## [1] 3.140327
```

```r
log_s2 <- var(logtexts); log_s2
```

```
## [1] 1.180902
```

(k) Use your results from (i) and (j) to calculate the ML estimates for μ and σ^2.

```r
muML <- mean(logtexts); muML
```

```
## [1] 3.140327
```

```r
sig2ML <- sum((logtexts - mean(logtexts))^2) / n; sig2ML
```

```
## [1] 1.167781
```

(l) Use the `optim` function in R to calculate the maximum likelihood estimates of μ and σ^2. Do they agree with your previous calculations? If not, consider different starting values.

```r
mu_mom <- 2 * log(xbar) - log(sum(texts^2) / n) / 2
sig2_mom <- log(sum(texts^2) / n) - 2 * log(xbar)

mles <- optim(par = c(mu_mom, sqrt(sig2_mom)), fn = lognorm_loglik,
              control = list(fnscale = -1), x = texts)
mles$par
```

```
## [1] 3.140374 1.080732
```

The ML estimates are: $\widehat{\mu} = 3.1403736$ and $\widehat{\sigma}^2 = 1.1679821$ (notice that $\widehat{\sigma}$ was returned.

$\boxed{6.6}$

$X_1, \dots, X_n$ iid w/ PDF $f(x_i|\theta) = e^{\theta - x}$, $x > \theta > 0$

(a) $L(\theta | x_1, \dots, x_n) = \prod_{i=1}^{n} e^{\theta - x_i} = e^{n\theta - \Sigma x_i}$

So $\ell(\theta) = \log(L(\theta|x_1, \dots, x_n)) = n\theta - \Sigma x_i$

Taking the derivative we find

$$\frac{d}{d\theta} \ell(\theta) = n$$

indicating that $\ell(\theta)$ is strictly increasing for $\theta < X_{min}$. Since θ can never equal X_{min}, $\ell(\theta)$ does not achieve a maximum.

(b) From part (a) we know that the likelihood is increasing through $(0, X_{min})$. Since the likelihood is continuous at $\theta = X_{min}$, the max. occurs at X_{min}.

Math 445, Spring 2016, HW #4 Solution

Problem 1: 6.11

$X_1, \ldots, X_n \overset{iid}{\sim} Exp(\lambda)$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} Exp(2\lambda)$. X_i indep Y_j $\forall i, j$

$$L(\lambda \mid X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i} \prod_{j=1}^{m} 2\lambda e^{-2\lambda Y_j}$$

$$= \lambda^n e^{-\lambda \sum_i X_i} (2\lambda)^m e^{-2\lambda \sum_j Y_j}$$

$$\ell(\lambda) = \log L(\lambda \mid X_1, \ldots, X_n, Y_1, \ldots, Y_m)$$

$$= n \log \lambda - \lambda \sum_i X_i + m \log 2 + m \log \lambda - 2\lambda \sum_j Y_j$$

$$\frac{d}{d\lambda} \ell(\lambda) = \frac{n}{\lambda} - \sum_i X_i + \frac{m}{\lambda} - 2 \sum_j Y_j = 0$$

$$\frac{n+m}{\lambda} = \sum_i X_i + 2 \sum_j Y_j$$

$$\boxed{\hat{\lambda} = \frac{n+m}{\sum_i X_i + 2 \sum_j Y_j}}$$

* Check that it's a max!

$\boxed{\text{Problem 2: 6.16}}$

Assume that $X_1, \dots, X_n \overset{iid}{\sim} \text{Gamma}(r, \lambda)$

(a) MOM estimates

$$E(X_1) = \frac{r}{\lambda} = \bar{X}_n \qquad E(X_1^2) = \frac{r^2 + r}{\lambda^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$r = \lambda \bar{X}_n \qquad \frac{1}{\lambda^2}\left[\lambda^2 \bar{X}_n^2 + \lambda \bar{X}_n\right] = \frac{1}{n}\sum_i X_i^2$$

$$\hat{r}_{MOM} = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_i X_i^2 - \bar{X}_n^2} \qquad \bar{X}_n^2 + \frac{1}{\lambda}\bar{X}_n = \frac{1}{n}\sum_i X_i^2$$

$$\frac{1}{\lambda} = \frac{\frac{1}{n}\sum_i X_i^2 - \bar{X}_n^2}{\bar{X}_n}$$

$$\hat{\lambda}_{MOM} = \frac{\bar{X}_n}{\frac{1}{n}\sum_i X_i^2 - \bar{X}_n^2}$$

*Since we have data in hand, we could simply have equated the pop. moments with the values of the sample moments and solve for r and λ.

See R Code.

$\hat{r}_{MOM} = 2.67 \qquad \hat{\lambda}_{MOM} = 3.84$

$\boxed{\text{Problem 3: 6.19}}$

$X_1, \ldots, X_n$ iid from $f(x|\theta) = \theta 2^{\theta} / x^{\theta+1}$, $x \geq 2$, $\theta > 1$.

(a) MOM estimator

$$E(X_1) = \int_2^{\infty} x \cdot \frac{\theta 2^{\theta}}{x^{\theta+1}} dx = \theta 2^{\theta} \int_2^{\infty} x^{-\theta} dx$$

$$= \theta 2^{\theta} \left[\frac{-1}{\theta-1} x^{-\theta+1} \Big|_2^{\infty} \right]$$

$$= \frac{\theta}{\theta-1} 2^{\theta} \left(2^{-\theta+1} \right)$$

$$= \frac{2\theta}{\theta-1}$$

