Stat 513
Emerson, Win 16
Midterm Examination
February 10, 2016

Instructions:

- **This exam is closed book, closed notes. No use of calculators, computers, or cell phones is permitted.**
- **Write answers to the following questions on separate sheets of paper, starting each problem at the top of a new page. Use only the front side of each page. Be sure to write your name on the top of each page.**
- **In order to receive full credit, you must make clear how you derived the answers to the problems.**
- **You are allowed 50 minutes for this exam. When time is called, you must put down your pencils**

1. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. normal random variables with $Y_i \sim \mathcal{N}(\theta, \theta^2)$ with $\theta \in (0, \infty)$.

   (a) What is the Cramér-Rao lower bound for the variance of an unbiased estimator of θ.
   <u>**Ans**</u>: We first note that this is a regular probability model for $\theta \in (0, \infty)$. *(There was a cut and paste error on the exam, where I gave that θ could be any real number. However, at $\theta = 0$, we have a degenerate distribution, and for a negative θ the density as written is negative. We could allow $\theta < 0$ if we put $|\theta|$ in the denominator of the density. But even then, we cannot have a regular problem when $\theta = 0$ is in the parameter space, because the log likelihood is not differentiable.)*
   We thus find the information for the i-th observation

$$L_i(\theta) = \frac{1}{\sqrt{2\pi}\theta} \exp\left\{-\frac{(Y_i - \theta)^2}{2\theta^2}\right\}$$

$$\mathcal{L}_i(\theta) = -\frac{\log(2\pi)}{2} - \log(\theta) - \frac{(Y_i - \theta)^2}{2\theta^2} = -\frac{\log(2\pi)}{2} - \log(\theta) - \frac{Y_i^2}{2\theta^2} + \frac{Y_i}{\theta} - \frac{1}{2}$$

$$\mathcal{U}_i(\theta) = -\frac{1}{\theta} + \frac{Y_i^2}{\theta^3} - \frac{Y_i}{\theta^2} = \frac{1}{\theta^3}\left(Y_i^2 - \theta Y_i - \theta^2\right)$$

$$J_i(\theta) = -E\left[\frac{1}{\theta^2} - \frac{3Y_i^2}{\theta^4} + \frac{2Y_i}{\theta^3}\right] = -\frac{1}{\theta^2} + \frac{3(\theta^2 + \theta^2)}{\theta^4} - \frac{2\theta}{\theta^3} = \frac{3}{\theta^2}$$

which yields the total score and total information as

$$\mathcal{U}(\theta) = \frac{1}{\theta^3}\left(\sum_{i=1}^{n} Y_i^2 - \theta\sum_{i=1}^{n} Y_i - n\theta^2\right) = \frac{n}{\theta^3}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \theta\overline{Y} - \theta^2\right)$$

$$J(\theta) = \frac{3n}{\theta^2}$$

Thus, an unbiased estimator $T(\vec{Y})$ of θ must have

$$Var(\left(T(\vec{Y})\right) \geq \frac{\theta^2}{3n}.$$

(b) Is there a function $g(\theta)$ for which you can derive a best regular unbiased estimator (BRUE)? Justify your answer.
<u>**Ans**</u>: From the total score function given in part (a), we see that

$$\mathcal{U}(\theta) \neq A(\theta)\left(T(\vec{Y}) - g(\theta)\right)$$

for any $T(\vec{Y})$ or any $g(\theta)$, so there is no BRUE for any function of θ.

(c) Find the maximum likelihood estimate $\hat{\theta}$ of θ.
<u>**Ans**</u>: By setting the total score function equal to 0 in order to find an MLE $\hat{\theta}$, we find

$$\mathcal{U}(\hat{\theta}) = \frac{n}{\hat{\theta}^3}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \hat{\theta}\overline{Y} - \hat{\theta}^2\right) = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{-\overline{Y} \pm \sqrt{(\overline{Y})^2 + \frac{4}{n}\sum_{i=1}^{n} Y_i^2}}{2}$$

and we take the positive square root of the discriminant function to obtain an MLE that is consistent for θ.

