STAT 135 CONCEPTS OF STATISTICS
HOMEWORK 2

Assigned June 29, 2021, due July 6, 2021

This homework pertains to materials covered in Lecture 3, 4 and 5. The assignment can be typed or handwritten, with your name on the document, and **with properly labeled input code and computer output for those problems that require it**. To obtain full credit, please write clearly and show your reasoning. If you choose to collaborate, the write-up should be your own. Please show your work! Upload the file to the Week 2 Assignment on bCourses.

Note in this homework, we use the following abbreviations: Standard Error (SE), Method of Moments (MM), Maximum likelihood estimators (MLE) and Mean Suqared Error (MSE).

Problem 1. Consider the i.i.d random variables $X_1, \dots, X_n$ from some population $f(x|\theta)$.

(1) Show that the sample moments $\hat{\mu}_k = n^{-1}\sum_{i=1}^n X_i^k$ is an unbiased estimator of the kth population moment, for any natural number k.

(2) Despite (1), the parameter estimators that are obtained through the method of moments are sometimes biased, which happens when the functions inverted are nonlinear. However, the MM estimators always have the asymptotic unbiasedness, i.e.

$$E(\hat{\theta}_{MM}) \to \theta, \text{ as } n \to \infty. \tag{1}$$

Use Theorem E of Lecture 3 to prove the above limit (1).

Solution.

(1) By the i.i.d assumption,

$$E(\hat{\mu}_k) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \mu_k,$$

which shows sample moments are unbiased estimators of the population moments.

(2) Theorem E ensures that MM estimators are consistent as long as functions inverted are continuous, and consistency ensures asymptotic unbiasedness (Page 18 of Lecture 3):

$$E(\hat{\theta}_{MM}) \to \theta, \text{ as } n \to \infty.$$

Problem 2. Suppose $X_1, \dots, X_n$ are independently sampled from a Kumaraswamy distribution with parameter β. The probability density function of this distribution is

$$f(x|\beta) = \begin{cases} \beta(1-x)^{\beta-1}, & \text{if } x \in (0,1), \\ 0, & \text{otherwise.} \end{cases}$$

(1) Find the MM estimator $\hat{\beta}_{MM}$ for β.
(*Hint:* $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = \text{Beta}(\alpha,\beta)$. *See this Wikipedia page for the definition of the Beta function.*)
(2) Suppose a researcher collected 100 observations and the sample mean is 0.45. Calculate the MM estimate for β, and use the Delta Method to approximate the SE of your estimate $\hat{\beta}_{MM}$.

Solution.

(1) It is easy to calculate the population mean of the Kumaraswamy distribution:

$$\mu = \int_0^1 \beta x(1-x)^{\beta-1}dx = \beta\frac{\Gamma(2)\Gamma(\beta)}{\Gamma(\beta+2)} = \frac{1}{\beta+1}.$$

Thus, the MM estimator can be obtained through solving

$$\bar{X}_n = \frac{1}{\hat{\beta}_{MM}+1},$$

which gives $\hat{\beta}_{MM} = 1/\bar{X}_n - 1$.

(2) From (1), $\hat{\beta}_{MM} = 1/\bar{X}_{100} - 1 = 1/0.45 - 1 = 1.222$.
Denote $g(\mu) = 1/\mu - 1$. By the Delta Method,

$$\sqrt{n}\{g(\bar{X}_n) - g(\mu) \to g'(\mu)N(0,\sigma^2)\}.$$

It is easy to calculate the variance of the Kumaraswamy distribution:

$$\sigma^2 = \frac{\beta}{(1+\beta)^2(2+\beta)}.$$

Therefore,

$$\text{Var}[\hat{\beta}_{MM}] \approx \frac{[g'(\mu)]^2\sigma^2}{n} = \frac{\sigma^2}{n\mu^4} = \frac{\beta(1+\beta)^2}{n(2+\beta)},$$

and

$$\text{SE}(\hat{\beta}_{MM}) = \sqrt{\frac{\hat{\beta}_{MM}(1+\hat{\beta}_{MM})^2}{n(2+\hat{\beta}_{MM})}} \approx 0.1368626.$$

Problem 3. In Example 6 of Lecture 3, we found the MLE of the Normal population parameters μ and σ from its i.i.d observations. However, we only dealt with the equations constructed with the gradient of the log-likelihood

$$\nabla l(\mu,\sigma) = \left(\frac{\partial l}{\partial \mu}, \frac{\partial l}{\partial \sigma}\right) = (0,0),$$

whose solutions can be either a maximum point or a minimum point. To complete the proof that the solution is indeed a maximum point, show that the Hessian matrix of the log-likelihood satisfies

$$\nabla^2 l(\mu, \sigma) < 0$$

for all $\mu \in \mathbb{R}$ and $\sigma > 0$ (see how to check negative definiteness from Theorem 3.1 of this document).