$$\frac{2\theta}{\theta-1} = \bar{X}_n \iff 2\theta = (\theta-1)\bar{X}_n$$

$$\frac{2}{\bar{X}_n} = 1 - \frac{1}{\theta}$$

$$\frac{\bar{X}_n - 2}{\bar{X}_n} = \frac{1}{\theta}$$

$$\boxed{\hat{\theta}_{MOM} = \frac{\bar{X}_n}{\bar{X}_n - 2}}$$

(b) ML estimator

$$L(\theta | X_1, \ldots, X_n) = \prod_{i=1}^{n} \frac{\theta 2^{\theta}}{X_i^{\theta+1}} \quad \frac{\theta^n 2^{n\theta}}{\prod_{i=1}^{n} X_i^{\theta+1}}$$

$$l(\theta) = \log L(\theta | X_1, \ldots, X_n) = n \log \theta + n\theta \log 2 - \sum_i (\theta+1) \log(X_i)$$

$$\frac{d}{d\theta} l(\theta) = \frac{n}{\theta} + n \log 2 - \sum_i \log X_i = 0$$

$$\boxed{\hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^{n} \log X_i - n \log 2}}$$

* Check that it's a max.

$\boxed{6.23}$

$E(\bar{X}) = E(X_i) = \theta + 1 \implies \bar{X}$ is biased because $E(\bar{X}) \neq \theta$

$$E(X_i) = \int_\theta^\infty x \cdot e^{\theta - x}\, dx = e^\theta \int_\theta^\infty x e^{-x}\, dx$$

using integration by parts

$$= e^\theta \left[-x e^{-x} \Big|_\theta^\infty + \int_\theta^\infty e^{-x}\, dx \right]$$

$$= e^\theta \left[\theta e^{-\theta} + \left(-e^{-x} \Big|_\theta^\infty \right) \right]$$

$$= e^\theta \left[\theta e^{-\theta} + e^{-\theta} \right]$$

$$= \theta + 1$$

$\text{Bias}(\bar{X}) = E(\bar{X}) - \theta = 1$

Since $\text{Bias}(\bar{X}) = 1$, an unbiased estimator of θ is given by $\hat{\theta} = \bar{X} - 1$.

Problem 5

(a) Let $Y \sim N(\mu, \sigma^2)$ and $X = e^Y$.

$$F_X(x) = P(X \leq x) = P(e^Y \leq x) = P(Y \leq \log(x)) = F_Y(\log(x))$$

Thus, the PDF of X is

$$f_X(x) = \frac{d}{dx} F_Y(\log(x)) = f_Y(\log(x)) \left(\frac{1}{x}\right)$$

$$= \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\log(x) - \mu)^2} \quad \text{for } x > 0.$$

(b) Recall that the MGF of $Y \sim N(\mu, \sigma^2)$ is $M_Y(t) = e^{\mu t + \sigma^2 t^2/2}$

Continuing to consider $X = e^Y$,

$$E(X) = E(e^Y) = M_Y(1) = e^{\mu + \sigma^2/2}$$

(c) Using a similar strategy

$$E(X^2) = E(e^{2Y}) = M_Y(2) = e^{2\mu + 2\sigma^2}$$

$$Var(X) = E(X^2) - [E(X)]^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

(d) $$e^{\mu + \sigma^2/2} = \bar{X}_n$$

$$\mu = \log(\bar{X}_n) - \frac{1}{2}\sigma^2$$

$$\hat{\mu}_{MOM} = \log(\bar{X}_n) - \frac{1}{2}\left[\log\left(\frac{1}{n}\sum X_i^2\right) - 2\log(\bar{X}_n)\right]$$

$$= 2\log(\bar{X}_n) - \frac{1}{2}\log\left(\frac{1}{n}\sum X_i^2\right)$$

$$e^{2\mu + 2\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$2\sigma^2 = \log\left(\frac{1}{n}\sum X_i^2\right) - 2\mu$$

$$\sigma^2 = \frac{1}{2}\log\left(\frac{1}{n}\sum X_i^2\right) - \log(\bar{X}_n) + \frac{1}{2}\sigma^2$$

$$\frac{1}{2}\sigma^2 = \frac{1}{2}\log\left(\frac{1}{n}\sum X_i^2\right) - \log(\bar{X}_n)$$

$$\hat{\sigma}^2_{MOM} = \log\left(\frac{1}{n}\sum X_i^2\right) - 2\log(\bar{X}_n)$$

(e) $$L(\mu,\sigma^2 \mid X_1,\dots,X_n) = \prod_{i=1}^{n} \frac{1}{X_i\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\log(X_i)-\mu)^2}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \left(\prod_{i=1}^{n} \frac{1}{X_i}\right) e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log(X_i)-\mu)^2}$$

(f) $$l(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n}\log(X_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log(X_i)-\mu)^2$$

$$= -\frac{n}{2}\left[\log(2\pi) + \log(\sigma^2)\right] - \sum_i \log(X_i) - \frac{1}{2\sigma^2}\sum_i(\log(X_i)-\mu)^2$$

(g) $$\frac{\partial}{\partial\mu} l(\mu,\sigma^2) = \frac{1}{\sigma^2}\sum_i(\log(X_i)-\mu) = 0$$

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_i \log(X_i)$$

$$\frac{\partial}{\partial\sigma^2} l(\mu,\sigma^2) = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_i(\log(X_i)-\mu)^2 = 0$$

$$\sum_i\left(\log(X_i) - \frac{1}{n}\sum_i\log(X_i)\right)^2 = n\sigma^2$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_i\left(\log(X_i) - \frac{1}{n}\sum_i\log(X_i)\right)^2$$

* You can check that it's a max.