(d) Derive an asymptotic distribution for $\hat{\theta}$.
<u>**Ans**</u>: We can use the asymptotic results for MLE in regular problems with i.i.d. data to find

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \rightarrow_d \mathcal{N}\left(0,\, J_1^{-1}(\theta)\right) \quad \Rightarrow \quad \sqrt{n}\left(\hat{\theta} - \theta\right) \rightarrow_d \mathcal{N}\left(0,\, \frac{\theta^2}{3}\right)$$

(e) Find the asymptotic relative efficiency of the MLE $\hat{\theta}$ compared to the sample mean $\overline{Y}$.
<u>**Ans**</u>: Using the Levy CLT we find

$$\sqrt{n}\left(\overline{Y} - \theta\right) \rightarrow_d \mathcal{N}\left(0,\, \theta^2\right),$$

so by taking the ratio of the asymptotic variances we find that the sample mean has one-third $(\frac{1}{3})$ the asymptotic efficiency of the MLE.

(f) Find a MME estimator $\tilde{\theta}$ based on the sample variance, and derive its asymptotic distribution. What is the asymptotic relative efficiency of the MLE $\hat{\theta}$ compared to the MME $\tilde{\theta}$?

<u>**Ans**</u>: (Using properties of normally distributed data) We know that the sample variance s^2 computed from normally distributed data is related to a chi square distribution by

$$(n-1)\frac{s^2}{Var(Y)} = (n-1)\frac{s^2}{\theta^2} \sim \chi^2_{n-1},$$

and we also know that asymptotically, if $V_n \sim \chi^2_n$, then

$$\sqrt{n}\left(\frac{1}{n}V_n - 1\right) \to_d \mathcal{N}(0,2).$$

Hence

$$\sqrt{n-1}\left(\frac{1}{n-1}\frac{(n-1)s^2}{\theta^2} - 1\right) \to_d \mathcal{N}(0,2) \quad \Rightarrow \quad \sqrt{n}\left(s^2 - \theta^2\right) \to_d \mathcal{N}(0, 2\theta^4).$$

Then using the delta method with $g(x) = \sqrt(x)$ (so $g'(x) = 1/(2\sqrt{x})$)

$$\sqrt{n}\left(s - \theta\right) \to_d \mathcal{N}\left(0, \frac{\theta^2}{2}\right).$$

Taking the ratio of asymptotic variances, we find that the sample variance has two-thirds ($\frac{2}{3}$) the efficiency of the asymptotic efficiency of the MLE.

(Alternative approach using the asymptotic distribution for s^2 in a distribution-free problem) For i.i.d. $Y_i \sim (\mu, \sigma^2)$ with fourth central moment γ_4, we know

$$\sqrt{n}(s^2 - \sigma^2) \to_d \mathcal{N}\left(0,\ \gamma_4 - \sigma^4\right).$$

For a normally distributed random variable, we know $\gamma_4 = 3\sigma^4$, so for $\sigma^2 = \theta^2$ we have

$$\sqrt{n}(s^2 - \theta^2) \to_d \mathcal{N}\left(0,\ 2\theta^4\right).$$

Applying the delta method as before gives us our same result for the asymptotic distribution of s

$$\sqrt{n}\left(s - \theta\right) \to_d \mathcal{N}\left(0, \frac{\theta^2}{2}\right).$$

2. Consider a simple linear regression in which we presume that the mean of Y_i is related to known predictor X_i by

$$(Y_i \mid X_i = x_i) = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where we presume the ϵ_i's are independent with $\epsilon_i \sim (0, \sigma_i^2)$.

(a) Derive expressions for the ordinary least squares estimator $\hat{\vec{\beta}}_{OLS}$ of $\vec{\beta}$, and provide expressions for its mean and variance, making clear any special notation you use.