Solution. We know the MLE for μ and σ are

$$\hat{\mu}_n = \bar{X}_n, \quad \hat{\sigma}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Next we evaluate the Hessian matrix for the Normal log-likelihood at $(\hat{\mu}_n, \hat{\sigma}_n)$:

$$\begin{aligned}\begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \sigma} \\ \frac{\partial^2 l}{\partial \sigma \partial \mu} & \frac{\partial^2 l}{\partial \sigma^2} \end{pmatrix}\Bigg|_{(\hat{\mu}_n, \hat{\sigma}_n)} &= \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{2}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu) \\ -\frac{2}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu) & \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2 \end{pmatrix}\Bigg|_{(\hat{\mu}_n, \hat{\sigma}_n)} \\ &= \begin{pmatrix} -\frac{n}{\hat{\sigma}_n^2} & 0 \\ 0 & -\frac{2n}{\hat{\sigma}_n^2} \end{pmatrix}.\end{aligned}$$

Since the first diagonal element $-n/\hat{\sigma}_n^2 < 0$ and the determinant

$$|\nabla^2 l(\hat{\mu}_n, \hat{\sigma}_n)| = 2\left(\frac{n}{\hat{\sigma}_n}\right)^2 > 0,$$

$\nabla^2 l(\hat{\mu}_n, \hat{\sigma}_n)$ is negative definite, and thus MLE found through the gradient equations are indeed the maximum point.

[*Optional*] To examine $|\nabla^2 l(\mu, \sigma^2)|$ for any $\mu \in \mathbb{R}$ and $\sigma > 0$, we need to have sufficiently large sample size and the asymptotic properties. Assume (μ_0, σ_0) are the underlying true parameters,

$$\begin{aligned}|\nabla^2 l(\mu, \sigma)| &= -\frac{n^2}{\sigma^4} + \frac{3n}{\sigma^6}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{4n^2}{\sigma^6}(\bar{X}_n - \mu)^2 \\ &\overset{\bar{X}_n \to \mu_0 \text{ a.s.}}{\approx} -\frac{n^2}{\sigma^4} + \frac{3n}{\sigma^6}\sum_{i=1}^{n}(X_i - \mu_0)^2 + \frac{3n^2}{\sigma^6}(\mu_0 - \mu)^2 - \frac{4n^2}{\sigma^6}(\mu_0 - \mu)^2 \\ &\overset{\hat{\sigma}_n \to \sigma_0 \text{ a.s.}}{\approx} -\frac{n^2}{\sigma^4} + \frac{3n\sigma_0^2}{\sigma^6} - \frac{n^2}{\sigma^6}(\mu_0 - \mu)^2,\end{aligned}$$

with which we can see that $|\nabla^2 l(\mu, \sigma)| > 0$ only holds when (μ, σ) is in a close neighborhood of (μ_0, σ_0). Therefore, it is not true that $\nabla^2 l(\mu, \sigma^2)$ is negative definite for any $\mu \in \mathbb{R}$ and $\sigma > 0$.

Although we just showed the Normal log-likelihood is not a concave function globally, the MLE obtained earlier is still the global maximum because there is only one solution to the gradient equations.

