1 Question 1

Conditioned upon X, $Y_i \sim \mathcal{N}(X_i'\theta, \sigma^2)$. This tells us that $\frac{Y_i - X_i'\theta}{\sigma} \sim \mathcal{N}(0,1)$. As a result, the density of Y_i given X_i is given by: $\frac{1}{\sigma}\phi(\frac{z-X_i'\theta}{\sigma})$.

1.1 What is the (condition) likelihood function? What is the (conditional) log-likelihood function?

$$\begin{aligned}
\ell_n(\theta) &= \prod_{i=1}^{n} p_\theta(y_i|x_i) \\
&= \prod_{i=1}^{n} \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\theta}{\sigma}\right) \\
&= \frac{1}{\sigma^n}\prod_{i=1}^{n} \phi\left(\frac{y_i - x_i'\theta}{\sigma}\right)
\end{aligned}$$

$$L_n(\theta) = -\log(\sigma) + \frac{1}{n}\sum_{i=1}^{n} \log\left(\phi\left(\frac{y_i - x_i'\theta}{\sigma}\right)\right)$$

1.2 Find the ML estimators

$$\begin{aligned}
\frac{\partial L_n(\theta)}{\partial \theta_j} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\phi'\left(\frac{y_i - x_i'\theta}{\sigma}\right)(-x_{i,j})}{\phi\left(\frac{y_i - x_i'\theta}{\sigma}\right)} = 0 \\
\frac{\partial L_n(\theta)}{\partial \sigma} &= \frac{-1}{\sigma} + \frac{1}{n}\sum_{i=1}^{n} \frac{\phi'\left(\frac{y_i - x_i'\theta}{\sigma}\right)\left(\frac{-(y_i - x_i'\theta)}{\sigma^2}\right)}{\phi\left(\frac{y_i - x_i'\theta}{\sigma}\right)} = 0
\end{aligned}$$

Let us simplify the notation by using

$$A_i = \frac{\phi'\left(\frac{y_i - x_i'\theta}{\sigma}\right)}{\phi\left(\frac{y_i - x_i'\theta}{\sigma}\right)} = \frac{y_i - x_i'\theta}{\sigma}$$

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i - x_i'\theta}{\sigma}(-x_{i,j})$$

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i - x_i'\theta}{\sigma}(y_i - x_i'\theta)$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2$$

$$0 = \sum_{i=1}^{n} x_{i,j} u_i \quad \forall j$$

Let us denote the matrix containing X_i on the i^{th} row as $\boldsymbol{X}$. Then our condition simplifies to:

$$0 = \boldsymbol{X}\boldsymbol{u}$$

This is exactly the orthogonality condition found in linear regression, and as such has the same solution.

$$\widehat{\theta}_n = (X'X)^{-1}X'Y$$

$$\widehat{\sigma^2}_n = \frac{1}{n}\sum_{i=1}^{n} u_i^2$$

1.3 (c)

Consider testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Describe how you would carry this out using each of the three hypothesis testing methods described in class. How do they compare with the Wald tests and LM tests described earlier in class using ordinary least squares?

Let $f(\theta)$ denote the log-likelihood function evaluated at θ. Let $g(\theta)$ denote the gradient of the log-likelihood function evaluated at θ.

- Wald Test

  Let us calculate an estimator of the variance of $\widehat{\theta}_n$, $\widehat{\Omega}_n$. We may estimate this using an estimate of the outer-product gradient of the log-likelihood function.

$$\widehat{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n} \nabla L_n(\widehat{\theta}_n)\nabla L_n(\widehat{\theta}_n)$$

Then a Wald-Test would use the fact that $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \widehat{\Omega}_n)$. As $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$ and $\widehat{\Omega}_n \xrightarrow{p} \Omega$. By our choice of construction of $\widehat{\Omega}_n$, it is positive definite, and therefore has a choleksy decomposition. Then we may write $\widehat{\Omega}_n = \widehat{F}_n \widehat{F}_n'$ where $\widehat{F}_n$ is a lower-triangular matrix.