<u>**Ans**</u>: We define response vector $\vec{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ and n by 2 design matrix $\mathbf{X}$ with the first column a vector of 1's and the second column the vector of predictors (so $\mathbf{X}_{i1} = 1$ and $\mathbf{X}_{i2} = X_i$). We further define covariance matrix $\mathbf{V}$ as the diagonal matrix having $\mathbf{V}_{ii} = \sigma_i^2$. We thus have for $\vec{\beta} = (\beta_0, \beta_2)^T$

$$\vec{Y} \,|\, \mathbf{X} \;\sim\; \left(\mathbf{X}\vec{\beta},\; \mathbf{V}\right)$$

and providing that $\exists i, j$ such that $X_i \neq X_j$, the OLSE is given by

$$\hat{\vec{\beta}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y} = \begin{pmatrix} \overline{Y} - \hat{\beta}_1\overline{X} \\ \\ \frac{S_{XY}}{S_{XX}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \left\{ \left(\frac{1}{n} - \frac{(X_i - \overline{X})}{\sum_{j=1}^n (X_j - \overline{X})^2} \right) Y_i \right\} \\ \\ \frac{\sum_{i=1}^n (X_i - \overline{X})\, Y_i}{\sum_{i=1}^n (X_i - \overline{X})^2} \end{pmatrix},$$

where the expressions based on $\overline{X}$, $\overline{Y}$, S_{XX} and S_{XY} presume that the first column of $\mathbf{X}$ is the constant 1. Using properties of expectation, we thus have

$$E\left[\hat{\vec{\beta}} \,|\, \mathbf{X}\right] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\, E[\vec{Y} \,|\, \mathbf{X}\,] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\vec{\beta} = \vec{\beta}$$

$$Var\left(\hat{\vec{\beta}} \,|\, \mathbf{X}\right) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\, Var(\vec{Y} \,|\, \mathbf{X}\,)\, \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$$

$$= \begin{pmatrix} \sum_{i=1}^n \left\{ \left(\frac{1}{n} - \frac{(X_i - \overline{X})}{S_{XX}} \right)^2 \sigma_i^2 \right\} & \sum_{i=1}^n \left\{ \left(\frac{(X_i - \overline{X})}{nS_{XX}} - \frac{(X_i - \overline{X})^2}{S_{XX}^2} \right) \sigma_i^2 \right\} \\ \\ \sum_{i=1}^n \left\{ \left(\frac{(X_i - \overline{X})}{nS_{XX}} - \frac{(X_i - \overline{X})^2}{S_{XX}^2} \right) \sigma_i^2 \right\} & \frac{\sum_{i=1}^n (X_i - \overline{X})^2 \sigma_i^2}{S_{XX}^2} \end{pmatrix}$$

where the expressions in $Var(\hat{\vec{\beta}}|\mathbf{X})$ based on $\overline{X}$, $\overline{Y}$, and S_{XX} presume that the first column of $\mathbf{X}$ is the constant 1 and that $\mathbf{V}$ is diagonal, in which case the equations are derived easily by considering the corresponding equations for $\hat{\vec{\beta}}$.
It was sufficient just to provide estimates and variances in matrix notation.

(b) Derive expressions for the best linear unbiased estimator $\hat{\vec{\beta}}_B$ of $\vec{\beta}$, and provide expressions for its mean and variance, again making clear any special notation you use.
<u>**Ans**</u>: By the Gauss-Markov Theorem, the BLUE is the generalized least squares estimate with weights derived from the inverse of the variance-covariance matrix for $\vec{Y}$. We can work the problem in either the untransformed state, or, because the variance-covariance matrix for $\vec{Y}$ is diagonal, under a transformation

$$\left.\begin{aligned} W_i &= \tfrac{1}{\sigma_i} X_i \\ \\ Z_i &= \tfrac{1}{\sigma_i} Y_i \end{aligned}\right\} \quad \Rightarrow \quad \vec{Z} \;\sim\; \left(\mathbf{W}\vec{\beta},\; \mathbf{I}_n\right),$$

where $\mathbf{I}_n$ is the n-dimensional identity matrix and design matrix $\mathbf{W}$ satisfies $\mathbf{W}_{i1} = \frac{1}{\sigma_i}$ and $\mathbf{W}_{i2} = W_i$.
Using the same notation as in part a we thus have

$$\hat{\vec{\beta}}_B = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\vec{Y} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\vec{Z}.$$