Problem 4. Let $X_1, \dots, X_n$ be a sample from the inverse Gaussian pdf,

$$f(x|\mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\{-\lambda(x-\mu)^2/(2\mu^2 x)\}, \quad x > 0.$$

Show that the MLEs of μ and λ are

$$\hat{\mu}_{MLE} = \bar{X}_n, \quad \hat{\lambda}_{MLE} = n \bigg/ \sum_{i=1}^{n} \left(\frac{1}{X_i} - \frac{1}{\bar{X}_n}\right).$$

Solution. First we write out the log-likelihood function

$$l(\mu, \lambda) = -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log\lambda - \frac{3}{2}\sum_{i=1}^{n}\log X_i - \frac{\lambda}{2\mu^2}\sum \frac{(X_i-\mu)^2}{X_i}.$$

Calculate the gradient of the above log-likelihood and solve its roots to get the MLE:

$$\begin{cases} \dfrac{\partial l}{\partial \mu} &= \dfrac{\lambda}{\mu^3}\sum \dfrac{(X_i-\mu)^2}{X_i} + \dfrac{\lambda}{\mu^2}\sum \dfrac{(X_i-\mu)}{X_i} = \dfrac{\lambda}{\mu^2}\sum_{i=1}^{n} X_i - \dfrac{n\lambda}{\mu^2} = 0, \\ \dfrac{\partial l}{\partial \lambda} &= \dfrac{n}{2\lambda} - \dfrac{1}{2\mu^2}\sum \dfrac{(X_i-\mu)^2}{X_i} = 0. \end{cases} \tag{2}$$

Solving the first equation in (2), we immediately get

$$\hat{\mu}_{MLE} = \bar{X}_n.$$

Plug this result in the second equation, and we have

$$\hat{\lambda}_{MLE} = n \bigg/ \left\{\frac{1}{2\bar{X}_n^2}\sum \frac{(X_i-\bar{X}_n)^2}{X_i}\right\} \overset{\text{Expand and simplify}}{=} n \bigg/ \sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}_n}\right).$$

Problem 5. Let X have the distribution given below, in which $0 < \theta < 1/6$:

value	1	2	3	4
probability	2θ	θ	3θ	$1-6\theta$

Let $X_1, \dots, X_n$ be i.i.d observations drawn from the distribution above, each of the type $X_i = x_i$ where $x_i = \{1, 2, 3, 4\}$. Express the likelihood function $L(\theta)$ and the log-likelihood $l(\theta)$ in terms of

$$\begin{aligned} n_1 &= \text{ number of samples equal to 1,} \\ n_2 &= \text{ number of samples equal to 2,} \\ n_3 &= \text{ number of samples equal to 3,} \\ n_4 &= \text{ number of samples equal to 4.} \end{aligned}$$

Note that $n = n_1 + n_2 + n_3 + n_4$. Find $\hat{\theta}_{MLE}$ in terms of the above n_i's.

Solution. Denote the indicator function by $\mathbb{1}\{\text{condition}\}$. Then

$$n_1 = \sum_{i=1}^{n} \mathbb{1}\{X_i = 1\},$$
$$n_2 = \sum_{i=1}^{n} \mathbb{1}\{X_i = 2\},$$
$$n_3 = \sum_{i=1}^{n} \mathbb{1}\{X_i = 3\},$$
$$n_4 = \sum_{i=1}^{n} \mathbb{1}\{X_i = 4\}.$$

Therefore the likelihood function can be written as

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} p(X_i|\theta) = \prod_{i=1}^{n} (2\theta)^{\mathbb{1}\{X_i=1\}} \theta^{\mathbb{1}\{X_i=2\}} (3\theta)^{\mathbb{1}\{X_i=3\}} (1-6\theta)^{\mathbb{1}\{X_i=4\}} \\ &= (2\theta)^{\sum_{i=1}^{n} \mathbb{1}\{X_i=1\}} \theta^{\sum_{i=1}^{n} \mathbb{1}\{X_i=2\}} (3\theta)^{\sum_{i=1}^{n} \mathbb{1}\{X_i=3\}} (1-6\theta)^{\sum_{i=1}^{n} \mathbb{1}\{X_i=4\}} \\ &= (2\theta)^{n_1} \theta^{n_2} (3\theta)^{n_3} (1-6\theta)^{n_4}. \end{aligned}$$

The log-likelihood is

$$l(\theta) = n_1 \log(2\theta) + n_2 \log\theta + n_3 \log(3\theta) + n_4 \log(1-6\theta).$$

Taking the first derivative of the log-likelihood to get

$$\frac{\partial l}{\partial \theta} = \frac{n_1+n_2+n_3}{\theta} - \frac{6n_4}{1-6\theta}.$$

It is easy to see that $\frac{\partial l}{\partial \theta} > 0$ if and only if $\theta < \frac{n_1+n_2+n_3}{6(n_1+n_2+n_3+n_4)}$. Therefore,

$$\hat{\theta}_{MLE} = \frac{n_1+n_2+n_3}{6(n_1+n_2+n_3+n_4)} = \frac{n_1+n_2+n_3}{6n}.$$

Problem 6. In Lecture 3, we looked at the session duration observations from Design A of the Youtube A/B testing, and computed the MM estimates and approximated their SE using the bootstrap plug-in method with the help of random samplers on the `R` shiny app. Now let's look at Design B.