This test would take the form of $\mathbb{1}_{\{T_n > z_{1-\alpha}\}}$ where $T_n := \widehat{F}_n \sqrt{n}(\widehat{\theta}_n - \theta_0)$, and $z_{1-\alpha}$ is the critical value for the normal distribution.

The question of interest is whether or not $\widehat{\Omega}_n \xrightarrow{p} \sigma^2 (X'X)^{-1}$, the variance estimator used in OLS.

$$\widehat{\Omega}_n \xrightarrow{p} \Omega$$

- LR Test

Let us define a new estimator of the variance, $\tilde{\sigma^2}_n$, which is our estimate of the variance when we hold θ fixed at θ_0.

$$\tilde{\sigma^2}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i^2$$

Define the log-likelihood function as $f(\theta, \sigma^2)$. We wish to compare the constrained log-likelihood, $f(\theta_0, \tilde{\sigma^2}_n)$ to the unconstrained log-likelihood: $f(\widehat{\theta}_n, \widehat{\sigma^2}_n)$.

It is known that

$$2(f(\widehat{\theta}_n, \widehat{\sigma^2}_n) - f(\theta_0, \tilde{\sigma^2}_n)) \xrightarrow{d} \chi^2(k+1)$$

Therefore we can define a test to take the form of $\mathbb{1}_{\{T_n > c_{1-\alpha}\}}$ where $T_n := 2(f(\widehat{\theta}_n, \widehat{\sigma^2}_n) - f(\theta_0, \tilde{\sigma^2}_n))$ and $c_{1-\alpha}$ is the critical value of a chi-squared distribution with $k+1$ degrees of freedom.

We have not examined an analog for this test in OLS.

2 Question 2

Let $X_1, ..., X_n$ be an i.i.d. sequence of random variables with distribution $U(\theta, 2\theta)$ with $\theta > 0$ i.e. uniformly distributed on the interval $[\theta, 2\theta]$.

2.1 (a)

Show that the maximum likelihood estimator of θ is given by

$$\widehat{\theta}_n = \frac{1}{n} \max_{1 \le i \le n} X_i$$

The density of X_i is given by: $\frac{1}{\theta}\mathbb{1}_{\{X_i\in[\theta,2\theta]\}}$. The likelihood function is then given by:

$$\begin{aligned}\ell_n(\theta) &= \prod_{i=1}^{n}\frac{1}{\theta}\mathbb{1}_{\{X_i\in[\theta,2\theta]\}}\\ &= \frac{1}{\theta^n}\mathbb{1}_{\{X_i\in[\theta,2\theta] \quad \forall i\}}\end{aligned}$$

The log-likelihood function is then given by:

$$L_n(\theta) = -n\log(\theta) + \log(\mathbb{1}_{\{X_i\in[\theta,2\theta]\}})$$

This function is decreasing in θ, provided that all $X_i \in [\theta, 2\theta]$. This means that we wish to choose the smallest possible θ value that maintains all $\theta \le X_i \le 2\theta$. The upper bound is the relevant bound, so we allow for $\widehat{\theta}_n = \frac{1}{2}\max_{1\le i\le n} X_i$.

2.2 (b)

Let us determine the distribution of $Z = \max X_i$. Let $z \in [\theta, 2\theta]$.

$$\begin{aligned}\Pr(Z \le z) &= \Pr(\max X_i \le z)\\ &= \cap \Pr(X_i \le z)\\ &= \left(\frac{z-\theta}{\theta}\right)^n\\ f_z(z) &= n\left(\frac{z-\theta}{\theta}\right)^{n-1}\frac{1}{\theta}\\ \mathbb{E}\left[Z\right] &= \frac{n}{\theta}\int_{\theta}^{2\theta} z\left(\frac{z}{\theta}-1\right)^{n-1}dz\\ &= n\int_0^1 (\theta u+\theta)u^{n-1}du\\ &= n\int_0^1 \theta u^n + \theta u^{n-1}du\\ &= \theta\frac{2n+1}{n+1}\end{aligned}$$

From this we can see that:

$$\mathbb{E}\left[\widehat{\theta}_n\right] = \frac{1}{2}\mathbb{E}\left[\max X_i\right] = \theta\frac{2n+1}{2n+2} \neq \theta$$

So $\widehat{\theta}_n$ is not an unbiased estimator of θ.

