7.2 Recall the gamma density: with \( \alpha, \beta > 0 \), we have

\[
f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad x > 0
\]

(a) When \( \alpha \) is known, the likelihood function is given by

\[
L(\beta|x) = \prod_{i=1}^{n} f(x_i|\alpha, \beta) = \frac{1}{(\Gamma(\alpha))^n \beta^n} \left( \prod_{i=1}^{n} x_i^{\alpha-1} \right) \exp \left( -\frac{1}{\beta} \sum_{i=1}^{n} x_i \right)
\]

The log-likelihood is then

\[
l \equiv \log L(\beta|x) = -\log (\Gamma(\alpha))^n - n\alpha \log \beta + (\alpha - 1) \left( \log \prod_{i=1}^{n} x_i \right) - \frac{1}{\beta} \sum_{i=1}^{n} x_i
\]

Hence

\[
\frac{\partial l}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i
\]

Setting the above derivative to zero, we find

\[
\beta n\alpha = \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \hat{\beta} = \frac{\sum_{i=1}^{n} x_i}{n\alpha}
\]

To check to see if the above gives a local maximum, we find

\[
\frac{\partial^2 l}{\partial \beta^2} = \frac{n\alpha}{\beta^2} - \frac{2}{\beta^3} \sum_{i=1}^{n} x_i
\]

Hence

\[
\frac{\partial^2 l}{\partial \beta^2} \bigg|_{\beta=\hat{\beta}} = n\alpha \left( \frac{n\alpha}{\sum_{i=1}^{n} x_i} \right)^2 - \frac{2(n\alpha)^3}{\left( \sum_{i=1}^{n} x_i \right)^2} = -\frac{(n\alpha)^3}{\left( \sum_{i=1}^{n} x_i \right)^2} < 0
\]

Thus it is a local maximum. Since \( \hat{\beta} \) is unique, it is also the global maximum.
We can solve the problem by using the result obtained in (a). For a given \( \alpha \), the value of \( \beta \) that maximizes \( L \) is given by (1). Hence the likelihood function as a function of the single variable is given by

\[
g(\alpha) = \frac{1}{(\Gamma(\alpha))^n b^{\alpha n}} \left( \prod_{i=1}^{n} x_i \right)^{\alpha - 1} \exp \left( -\frac{1}{\beta} \sum_{i=1}^{n} x_i \right) \bigg|_{\beta = \hat{\beta}}
\]

\[
= \frac{1}{(\Gamma(\alpha))^n} \left( \frac{n}{\sum_{i=1}^{n} x_i / n} \right)^{\alpha - 1} \exp \left( -\frac{n\alpha}{\sum_{i=1}^{n} x_i} \right)
\]

\[
= \frac{1}{(\Gamma(\alpha))^n} \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^{\alpha - 1} \exp (-n\alpha).
\]

So we first find

\[ \hat{\alpha} = \arg \max g(\alpha) \]

and then use (1) to find \( \hat{\beta} \). The above function cannot be evaluated numerically in that \( \Gamma(\alpha) \) will overflow for large \( \alpha \). So we do the log-likelihood instead:

\[
h(\alpha) = \log g(\alpha)
\]

\[
= -n \log (\Gamma(\alpha)) - n\alpha \left( \log \left( \sum_{i=1}^{n} x_i \right) - \log (n\alpha) \right) + (\alpha - 1) \left( \log \prod_{i=1}^{n} x_i \right) - n\alpha
\]

In R:

```r
> lst <- hw7.2()
[[1]]:
function()
{
  # Problem 7.2 in Cassella and Berger
  x <- c(22, 23.9, 20.9, 23.8, 25.1, 21.7, 23.8, 22.8, 23.1, 23.1, 23.6, 23.6, 23.1, 23.2)
  sum.x <- sum(x)
  prod.x <- prod(x)
  n <- length(x)
  x <- c(sum.x, prod.x, n)
  out <- nlminb(start = c(25), objective = obj.hw7.2, lower = c(5), x = x)
  alpha.h <- out$par
  beta.h <- sum.x/(n * alpha.h)
  list("alpha-hat" = alpha.h, "beta-hat" = beta.h)
}
```

```r
> hw7.2()
"alpha-hat":
[1] 514.3386
"beta-hat":
[1] 0.04234889
```
Thus the answers are $b=514.33$ and $b=0.045$.