(1) There are also 200 observations of session duration (hours) from Design B. The values of these observations are summarized in Figure 1, which indicates a right-skewed distribution. We also use Gamma(α, β) as the population distribution. For this sample, $\bar{X}_n = 0.9537$ and $\frac{1}{n}\sum_{i=1}^{200}(X_i - \bar{X}_n)^2 = 0.4103$. Please give the MM estimates $\hat{\alpha}_{MM}$ and $\hat{\beta}_{MM}$.

(2) To estimate SE for $\hat{\alpha}_{MM}$ and $\hat{\beta}_{MM}$, we use the bootstrap method and pretend that $\hat{\alpha}_{MM}$ and $\hat{\beta}_{MM}$ are the true population parameters. Then the shiny app can be used to repeat the test many times (which would be expensive to do in real life), and calculate the MM estimates for α and β from each sample of each test. Click on this

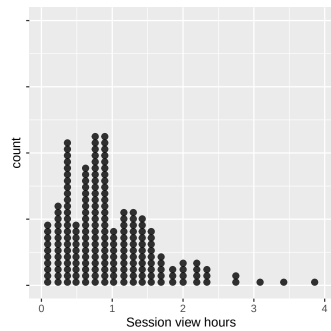


FIGURE 1. The dotplot of session duration observations from Design B

link to the app, and choose different number of repeats. You can start from 8 tests, and report the SE for $\hat{\alpha}_{MM}$. Then gradually increase the number to 50, 100, 400, 800, 1000, and see how the value of SE changes.

(3) Every time the sliding widget for the number of samples is moved, the app will re-generate all the samples. Slide the widget multiple times to see whether values of SE for 400, 800 and 1000 change. If SE stabilizes after 400, we can report that value as our SE for $\hat{\alpha}_{MM}$.

Solution.

(1) The MM estimates are

$$\hat{\alpha}_{MM} = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{200}(X_i - \bar{X}_n)^2} = \frac{0.9537^2}{0.4103} = 2.2168,$$

$$\hat{\beta}_{MM} = \frac{\bar{X}_n}{\frac{1}{n}\sum_{i=1}^{200}(X_i - \bar{X}_n)^2} = \frac{0.9537^2}{0.4103} = 2.3244.$$

(2) Gradually increasing the number of samples while fixing the sample size $n = 200$, I got the following SE values (it can vary from your values by 0.02 at most):

# of samples	SE, 1st slide	SE, 2nd slide	SE, 3rd slide	SE, 4th slide
8	0.23	0.3	0.16	0.29
50	0.29	0.3	0.24	0.25
100	0.28	0.26	0.22	0.27
400	0.27	0.27	0.25	0.27
800	0.27	0.27	0.28	0.27
1000	0.27	0.27	0.27	0.27

(3) Yes my SE values started stabilizing after the number of samples reached 400.