2.3 (c)

Prove that $\widehat{\theta}_n$ is a consistent estimator of θ.

$$
\begin{aligned}
\Pr(\left|\widehat{\theta}_n - \theta\right| < \epsilon) &= \Pr(\left|\frac{1}{2}\max X_i - \theta\right| < \epsilon) \\
&= \Pr(\theta - \frac{1}{2}\max x_i < \epsilon) \\
&= \Pr(\max X_i > 2(\theta - \epsilon)) \\
&= 1 - \Pr(\max X_i < 2(\theta - \epsilon)) \\
&1 - \left(\frac{\theta - 2\epsilon}{\theta}\right)^n \\
&\to 1
\end{aligned}
$$

2.4 (d)

Show that $n(\theta - \widehat{\theta}_n)$ converges in distribution to an exponential distribution with parameter λ as $n \to \infty$. Express λ in terms of θ.

Firstly we can note that the Probability that $\theta - \widehat{\theta}_n < 0$ is zero. There is no support for $X_i > \theta$, so the maximum cannot be greater than θ.

$$
\begin{aligned}
\Pr(n(\theta - \widehat{\theta}_n) \leq x) &= \Pr(\theta - \widehat{\theta}_n \leq \frac{x}{n}) \\
&= 1 - \left(\frac{\theta - \frac{2x}{n}}{\theta}\right)^n \\
&= 1 - \left(1 - \frac{2x}{n\theta}\right)^n \\
&\to 1 - \exp(-\frac{2x}{\theta})
\end{aligned}
$$

If we let $\lambda = \frac{\theta}{2}$, then we see that $n(\theta - \widehat{\theta}_n) \xrightarrow{d} \exp(\frac{\theta}{2})$.

2.5 (e)

Construct an (approximate) 95% confidence interval for θ. Justify your answer.

We are certain that $\theta > \frac{1}{2}\max X_i$, so this is a logical lower bound for our interval. We wish to use part (d) to determine the upper bound.

We are interested in an interval $C_n := [\frac{1}{2}\max X_i, c]$ such that $\Pr(\theta \in C_n) \leq .95$.

$$
\begin{aligned}
\Pr(\theta < c) &= \Pr(n\theta < nc + n\widehat{\theta}_n - n\widehat{\theta}_n) \\
&= \Pr(n(\theta - \widehat{\theta}_n) < n\left(c - \widehat{\theta}_n\right)) \\
&\to 1 - \exp\left(-\frac{2n(c - \widehat{\theta}_n)}{\theta}\right) \\
&\leq 1 - \exp\left(-\frac{2n(c - \widehat{\theta}_n)}{\widehat{\theta}_n}\right)
\end{aligned}
$$

Since we know that $\widehat{\theta}_n < \theta$. Setting this probability equal to .95.

$$\begin{aligned} 1 - \exp\left(-\frac{2n(c - \widehat{\theta}_n)}{\widehat{\theta}_n}\right) &= .95 \\ \exp\left(-\frac{2n(c - \widehat{\theta}_n)}{\widehat{\theta}_n}\right) &= .05 \\ \frac{2n(c - \widehat{\theta}_n)}{\widehat{\theta}_n} &= -\log .05 \\ c &= \widehat{\theta}_n - \frac{\widehat{\theta}_n \log(.05)}{2n} \end{aligned}$$

From this we know that: $\Pr(\theta \in [\widehat{\theta}_n, c]) \leq .95$. as $n \to \infty$.

3 Question 3

3.1 (a)

Suggest an unbiased estimator $\tilde{\theta}_n$ of θ. Justify your answer.

Using a Method of Moments estimator, we may set the sample mean equal to the population mean.