7.9 (a) The method of moment approach: Here $E(X) = \theta/2$. We set $\overline{X} = \theta/2$. This means $\overline{\theta} = 2\overline{X}$ is the MM estimator. For this estimator, we have

$$E(\overline{\theta}) = E(2\overline{X}) = 2E(\overline{X}) = 2\frac{\theta}{2} = \theta$$

and

$$Var(\overline{\theta}) = Var(2\overline{X}) = 4Var(\overline{X}) = 4\frac{Var(X)}{n} = \frac{4\theta^2}{n} \quad \text{by (5.2.6)}$$

$$= \frac{4\theta^2}{n12} \quad \text{by p. 626 of the text}$$

$$= \frac{\theta^2}{3n}$$

(b) The MLE approach: The likelihood function is given by

$$L(\theta|x) = \prod_{i=1}^{n} \left( \frac{1}{\theta} \right) I_{(0,\theta)}(x_i)$$

Since we require $X_{(1)} > 0$ and $0 < X_{(n)} < \theta$, the above can be restated as

$$L(\theta|x) = \frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)}) I_{(0,\infty)}(x_{(1)})$$

For $\theta \geq x_{(n)}$, $1/\theta^n$ is decreasing in $\theta$. Thus the likelihood function is maximized when $\overline{\theta}$ is set at the smallest possible value $x_{(n)}$. So we conclude that the MLE estimator $\hat{\theta} = X_{(n)}$.  

3
Recall in Example 7.3.13, we found the density of \( X(n) \):

\[ f_{X(n)}(x) = \frac{n x^{n-1}}{\theta^n} \quad 0 < x < \theta \]

Also we found that the first two moments of \( X(n) \):

\[ \begin{align*}
E(X(n)) &= \frac{n}{n+1} \theta \\
E(X^2(n)) &= \frac{n}{n+2} \theta^2
\end{align*} \]

We thus conclude that

\[ E(\hat{\theta}) = \frac{n}{n+1} \theta \]

and

\[ Var(\hat{\theta}) = Var(X(n)) = \frac{n}{n+2} \theta^2 - \left( \frac{n}{n+1} \right)^2 \theta^2 \]

\[ = \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2} \theta^2 = \frac{n\theta^2}{(n+2)(n+1)^2}. \]

We see that \( \hat{\theta} \) is an unbiased estimator and \( \tilde{\theta} \) is a biased estimator and \( Var(\tilde{\theta}) < Var(\hat{\theta}) \) for all \( \theta \). When \( n \) is not too small, \( n/(n+1) \) is close to 1. But the reduction in the variance can be appreciable. This makes the MLE a more attractive choice for moderate to large sample size. \( \square \)

7. 10 (a) The density of \( X_i \) is

\[ f_{X_i}(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \quad 0 < x < \beta \]

Thus the joint density is

\[ f(x|\Theta) = \prod_{i=1}^{n} \frac{\alpha}{\beta^\alpha} x_i^{\alpha-1} I_{(0,\beta)}(x_i) = \left( \frac{\alpha}{\beta^\alpha} \right)^n \left( \prod_{i=1}^{n} x_i \right)^{\alpha-1} I_{(-\infty,\beta)}(x(n)) I_{(0,\infty)}(x(1)) = L(\Theta|x) \]

By the Factorization Theorem, we see that \( \prod_{i=1}^{n} x_i, X(n) \) are sufficient statistics for \( (\alpha, \beta) \).

(b) For any fixed \( \alpha \), \( L(\alpha, \beta|x) = 0 \) if \( \beta < x(n) \). On the other hand, if \( \beta \geq x(n) \), then \( L \) is decreasing in \( \beta \). This means that we should set \( \beta \) at the smallest possible value, namely, the MLE estimator for \( \beta \) is \( X(n) \) for any given \( \alpha \). Now

\[ l \equiv \log L = n \log \alpha - n \alpha \log \beta + (\alpha - 1) \log \prod_{i=1}^{n} x_i \]

Thence

\[ \frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} - n \log \beta + \log \prod_{i=1}^{n} x_i. \]
Now we set the above to zero and solve for $\alpha$:

$$\hat{\alpha} = \frac{n}{n \log \beta - \log \prod_{i=1}^{n} X_i} = n \left[ n \log X_{(n)} - \sum_{i=1}^{n} \log X_i \right]^{-1}$$

$$= n \left[ \sum_{i=1}^{n} \log X_{(n)} - \sum_{i=1}^{n} \log X_i \right]^{-1} = n \left[ \sum_{i=1}^{n} \log X_{(n)} - \log X_i \right]^{-1}$$

The second derivative is $-n/\alpha^2 < 0$. Hence the above is the MLE.