Problem 7. Copy and paste the following `R` script which reads in the all session view hours shown in Figure 1:

```
B_view_hours <- c(1.17, 1.417, 0.419, 1.426, 1.203, 0.136, 0.57,
    0.844, 0.78, 1.014, 0.75, 0.075, 0.318, 0.584, 1.632, 1.738,
    0.624, 0.609, 0.72, 1.504, 1.583, 0.429, 0.866, 1.214, 0.947,
    1.546, 1.148, 0.345, 1.384, 0.884, 1.648, 0.151, 0.143, 1.203,
    0.699, 1.516, 2.254, 0.297, 0.767, 1.147, 0.569, 3.088, 0.617,
    1.575, 1.211, 0.917, 0.025, 0.979, 0.277, 2.751, 3.859, 1.314,
    1.073, 0.424, 0.119, 0.726, 0.322, 0.854, 0.844, 0.238, 0.991,
    1.366, 1.349, 0.387, 0.957, 0.398, 0.993, 0.678, 0.431, 0.8,
    0.719, 0.173, 0.399, 0.201, 2.276, 0.702, 1.911, 0.481, 0.541,
    1.051, 0.82, 1.285, 0.322, 2.134, 0.775, 1.565, 2.399, 0.576,
    0.262, 0.612, 0.932, 0.747, 0.936, 1.748, 0.81, 1.225, 1.277,
    1.386, 0.431, 0.508, 0.146, 0.591, 0.97, 1.104, 0.418, 1.501,
    0.702, 0.108, 0.51, 2.014, 0.426, 0.87, 1.36, 0.864, 0.412,
    0.802, 0.215, 0.243, 0.658, 0.671, 1.31, 0.13, 0.223, 1.294,
    1.601, 0.966, 0.415, 0.529, 0.481, 0.416, 1.347, 1.345, 0.192,
    1.213, 0.938, 0.317, 0.341, 0.78, 1.945, 1.004, 1.407, 0.579,
    0.621, 0.842, 0.307, 0.227, 0.842, 0.836, 1.386, 0.841, 1.201,
    0.438, 1.425, 2.186, 0.327, 0.197, 0.436, 0.606, 1.797, 0.809,
    0.587, 2.253, 1.294, 0.885, 0.82, 1.235, 0.31, 2.275, 1.521,
    0.719, 2.741, 0.256, 0.614, 0.747, 1.367, 1.333, 2.02, 0.769,
    1.636, 0.864, 0.131, 0.648, 0.846, 0.304, 0.751, 1.814, 0.838,
    0.82, 0.935, 0.58, 0.907, 0.832, 1.161, 0.517, 0.775, 1.482,
    0.458, 1.449, 3.422, 2.058)
```

Follow Example 1 of Lecture 4 and compute in `R` the MLE estimates $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$ for Design B. Also calculate the 95% bootstrap CIs for α and β separately using the MLE (*You need to calculate the Fisher information matrix first*).

Solution. From `R`, we have $\hat{\alpha}_{MLE} = \mathbf{2.172}$ and $\hat{\beta}_{MLE} = \mathbf{2.278}$. We demonstrate using two approaches.

In class, we showed that the individual 95% bootstrap CIs for α and β are

$$\hat{\alpha}_{MLE} \pm \frac{z_{0.05/2}}{\sqrt{n\{\text{trigamma}(\alpha)\text{-}1/\alpha\}}}$$
$$\hat{\beta}_{MLE} \pm \frac{z_{0.05/2}\sqrt{\beta^2 * \text{trigamma}(\alpha)}}{\sqrt{n\{\alpha * \text{trigamma}(\alpha)\text{-}1\}}}.$$

Therefore,

$$95\% \text{ CI for } \alpha: \quad [\mathbf{1.775,\ 2.569}],$$
$$95\% \text{ CI for } \beta: \quad [\mathbf{1.809,\ 2.746}].$$

(1) Approach 1: use `uniroot()`

Define the target function:

```
alpha_target_func <- function(alpha, data){
  n <- length(data)
  res <- n*log(alpha)-n*digamma(alpha)-n*log(mean(data)) +
      sum(log(data))
  return(res)
}
```

Then find the root and obtain the MLEs:

```
# uniroot() searches in a specified interval for a root (i.e.,
    zero) of a function
solution <- uniroot(alpha_target_func, interval = c(2,2.5), data =
    B_view_hours)
alpha_MLE <- solution$root
beta_MLE <- alpha_MLE/mean(A_view_hours)
c('alpha_MLE'=alpha_MLE, 'beta_MLE'=beta_MLE)
> alpha_MLE beta_MLE
>  2.172311 2.278298
```

(2) Approach 2: use `optim()`

First define the log-likelihood function:

```
log_lik_for_optim <- function(par, data){
  alpha = par[1]; beta = par[2]
  n = length(data)
  log_lik_result = n*alpha*log(beta) - n*log(gamma(alpha)) +
      (alpha-1)*sum(log(data)) - beta*sum(data)
  return(log_lik_result)
}
```