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} X_i &= \mathbb{E}\left[X\right] \\ \overline{X}_N &= \frac{3\theta}{2} \\ \tilde{\theta}_n &= \frac{2\overline{X}_N}{3} \end{aligned}$$

To show this is an unbiased estimator:

$$\begin{aligned} \mathbb{E}\left[\tilde{\theta}_n\right] &= \frac{2}{3}\mathbb{E}\left[\overline{X}_N\right] \\ &= \frac{2}{3}\frac{3\theta}{2} \\ &= \theta \end{aligned}$$

```
using Random
using Latexify
using Distributions

```

```
thetapos = [0.5,1,10]
npos = [2,5,20,100]

Random.seed!( 235711 )

#This just allocates a bunch of arrays to fit all the data we're asked to
↪  simulate
X = zeros( length(thetapos), length(npos) )
hatABS = zeros( length(thetapos), length(npos) )
tildeABS = zeros( length(thetapos), length(npos) )
hatMSE = zeros( length(thetapos), length(npos) )
tildeMSE = zeros( length(thetapos), length(npos) )
ind = zeros( length(thetapos), length(npos) )

tempx = zeros(10000,maximum(npos))
tempHatTheta = zeros(10000)
tempTildeTheta = zeros(10000)

for i in 1:length(thetapos)
    theta = thetapos[i]
    dist = Uniform( theta, 2*theta)
    for j in 1:length(npos)
        n = npos[j]
        for k in 1:10000
            tempx[k,1:n] = rand( dist, n)
            tempHatTheta[k] = maximum(tempx[k,1:n])/2.0
            tempTildeTheta[k] = (2.0/3.0)*mean(tempx[k,1:n])
        end

        tempHatABS = abs.( tempHatTheta .- theta)
        hatABS[i,j] = mean( tempHatABS )

        tempTildeABS = abs.( tempTildeTheta .- theta)
        tildeABS[i,j] = mean(tempTildeABS)

        tempHATMSE = tempHatABS .* tempHatABS
        hatMSE[i,j] = mean(tempHATMSE)

        tempTildeMSE = tempTildeABS .* tempTildeABS
        tildeMSE[i,j] = mean( tempTildeMSE)

        temp = tempTildeMSE .> tempHATMSE
        ind[i,j] = mean(temp)
    end
end
```

```
latexify( hatABS )
latexify( tildeABS )
latexify( hatMSE )
latexify( tildeMSE )
latexify( ind )
```

Let the rows of the tables vary on θ, and the columns to vary on n.

$$\mathbb{E}\left[\left|\widehat{\theta}_n - \theta\right|\right] = \begin{bmatrix} 0.08406155 & 0.04192225 & 0.01200721 & 0.00244757 \\ 0.16782184 & 0.08347129 & 0.02412170 & 0.00491011 \\ 1.65429116 & 0.83625597 & 0.23698989 & 0.04959539 \end{bmatrix} \tag{1}$$

$$\mathbb{E}\left[\left|\tilde{\theta}_n - \theta\right|\right] = \begin{bmatrix} 0.055846444 & 0.03451907 & 0.01698833 & 0.00760587 \\ 0.111192510 & 0.06964597 & 0.03418168 & 0.01506232 \\ 1.119239886 & 0.69265960 & 0.34332960 & 0.15368077 \end{bmatrix} \tag{2}$$

$$\mathbb{E}\left[\left(\widehat{\theta}_n - \theta\right)^2\right] = \begin{bmatrix} 0.01024438 & 0.00291720 & 0.00027466 & 1.18647824 \times 10^{-5} \\ 0.04150057 & 0.01226166 & 0.00110024 & 4.83207145 \times 10^{-5} \\ 4.13854889 & 1.18343059 & 0.10500403 & 0.00480661 \end{bmatrix} \tag{3}$$

$$\mathbb{E}\left[\left(\tilde{\theta}_n - \theta\right)^2\right] = \begin{bmatrix} 0.00466503 & 0.0018422 & 0.00045143 & 9.03422956 \times 10^{-5} \\ 0.01849642 & 0.0074670 & 0.00181460 & 0.00036373 \\ 1.86594666 & 0.7375063 & 0.18430868 & 0.03693727 \end{bmatrix} \tag{4}$$

$$\Pr(\left|\tilde{\theta}_n - \theta\right| < \left|\widehat{\theta}_n - \theta\right|) \begin{bmatrix} 0.292 & 0.4114 & 0.6225 & 0.8097 \\ 0.2904 & 0.4081 & 0.6311 & 0.7977 \\ 0.2952 & 0.4149 & 0.6242 & 0.8045 \end{bmatrix} \tag{5}$$