(Note that with a non-constant first column of $\mathbf{W}$, there is not a particularly simple expression for $\hat{\vec{\beta}}_B$ in terms of univariate summary measures. Instead they would involve weighted means and sums.)
Using properties of expectation, we thus have

$$E\left[\hat{\vec{\beta}}\,|\,\mathbf{X}\right] = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\,E[\vec{Y}\,|\,\mathbf{X}\,] = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\vec{\beta} = \vec{\beta}$$

$$E\left[\hat{\vec{\beta}}\,|\,\mathbf{W}\right] = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\,E[\vec{Z}\,|\,\mathbf{W}\,] = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{W}\vec{\beta} = \vec{\beta}$$

$$Var\left(\hat{\vec{\beta}}\,|\,\mathbf{X}\right) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\,Var(\vec{Y}\,|\,\mathbf{X}\,)\,\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$
$$= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$
$$= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

$$Var\left(\hat{\vec{\beta}}\,|\,\mathbf{W}\right) = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\,Var(\,\vec{Z}\,|\,\mathbf{W}\,)\,\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}$$
$$= (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{I}_n\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1} = (\mathbf{W}^T\mathbf{W})^{-1}$$

It was sufficient just to provide estimates and variances in matrix notation in either the untransformed (preferable) or transformed model.

(c) Suppose that $X_i = x_i > 0$ for all i and $\sigma_i^2 = \theta^2 x_i$. Write down an expression for an unbiased estimator of θ^2.
Ans: *On the exam, there was a typo that suggested $X_i = x_1\ \forall i$. If X_i is constant, then this is a one sample problem, in which case the design matrix $\mathbf{X}$ is not full rank, and there is not a unique BLUE of $\vec{\beta}$, though there is still a BLUE of $E[Y_i]$. In that one sample problem, the sample variance $s^2 = \frac{1}{n-1}\sum(Y_i - \overline{Y})^2$ estimates $x_1\theta^2$, so s^2/x_1 would be an unbiased estimator of θ^2. I give below the answer to the problem as I wrote it above, which is the question I meant to ask.*

Under a slight modification of the notation in parts a and b, we define covariance matrix $\mathbf{V}$ as the diagonal matrix having $\mathbf{V}_{ii} = x_i$. We thus have for $\vec{\beta} = (\beta_0, \beta_2)^T$

$$\vec{Y}\,|\,\mathbf{X} \sim \left(\mathbf{X}\vec{\beta},\ \theta^2\mathbf{V}\right)$$

and for transformations

$$\left.\begin{array}{l} W_i = \frac{1}{\sqrt{x_i}}X_i \\ \\ Z_i = \frac{1}{\sqrt{x_i}}Y_i \end{array}\right\} \quad \Rightarrow \quad \vec{Z} \sim \left(\mathbf{W}\vec{\beta},\ \theta^2\mathbf{I}_n\right),$$

where $\mathbf{I}_n$ is the n-dimensional identity matrix and design matrix $\mathbf{W}$ satisfies $\mathbf{W}_{i1} = \frac{1}{\sqrt{x_i}}$ and $\mathbf{W}_{i2} = W_i$.
We then define

$$\hat{\vec{\beta}}_B = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\vec{Z}.$$

Then using a theorem given in class, for

$$\begin{aligned}
\hat{\theta}^2 &= \frac{1}{n-2}\left(\vec{Z} - \mathbf{W}\hat{\vec{\beta}}_B\right)^T\left(\vec{Z} - \mathbf{W}\hat{\vec{\beta}}_B\right) \\
&= \frac{1}{n-2}\left(\vec{Y} - \mathbf{X}\hat{\vec{\beta}}_B\right)^T\mathbf{V}^{-1}\left(\vec{Y} - \mathbf{X}\hat{\vec{\beta}}_B\right) \\
&= \frac{1}{n-2}\sum_{i=1}^{n}\frac{(Y_i - \hat{\beta}_{B0} - x_i\hat{\beta}_{B1})^2}{x_1}
\end{aligned}$$

we have $E\left[\hat{\theta}^2\,|\mathbf{X}\right] = \theta^2$.

3. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables with the double exponential distribution having density for $\mu \in (-\infty, \infty)$

$$f_Y(y\,|\,\mu) = \frac{1}{2}e^{-|y-\mu|}.$$

(a) Is this a regular probability model? Explain your answer.
Ans: No, the log likelihood is not differentiable at μ.

(b) Find a method of moments estimate $\tilde{\mu}$ for μ and derive its asymptotic distribution. (Hint: What is the distribution of a random variable $W_i = Y_i - \mu$.)
Ans: Using the hint, we first find the moments of $W_i = Y_i - \mu$ as

$$\begin{aligned}
E[W_i] &= \int_{-\infty}^{\infty} w\,f(w)\,dw = \int_{-\infty}^{\infty} w\,\frac{1}{2}e^{-|w|}\,dw = 0 && \text{(odd function)} \\
E[W_i^2] &= \int_{-\infty}^{\infty} w^2\,\frac{1}{2}e^{-|w|}\,dw = 2\int_{0}^{\infty} w^2\,\frac{1}{2}e^{-w}\,dw && \text{(even function)} \\
&= 2 && \text{(expectation of squared E(1) r.v.)} \\
Var(W_i) &= E[W_i^2] + E^2[W_i] = 2.
\end{aligned}$$

Thus we find

$$\begin{aligned}
E[Y_i] &= E[W_i + \mu] = \mu \\
Var(Y_i) &= Var(W_i + \mu) = Var(W_i) = 2
\end{aligned}$$

So a method of moments estimator is $\tilde{\mu} = \overline{Y}$, which by the Levy CLT has asymptotic distribution

$$\sqrt{n}\,(\tilde{\mu} - \mu) \;\to_d\; \mathcal{N}(0, 2).$$

(c) Find the asymptotic distribution of maximum likelihood estimate $\hat{\mu}$ (the sample median) for μ. (You do not need to prove the form of the MLE.)
Ans: Owing to the density f_Y being symmetric about μ, we know that μ is both the mean and the median of this continuous distribution. Letting $\hat{\mu}$ be the sample median, and noting that the density is positive at the median, we know the asymptotic distribution is

$$\sqrt{n}\,(\hat{\mu}-\mu) \ \to_d \ \mathcal{N}\left(0,\ \frac{1}{4f_Y^2(\mu)}=1\right).$$

(d) What is the asymptotic relative efficiency of the MME $\tilde{\mu}$ compared to MLE $\hat{\mu}$.
Ans: Taking the ratio of the asymptotic variances of the two $\sqrt{n}$-consistent normally distributed estimators, we find that the MME $\tilde{\mu}$ has $\frac{1}{2}=50\%$ the efficiency of MLE $\hat{\mu}$.

4. Suppose $W_1, W_2, \ldots, W_n$ are i.i.d. multinomial random variables with $W_i \sim Mult(1,\vec{\theta})$, such that $W_i \in \{1,2,3\}$ with $\vec{\theta}=(\theta_1,\theta_2,\theta_3)$ satisfying $\theta_k = Pr(W_i=k)$, $\theta_k \in (0,1)$, and $\sum_{k=1}^{3}\theta_k = 1$. For notational convenience, we can define random vector $\vec{Y}_i$ as

$$\vec{Y}_i = \begin{pmatrix} Y_{i1} = \mathbf{1}_{[W_i=1]} \\ Y_{i2} = \mathbf{1}_{[W_i=2]} \\ Y_{i3} = \mathbf{1}_{[W_i=3]} \end{pmatrix}$$