Then find the maximum

```
res <- optim(par = c(2.104, 2.298), log_lik_for_optim, data =
    B_view_hours,
          control = list(trace = TRUE, fnscale = -1)) #
              minimization to maximization
res$par
[1] 2.172098 2.278049
```

Problem 8. Suppose $X_1, \dots, X_n \sim N(\mu, \sigma^2)$. In Lab 2, you have proved that the MM estimators for μ and σ are

$$\hat{\mu}_{MM} = \bar{X}_n, \quad \hat{\sigma}_{MM} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}.$$

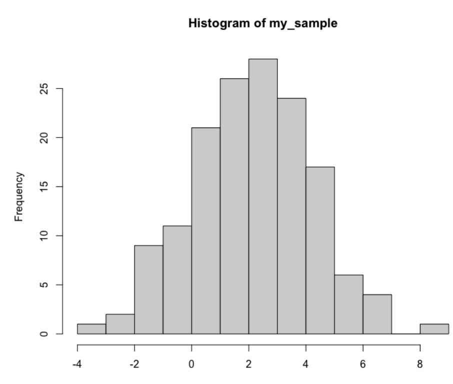
Now you are going to implement the bootstrap method in `R` to approximate the SE for $\hat{\mu}_{MM}$ and $\hat{\sigma}_{MM}$.

(1) Start from a population $N(2,4)$. Generate $n = 150$ i.i.d observations from this population and plot your observations in a histogram.
(2) Calculate your MM estimate for μ and σ.
(3) We have showed in Lecture 1 that the theoretical SE for $\hat{\mu}_{MM}$ is $\text{SE}(\hat{\mu}_{MM}) = \sigma/\sqrt{n}$. However, the theoretical SE for $\hat{\sigma}_{MM}$ is not easy to obtain, which is why the bootstrap method is used frequently to approximate SE for more complex estimators.
   (a) This way of bootstrapping is outlined on Page 11 of Lecture 3 slides: plug your sample variance in $\sigma/\sqrt{n}$ and get an estimate of SE for $\hat{\mu}_{MM}$.
   (b) Suppose the 150 observations you generated in (1) are collected in a real survey (which means you don't know about the true population parameters). Follow the steps outlined on Page 10 of the Lecture 3 slides to approximate the SE for $\hat{\mu}_{MM}$ and $\hat{\sigma}_{MM}$. (*Note you need to generate at least 500 samples, each of size ~~100~~ **150** from the Normal density whose parameters are given by $\hat{\mu}_{MM}$ and $\hat{\sigma}_{MM}$. Compute MM estimates of μ and σ from each sample. You now have ≥ 500 simulated estimates of μ and σ. Plot a histogram of all $\hat{\mu}_{MM}$ and $\hat{\sigma}_{MM}$ separately. The SEs for $\hat{\mu}_{MM}$ and $\hat{\sigma}_{MM}$ are simply the SEs of these* 500 *MM estimates.*)
   (c) Recall that you do know the true values of μ and σ. Mark each of them on the horizontal axis of the appropriate histograms. Is either of them in a surprising place?
   (d) Compare the SEs for $\hat{\mu}_{MM}$ you obtain in (a) and (b). Are the values close to each other?

Solution.

(1) Generate 150 i.i.d $N(2,4)$ samples in `R` and plot in a histogram:

```
my_sample <- rnorm(150, mean=2, sd=2)
hist(my_sample)
```



(2) The MM estimates are simply the sample mean and the sample variance (biased) which have values of **2.120** and **2.107** respectively:

```
mu_MM = mean(my_sample)
sigma_MM = sd(my_sample)
c('mu_MM' = mu_MM, 'sigma_MM' = sigma_MM)
>  mu_MM sigma_MM
> 2.120204 2.106537
```

(3) (a) $\mathrm{SE}(\hat{\mu}_{MM}) \approx \sigma_{MM}/\sqrt{n} = \mathbf{0.171998}$.