(c) Hence, we conclude $b_\alpha = 1.259$ and $b_\beta = 25$. □

7.19 (a) We see that the random variable $Y_i - \beta x_i = \epsilon_i$ where $\epsilon_i \sim n(0, \sigma^2)$. The joint density of $y$ is given by

$$f(y|\beta, \sigma^2) = \prod_{i=1}^{n} (2\pi \sigma^2)^{-1/2} \exp \left( -\frac{1}{2\sigma^2} (y_i - \beta x_i)^2 \right)$$

$$= (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2) \right)$$

$$= (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{\beta^2}{2\sigma^2} \sum_{i=1}^{n} x_i^2 \right) \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 \right) \exp \left( \frac{\beta}{\sigma^2} \sum_{i=1}^{n} x_i y_i \right)$$

By the Factorization Theorem, we see that $\left( \sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} x_i y_i \right)$ is sufficient for $(\beta, \sigma^2)$.

(b) The log-likelihood is

$$l \equiv \log L(\Theta|y) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{\beta^2}{2\sigma^2} \sum_{i=1}^{n} x_i^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2 + \frac{\beta}{\sigma^2} \sum_{i=1}^{n} x_i y_i$$
For any fixed value of $\sigma^2$, we have

$$
\frac{\partial l}{\partial \beta} = -\frac{\beta}{\sigma^2} \sum_{i=1}^{n} x_i^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i y_i
$$

So we set $\partial l/\partial \beta = 0$ and solve for $\beta$:

$$
\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}
$$

Also, we see

$$
\frac{\partial^2 l}{\partial \beta^2} = -\frac{1}{\sigma^2} \sum_{i=1}^{n} x_i^2 < 0
$$

Since $L \to 0$ as $\beta \to \pm\infty$ and $\hat{\beta}$ does not depend on $\sigma^2$, we conclude it is the MLE. Also

$$
E\hat{\beta} = E\left[ \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2} \right] = \frac{\sum_{i=1}^{n} x_i E[Y_i]}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i \beta x_i}{\sum_{i=1}^{n} x_i^2} = \beta
$$

Hence $\hat{\beta}$ is an unbiased estimator.

(c) We write

$$
\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2} = \sum_{i=1}^{n} a_i Y_i
$$

where

$$
a_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}
$$

and we see $a_i$s are constants. Since $Y_i$ is normal and a linear combination of normals is a normal, we only need the mean and variance to characterize it. The mean of $\hat{\beta}$ is given by $\beta$ as shown in (b). The variance is

$$
Var[\hat{\beta}] = Var\left[ \sum_{i=1}^{n} a_i Y_i \right] = \sum_{i=1}^{n} a_i^2 Var[Y_i] = \sum_{i=1}^{n} a_i^2 \sigma^2
$$

$$
= \sum_{i=1}^{n} \frac{x_i^2}{\sum_{i=1}^{n} x_i^2} \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} \square
$$
Problem E. (1) For each $W_i$, the density is given by

$$f(w|\mu, \sigma^2) = (2\pi\sigma^2 h)^{-1/2} \exp\left(-\frac{1}{2\sigma^2 h}(w - \mu h)^2\right)$$

Hence the likelihood function is given by

$$L(\Theta|w) = \prod_{i=1}^{N} (2\pi\sigma^2 h)^{-1/2} \exp\left(-\frac{1}{2\sigma^2 h}(w_i - \mu h)^2\right) = (2\pi\sigma^2 h)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2 h} \sum_{i=1}^{N} (w_i - \mu h)^2\right)$$

The log-likelihood function is then

$$l \equiv \log(L(\Theta|w)) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2 h) - \frac{1}{2\sigma^2 h} \sum_{i=1}^{N} (w_i - \mu h)^2$$