We can see that at very small values of n, unbiasedness is a valuable property, but as soon as we get to reasonable data sizes, the maximum likelihood estimator outperforms the Method of Moments estimator by a large margin. So for data sets of reasonable size, I would always prefer the maximum likelihood estimator for Uniform estimation.

Unbiasedness is not always the most important property. Maximum likelihood estimators trade off bias for much lower variance. This trade-off can be examined by looking at Mean-Squared error, which is considers this trade-off.

3.2 (c)

Can you justify your preference on theoretical grounds?

From the Central Limit Theorem, we know that $\sqrt{n}(\tilde{\theta}_n - \theta) \sim \mathcal{N}$. So we know that $\tilde{\theta}_n$ is $\sqrt{n}$-consistent for θ. However, $\widehat{\theta}_n$ is n-consistent for θ. This indicates that $\widehat{\theta}_n$ is converging to θ more quickly than $\tilde{\theta}_n$ is. We are able to "blow up" $\widehat{\theta}_n$ by more than $\tilde{\theta}_n$ and still have the sequence be tight.

From part (d) of question 2, we showed that $\widehat{\theta}_n$ was n-consistent, and that it converged to an exponential distribution. This tells us that the order of convergence of the maximum likelihood estimator is $\mathcal{O}(\frac{1}{n})$ rather than the order of converge of $\mathcal{O}(\frac{1}{\sqrt{n}})$ of the Method of Moments estimator.

4 Question 4

4.1 (a)

$$\begin{aligned} \ell_n(\theta) &= \prod_{i=1}^{n} p_\theta(y|x) \\ &= \prod_{i=1}^{n} G(x'\theta)^y (1 - G(x'\theta))^{1-y} \\ L_n(\theta) &= \frac{1}{n} \sum_{i=1}^{n} y \log(G(x'\theta)) + (1-y) \log(1 - G(x'\theta)) \end{aligned}$$

4.2 (b)

Uniqueness of the Maximum likelihood estimator requires that

$$\Pr(p_\theta(y|x) \neq p_{\theta_0}(y|x)) > 0$$
$$\Pr(G(x'\theta) \neq G(x'\theta_0) > 0)$$

This occurs when G is a one-to-one function, and there is no perfect colinearity in X. A class of functions with this property that also ensures this is a valid probability measure is the class of strictly increasing functions with range $[0, 1]$.

4.3 (c)

No, there is perfect separation between the data, so the logit regression will not reach a maximum.

We would like to predict the y for any value of x below 1 as 0, and for any value of x above 0.8 as 1. We can always choose a steeper logit function to increase the probability that we are correct for this data. This implies that there is no maximum, as the likelihood function is unbounded above.

Assume that x takes the form of $(1, X_i)'$ and that $\theta = (a, b)'$ For any proposed maximum, $0 < G(x'\theta) < 1$. We also know that $-\frac{a}{b}$ must be between .8 and 1, as this is the point of inflection. By increasing b, and ensuring that $-\frac{a}{b}$ is constant between .8 and 1, the probability of the $y = 1$ data increases, and the probability of the $y = 0$ data increases as well. This increases the likelihood function, and contradicts the assumption that it was a maximum.

4.4 (d)

$$\begin{aligned}
\log p_\theta(y|x) &= y\log(G(x'\theta)) + (1-y)\log(1-G(x'\theta)) \\
\frac{\partial \log p_\theta(y|x)}{\partial \theta_j} &= \frac{yG'(x'\theta)x_j}{G(x'\theta)} - \frac{(1-y)G'(x'\theta)x_j}{G(x'\theta)} \\
\nabla \log p_\theta(y|x) &= \frac{yG'(x'\theta) - (1-y)G'(x'\theta)}{G(x'\theta)} X \\
\nabla \log p_\theta(y|x) \nabla \log p_\theta(y|x)' &= \left(\frac{yG'(x'\theta) - (1-y)G'(x'\theta)}{G(x'\theta)}\right)^2 XX'
\end{aligned}$$

Call this outer-product matrix A. Under the appropiate regularity conditions we know that $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, A^{-1})$.