(b) The R code to repeat the survey 500 times is attached below:

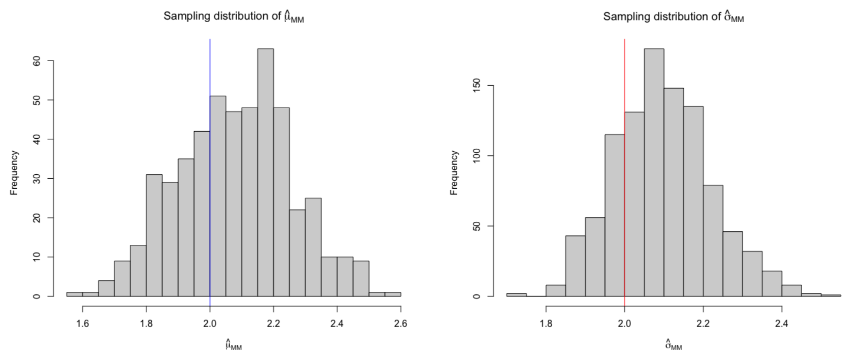
```
B=500

# the function is used to find the MM estimates of
    each size-150 sample
find_MM <- function(){
    resample <- rnorm(150, mean=mu_MM, sd=sigma_MM) #
        mimicking the sampling process in a real
        survey
    mu_MM_resample <- mean(resample) # Mean estimate
        from this sample
    sigma_MM_resample <- sd(resample)
    return(c(mu_MM_resample, sigma_MM_resample))
}

# we will run the above function 500 times
matrix <- replicate(B, find_MM())
SE_mu_MM <- sd(matrix[1,]) # mu estimates from 500
    samples
SE_sigma_MM <- sd(matrix[2,]) # mu estimates from
    500 samples
c('SE_mu_MM' = SE_mu_MM, 'SE_sigma_MM'=SE_sigma_MM)
>  SE_mu_MM SE_sigma_MM
>  0.1784958  0.1240255
```

Therefore, $\mathrm{SE}(\hat{\mu}_{MM}) \approx \mathbf{0.1784958}$ and $\mathrm{SE}(\hat{\sigma}_{MM}) = \mathbf{0.1240255}$.

(c) Plot the histograms with the true parameter values:

```
hist(matrix[1,], breaks=20,
    xlab=expression(hat(mu)[MM]),
    main=expression(paste("Sampling distribution of
    ", hat(mu)[MM])))
abline(v=2, col='blue')
```

```
hist(matrix[2,], breaks=20,
    xlab=expression(hat(sigma)[MM]),
    main=expression(paste("Sampling distribution of
    ", hat(sigma)[MM]))) # sigma estimates from 500
    samples
abline(v=2, col='red')
```



They are both close to the center of the sampling distributions, although true value of σ has more deviation. Overall, the results are not surprising.

(d) From (a) and (b), I got 0.171998 and 0.1784958 for SE($\hat{\mu}_{MM}$) respectively. They are fairly close to each other.

Problem 9. Let $X_1, \ldots, X_n$ be a random sample from the pdf

$$f(x|\theta) = \theta x^{-2}, 0 < \theta \leq x < \infty.$$

(1) What is the sufficient statistic for θ?
(2) Find the MM estimator of θ.
(3) Find the MLE of θ.

Solution.

(1) The likelihood function can be written as

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta) = \theta^n \prod_{i=1}^{n} X_i^2 \cdot \mathbb{1}\{\theta \leq X_{(1)}\}.$$

By the Fisher-Neyman Factorization theorem, we know the sufficient statistic for θ is $X_{(1)}$.

(2) Since any kth moment of the population is

$$\mu_k = \int_\theta^\infty \theta x^{-2+k} dx = \infty,$$

we can't find an MM estimator following the common approach. However, if we take the 1/2th moment,

$$\mu_{1/2} = \int_\theta^\infty \theta x^{-1.5} dx = -2x^{-1/2}|_\theta^\infty = 2\theta^{-1/2}.$$

Therefore, the corresponding MM estimator for θ is

$$\hat{\theta}_{MM} = 1 \Bigg/ \left(\frac{1}{2n} \sum_{i=1}^{n} \sqrt{X_i} \right)^2 .$$

(3) Since $L(\theta)$ is a monotonically increasing function of θ, the maximum is attained at $\hat{\theta}_{MLE} = X_{(1)}$.

Problem 10. Suppose $X_1, \ldots, X_n$ are i.i.d observations from a population with pmf

$$P(X = x|\theta) = \theta^x (1-\theta)^{1-x}, \quad , x = 0 \text{ or } 1, \quad 0 \le \theta \le \frac{1}{2}$$

.