We also note that owing to the constraints on the multinomial random variables, we will find it easiest to reparameterize our problem such that $\theta_3 = 1-\theta_1-\theta_2$ and $Y_{i3} = 1 - Y_{i1} - Y_{i2}$. Hence, we have density for $\vec{y}$ having exactly one nonzero element equal to 1

$$f_{\vec{Y}_i}(\vec{y}\,|\,\vec{\theta}) = \theta_1^{y_1}\theta_2^{y_2}\theta_3^{y_3} = \theta_1^{y_1}\theta_2^{y_2}(1-\theta_1-\theta_2)^{1-y_1-y_2}.$$

(a) Write down the likelihood and log likelihood of $\vec{\theta}$ based on observations $\vec{Y}_1, \vec{Y}_2, \ldots$.
Ans: Using the reparameterization for $\vec{Y}$ such that $Y_{ij} \in (0,1)$ for $j=1,2,3$ and $Y_{i1}+Y_{i2}+Y_{i3}=1$

$$\begin{aligned} L_i(\vec{\theta}\,|\,\vec{Y}_i=\vec{y}_i) &= \theta_1^{y_{i1}}\theta_2^{y_{i2}}(1-\theta_1-\theta_2)^{1-y_{i1}-y_{i2}}\mathbf{1}_{[\theta_1\in(0,1)]}\mathbf{1}_{[\theta_2\in(0,1)]}\mathbf{1}_{[\theta_1+\theta_2<1]} \\ \mathcal{L}_i(\vec{\theta}\,|\,\vec{Y}_i=\vec{y}_i) &= y_{i1}\log(\theta_1) + y_{i2}\log(\theta_2) + (1-y_{i1}-y_{i2})\log(1-\theta_1-\theta_2) \\ &\qquad + \log(\mathbf{1}_{[\theta_1\in(0,1)]}\mathbf{1}_{[\theta_2\in(0,1)]}\mathbf{1}_{[\theta_1+\theta_2<1]}) \\ &= y_{i1}\log\left(\frac{\theta_1}{1-\theta_1-\theta_2}\right) + y_{i2}\log\left(\frac{\theta_2}{1-\theta_1-\theta_2}\right) + \log(1-\theta_1-\theta_2) \\ &\qquad + \log(\mathbf{1}_{[\theta_1\in(0,1)]}\mathbf{1}_{[\theta_2\in(0,1)]}\mathbf{1}_{[\theta_1+\theta_2<1]}) \end{aligned}$$

where the second form of the log likelihood is included only to demonstrate the canonical parameters.

The full likelihood and log likelihood for these i.i.d. data are thus

$$
\begin{aligned}
L(\vec{\theta}\,|\,\vec{Y}_i = \vec{y}_i) &= \theta_1^{\sum y_{i1}}\theta_2^{\sum y_{i2}}(1-\theta_1-\theta_2)^{n-\sum y_{i1}-\sum y_{i2}}\mathbf{1}_{[\theta_1\in(0,1)]}\mathbf{1}_{[\theta_2\in(0,1)]}\mathbf{1}_{[\theta_1+\theta_2<1]} \\
\mathcal{L}(\vec{\theta}\,|\,\vec{Y}_i = \vec{y}_i) &= \sum y_{i1}\log(\theta_1) + \sum y_{i2}\log(\theta_2) + (n - \sum y_{i1} - \sum y_{i2})\log(1-\theta_1-\theta_2) \\
&\qquad + \log(\mathbf{1}_{[\theta_1\in(0,1)]}\mathbf{1}_{[\theta_2\in(0,1)]}\mathbf{1}_{[\theta_1+\theta_2<1]}) \\
&= \sum y_{i1}\log\left(\frac{\theta_1}{1-\theta_1-\theta_2}\right) + \sum y_{i2}\log\left(\frac{\theta_2}{1-\theta_1-\theta_2}\right) + n\log(1-\theta_1-\theta_2) \\
&\qquad + \log(\mathbf{1}_{[\theta_1\in(0,1)]}\mathbf{1}_{[\theta_2\in(0,1)]}\mathbf{1}_{[\theta_1+\theta_2<1]})
\end{aligned}
$$

We could write the likelihood and log likelihood in the original parameterization, though it would be important to include the constraint on the parameters.