5 Question 5

Let X_i be an i.i.d. sequence of random variables with pdf on $\mathbb{R}$ given by:

$$f_\theta(x) = \begin{cases} (1+\theta)x^\theta & \text{for } 0 < x < 1 \\ 0 & otherwise \end{cases}$$

for some $\theta > -1$.

5.1 (a)

$$\begin{aligned}
\mu = \mathbb{E}\left[X_i\right] &= \int_0^1 (1+\theta)x^{\theta+1} \\
&= \frac{1+\theta}{2+\theta} x^{\theta+2}|_0^1 \\
&= \frac{1+\theta}{2+\theta}
\end{aligned}$$

$$\begin{aligned}
\mu(2+\theta) &= 1+\theta \\
2\mu - 1 &= \theta(1-\mu) \\
\theta &= \frac{2\mu-1}{1-\mu} \quad \mu \in (0,1)
\end{aligned}$$

5.2 (b)

Write the log-likelihood function as a function of μ.

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log(1+\theta) + \theta\log(X_i)$$
$$L_n(\mu) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{\mu}{1-\mu}\right) + \frac{2\mu-1}{1-\mu}\log(X_i)$$
$$L_n(\mu) = \frac{1}{n}\sum_{i=1}^{n} \log\mu - \log(1-\mu) + \frac{2\mu-1}{1-\mu}\log(X_i)$$

5.3 (c)

We believe that the solution will be interior. This condition is equivalent to saying that $\mu \neq 0$ and $\mu \neq 1$.

$$0 = \frac{\partial L_n(\mu)}{\partial \mu} = \frac{1}{n}\sum_{i=1}^{n} \frac{\mu\log(X_i) - \mu + 1}{(1-\mu)^2\mu} = 0$$
$$0 = \frac{1}{n}\sum_{i=1}^{n} \mu\log(X_i) - \mu + 1$$
$$\widehat{\mu}_n = \frac{1}{1 - \frac{1}{n}\sum_{i=1}^{n}\log(X_i)}$$

As long as $\log(X_i) \neq 0$ and $\log(X_i) \neq 1$, which occurs with probability 0.

5.4 (d)

We may note that $\widehat{\mu}_n$ is a continuous function of $\frac{1}{n}\sum_{i=1}^{n}\log(X_i)$. Provided that $\mathbb{E}[\log(X_i)]$ exists, the Weak Law of Large Numbers tells us that $\frac{1}{n}\sum_{i=1}^{n}\log(X_i) \xrightarrow{p} \mathbb{E}[\log(X_i)]$.

$$\begin{aligned}\mathbb{E}[\log(X_i)] &= \int_0^1 (1+\theta)\log(x)x^{\theta} \\ &= \log(x)x^{\theta+1}|_0^1 - \int_0^1 \frac{x^{\theta+1}}{x}dx \\ &= -\frac{x^{\theta+1}}{\theta+1}|_0^1 \\ &= -\frac{1}{\theta+1}\end{aligned}$$

By the Continuous mapping theorem, we know that:

$$\begin{aligned}\widehat{\mu}_n &\xrightarrow{p} \frac{1}{1-\frac{-1}{\theta+1}} \\ &= \frac{\theta+1}{\theta+2} \\ &= \mu\end{aligned}$$

5.5 (e)

$$\begin{aligned}\mathbb{E}\left[\log^2(X_i)\right] &= \int_0^1 (1+\theta)\log^2(x)x^\theta dx \\ &= \log^2(x)x^{\theta+1}|_0^1 - \int_0^1 2\log(x)x^{\theta+1}dx \\ &= -2\frac{1}{\theta+2}\log(x)x^{\theta+2}|_0^1 + \frac{2}{\theta+2}\int_0^1 x^{\theta+1}dx \\ &= \frac{2}{(\theta+2)^2}\end{aligned}$$

We notice that

$$\mathbb{V}\left(\log(X_i)\right) = \mathbb{E}\left[\log(X_i)^2\right] - \mathbb{E}\left[\log(X_i)\right]^2 = \frac{1}{(1+\theta)^2}$$

Since we know that the variance of $\log(X)$ is finite, we may apply the central limit theorem.