(1) Find the MM estimator and MLE of θ.
(2) Try to find the MSE of each of the estimators. (*See Page 29 of Lecture 1 slides for the definition of MSE.*)
(3) Which estimator is preferred? Justify your choice.

Solution.

(1) The first moment of this population is

$$\mu = EX = \theta,$$

and thus $\hat{\theta}_{MM} = \bar{X}_n$.

To obtain the MLE, write out the log-likelihood function

$$l(\theta) = \left(\sum_{i=1}^{n} X_i \right) \log \theta + \left(n - \sum_{i=1}^{n} X_i \right) \log(1-\theta).$$

Take the first derivative, and we know

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^{n} X_i \Big/ \theta - \left(n - \sum_{i=1}^{n} X_i \right) \Big/ (1-\theta) > 0$$

is equivalent to $\theta < \bar{X}_n$. Note that $0 \le \theta \le 1/2$, which gives

$$\hat{\theta}_{MLE} = \min\{\bar{X}_n, 1/2\}.$$

(2) By the definition of MSE,

$$\begin{aligned} \mathrm{MSE}(\hat{\theta}_{MM}) =& E(\hat{\theta}_{MM} - \theta)^2 \\ =& \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n} - \theta \right)^2 \theta^k (1-\theta)^{n-k} = \frac{\theta(1-\theta)}{n}. \end{aligned} \tag{3}$$

For the MLE, let's first examine the sampling distribution. We know that all possible values of $\hat{\theta}_{MLE}$ are $1/n, 2/n, \ldots, \lfloor n/2 \rfloor / n, 1/2$, in which $\lfloor \cdot \rfloor$ is the floor function.

$$P(\hat{\theta}_{MLE} = p) = P(\hat{\theta}_{MLE} = p, \bar{X}_n \ge 1/2) + P(\hat{\theta}_{MLE} = p, \bar{X}_n < 1/2)$$

$$= \begin{cases} P(\bar{X}_n \geq 1/2), & p = 1/2, \\ P(\bar{X}_n = k), & p = 1/n, \ldots, \lfloor n/2 \rfloor / n, \end{cases}$$

$$\overset{\sum_{i=1}^n X_i \text{ is binomial}}{=} \begin{cases} \sum_{k=\lfloor n/2 \rfloor+1}^n \binom{n}{k} \theta^k (1-\theta)^{n-k}, & p = 1/2, \\ P(\bar{X}_n = k), & p = 1/n, \ldots, \lfloor n/2 \rfloor / n. \end{cases}$$

Therefore,

$$\begin{aligned} &\text{MSE}(\hat{\theta}_{MLE}) = E(\hat{\theta}_{MLE} - \theta)^2 \\ &= \sum_{k=\lfloor n/2 \rfloor+1}^{n} \binom{n}{k} \left(\frac{1}{2} - \theta\right)^2 \theta^k (1-\theta)^{n-k} + \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k} \left(\frac{k}{n} - \theta\right)^2 \theta^k (1-\theta)^{n-k}. \end{aligned} \tag{4}$$

(3) From equations (3) and (4), we have

$$\begin{aligned} \text{MSE}(\hat{\theta}_{MM}) - \text{MSE}(\hat{\theta}_{MLE}) &= \sum_{k=\lfloor n/2 \rfloor+1}^{n} \binom{n}{k} \left[\left(\frac{k}{n} - \theta\right)^2 - \left(\frac{1}{2} - \theta\right)^2 \right] \theta^k (1-\theta)^{n-k} \\ &= \sum_{k=\lfloor n/2 \rfloor+1}^{n} \binom{n}{k} \left(\frac{k}{n} - \frac{1}{2}\right) \left(\frac{k}{n} + \frac{1}{2} - 2\theta\right) \theta^k (1-\theta)^{n-k}. \end{aligned}$$

The facts that $k/n > 1/2$ when $k \geq \lfloor n/2 \rfloor + 1$ and $\theta \leq 1/2$ imply that every term in the sum is positive; that is,

$$\text{MSE}(\hat{\theta}_{MM}) - \text{MSE}(\hat{\theta}_{MLE}) > 0$$

for any θ in [0,1/2]. Therefore, MLE is better than MM estimator because it always has a smaller MSE.