(2) The first-order conditions are:

$$\frac{\partial l}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} (w_i - \mu h) d\mu(w_i - \mu h) = 0 \tag{2}$$

and

$$\frac{\partial l}{\partial \sigma^2} = -\frac{N}{2} h \sigma^4 + \frac{1}{2\sigma^4 h} \sum_{i=1}^{N} (w_i - \mu h)^2 = 0 \tag{3}$$

From (2), we obtain

$$\sum_{i=1}^{n} w_i - N\mu h = 0 \quad \text{or} \quad \mu = \frac{\sum_{i=1}^{n} W_i}{Nh} \tag{4}$$

We multiply (3) by $2\sigma^4$. This gives

$$N\sigma^2 = \frac{1}{h} \sum_{i=1}^{N} (w_i - \mu h)^2 \quad \text{or} \quad \sigma^2 = \frac{1}{Nh} \sum_{i=1}^{N} (W_i - \bar{\mu}h)^2 \tag{5}$$

(3) First, we calculate the Hessian

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{Nh}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{2}{2\sigma^6 h} \sum_{i=1}^{N} (w_i - \mu h)^2 = \frac{N}{2\sigma^4} - \frac{1}{\sigma^4 h} \sum_{i=1}^{N} (w_i - \mu h)^2$$

$$\frac{\partial^2 l}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{i=1}^{N} (w_i - \mu h)$$
We use (3) and (5) in the above, and obtain

\[ \frac{\partial^2 l}{\partial \mu^2} = -\frac{Nh}{\sigma^2} \]

\[ \frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}Nh^2 = \frac{N}{2\sigma^4} - \frac{N}{2\sigma^4} \quad \text{by (5)} \]

\[ \frac{\partial^2 l}{\partial \mu \partial \sigma^2} = 0 \quad \text{by (3)} \]

We see the observed information matrix is

\[ I_O(\Theta) = \begin{bmatrix} \frac{Nh}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \]

and its inverse is given by

\[ \Psi = \begin{bmatrix} \frac{\hat{\sigma}^2}{Nh} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{N} \end{bmatrix} \]

Thus we conclude that

\[ \text{Var}[\hat{\mu}] = \frac{\hat{\sigma}^2}{Nh} \]

\[ \text{Var}[\hat{\sigma}^2] = \frac{2\hat{\sigma}^4}{N} \]

(4) The first two population moments are

\[ E(W_i) = \mu h \quad E(W_i^2) = \sigma^2 h + \mu^2 h^2 \]

The first two sample moments are

\[ \bar{W} = \frac{1}{N} \sum_{i=1}^{n} W_i \quad S^2 = \frac{1}{N} \sum_{i=1}^{n} W_i^2 \]

Equating the two first moments, we find

\[ \bar{\mu} = \frac{1}{Nh} \sum_{i=1}^{n} W_i = \hat{\mu}. \]

Equating the two second moments, we find

\[ \sigma^2 h + \mu^2 h^2 = \frac{1}{N} \sum_{i=1}^{n} W_i^2 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{n} (W_i - \bar{\mu}h)^2 = \hat{\sigma}^2. \]

So for this case, the MLE’s and MM estimators are identical.
(5) I selected the INTEL price series. It has the symbol "INTC". It is important to note that the Yahoo data list the most recent data first. So we need to reverse the price series so that the prices are listed in increasing order of the date index $n$.

```matlab
function [T]=gbm_model(S)
%
% Geometric Brownian Motion - Parameter Estimation - MLE

% S = daily closing prices
n=length(S); x3=S; x3(1)=1; x3(1)=[]; a=log(x3(1));
B=1; b1=256; bmu=bmu(x); mu=mue(x); mue(mue(x)); mue(bmu(x));
S=x3(a,(x-n)'^2)/2);%
forint(1) mu = 0.44e var = 0.44e [n',n',n'];
var(1) = 0.44e var(1) = 0.44e [n',n',n']
end
forint(1) var(1) = 0.44e var(1) = 0.44e [n',n',n']
end
```

(a), (b) Since the moment estimators and MLE estimators are identical in this case, the following results only give one set of results.

```
>> [T]=gbm_model(D);
mu = 0.5007 var = 0.3514
var(mu) = 0.8516 var(s2) = 0.0024
95 pct CI for mu: (-1.3080, 2.3094)
95 pct CI for var: (0.2559, 0.4470)
```

(c) The actual and simulated price series are shown below.
(d) It seems that the simulated price series exhibited a similar pattern of variation in this case. □