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^n \log(X_i) - \frac{-1}{1+\theta}) \xrightarrow{d} \mathcal{N}(0, \frac{1}{(1+\theta)^2})$$

Applying the delta-method, with the function $f(x) = \frac{1}{1-x}$. Note that its derivative is: $g(x) = \frac{1}{(1-x)^2}$.

Thus: $g(\frac{-1}{1+\theta})^2 = \left(\frac{1}{(1-\frac{-1}{(1+\theta)})^2}\right)^2 = \left(\frac{\theta+1}{\theta+2}\right)^4$

$$\sqrt{n}(\widehat{\mu}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \frac{(\theta+1)^2}{(\theta+2)^4}) = \mathcal{N}(0, \mu^2(1-\mu)^2)$$

5.6 (f)

The log-likelihood function for μ is given by:

$$log(\mu) - \log(1-\mu) + \frac{2\mu-1}{1-\mu}\log(X_i)$$

$$g(\mu) = \frac{1}{\mu} + \frac{1}{1-\mu} + \frac{1}{(1-\mu)^2}\log(X_i)$$
$$h(\mu) = \frac{-1}{\mu^2} + \frac{1}{(1-\mu)^2} + \frac{2}{(1-\mu)^3}\log(X_i)$$

The Information matrix is given by the negative of the expected value of the second derivative.

$$\mathbb{E}\left[\log(X_i)\right] = \frac{-1}{1+\theta} = \frac{-(\theta+2)}{\mu} = \frac{-(1-\mu)}{\mu}$$

Combining:

$$\mathbb{E}\left[h(\mu)\right] = \frac{-1}{\mu^2} + \frac{1}{(1-\mu)^2} + \frac{2}{(1-\mu)^3}\frac{-(1-\mu)}{\mu}$$
$$= \frac{-1}{\mu^2(1-\mu)^2}$$

Fischer's information matrix is therefore:

$$-B = \frac{1}{\mu^2(1-\mu)^2}$$

This is the inverse of the variance computed via the delta-method.

5.7 (g)

Since we have a one-dimensional estimate, our constrained estimate of μ is $\tilde{\mu}_n = \frac{2}{3}$.

We know that ${H_n}^{-1} \xrightarrow{p} -B$ and $F_n \xrightarrow{p} 1$. So we estimate ${H_n}^{-1}$ with $\frac{1}{(\frac{2}{3})^2(\frac{1}{3})^2} = \frac{81}{4}$.

The log-likelihood is given by:

$$L_n(\mu) = \frac{1}{n}\sum_{i=1}^{n} \log\mu - \log(1-\mu) + \frac{2\mu-1}{1-\mu}\log(X_i)$$

Taking the derivative:

$$\frac{\partial L_n}{\partial \mu}(\tilde{\mu}_n) = \frac{3}{2} + 3 + 9\frac{1}{n}\sum_{i=1}^{n}\log X_i$$

The score test is based on the fact that

$$F_n {H_n}^{-1}\sqrt{n}\frac{\partial L_n}{\partial \mu}(\tilde{\mu}_n) \xrightarrow{d} \mathcal{N}(-B^{-1})$$

We then can construct a test based on the test statistic

$$T_n := \frac{9}{2}\sqrt{n}\left(\frac{9}{2} + 9\frac{1}{n}\sum_{i=1}^{n}\log X_i\right)$$

We know that under the null, this test statistic is distributed $\mathcal{N}(0,1)$ so we will use the test:

$$\mathbb{1}_{\{|T_n|>z_{1-\frac{\alpha}{2}}\}}$$

where $z_{1-\frac{\alpha}{2}}$ is a critical value from a normal distribution.