(b) Find the maximum likelihood estimate $\hat{\vec{\theta}}$ of $\vec{\theta}$, and derive its mean and variance. Is it unbiased?
Ans: We first note that in the over parameterized form, the moments of $\vec{Y}_i$ are found from

$$
\begin{aligned}
E[Y_{ik}] &= Pr[W_i = k] = \theta_k \\
E[Y_{ik}^2] &= E[Y_{ik}] = \theta_k \qquad\qquad (Y_{ik} \text{ is a 0-1 variable}) \\
E[Y_{ik}Y_{ik'}] &= 0 \quad \text{for } k \neq k' \\
Var(Y_{ik}) &= E[Y_{ik}^2] - E^2[Y_{ik}] = \theta_k(1-\theta_k) \\
Cov(Y_{ik}Y_{ik'}) &= E[Y_{ik}Y_{ik'}] - E[Y_{ik}]E[Y_{ik'}] = -\theta_k\theta_{k'} \quad \text{for } k \neq k'
\end{aligned}
$$

obtaining

$$
\vec{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} \sim \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \quad \begin{pmatrix} \theta_1(1-\theta_1) & -\theta_1\theta_2 & -\theta_1\theta_3 \\ -\theta_1\theta_2 & \theta_2(1-\theta_2) & -\theta_2\theta_3 \\ -\theta_1\theta_3 & -\theta_2\theta_3 & \theta_3(1-\theta_3) \end{pmatrix} \right)
$$

In the reparameterized form, we merely consider the first two elements of $\vec{Y}$ and $\vec{\theta}$ and the corresponding partition of the variance-covariance matrix.

$$
\vec{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim \left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad \begin{pmatrix} \theta_1(1-\theta_1) & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_2(1-\theta_2) \end{pmatrix} \right)
$$

Using the fact that this is a regular probability model, we then find the contributions to the efficient score function and information from the i-th observation as

$$
\begin{aligned}
\mathcal{U}_{i1}(\vec{\theta}) &= \frac{\partial}{\partial\theta_1}\mathcal{L}(\vec{\theta}) = \frac{Y_{i1}}{\theta_1} - \frac{(1 - Y_{i1} - Y_{i2})}{(1 - \theta_1 - \theta_2)} \\
\mathcal{U}_{i2}(\vec{\theta}) &= \frac{\partial}{\partial\theta_2}\mathcal{L}(\vec{\theta}) = \frac{Y_{i2}}{\theta_2} - \frac{(1 - Y_{i1} - Y_{i2})}{(1 - \theta_1 - \theta_2)} \\
\mathbf{J}_{i11}(\vec{\theta}) &= -E\left[\frac{\partial}{\partial\theta_1}\mathcal{U}_{i1}(\vec{\theta})\right] = -E\left[-\frac{Y_{i1}}{\theta_1^2} - \frac{(1 - Y_{i1} - Y_{i2})}{(1 - \theta_1 - \theta_2)^2}\right] \\
&= \frac{1}{\theta_1} + \frac{1}{(1 - \theta_1 - \theta_2)} = \frac{(1 - \theta_2)}{\theta_1(1 - \theta_1 - \theta_2)} \\
\mathbf{J}_{i12}(\vec{\theta}) &= -E\left[\frac{\partial}{\partial\theta_1}\mathcal{U}_{i2}(\vec{\theta})\right] = -E\left[-\frac{(1 - Y_{i1} - Y_{i2})}{(1 - \theta_1 - \theta_2)^2}\right] = \frac{1}{(1 - \theta_1 - \theta_2)} \\
\mathbf{J}_{i21}(\vec{\theta}) &= \mathbf{J}_{12}(\vec{\theta}) = \frac{1}{(1 - \theta_1 - \theta_2)} \\
\mathbf{J}_{i22}(\vec{\theta}) &= \frac{(1 - \theta_1)}{\theta_2(1 - \theta_1 - \theta_2)}
\end{aligned}
$$

Hence the score vector and information matrix are given by

$$
\vec{\mathcal{U}}(\vec{\theta}) = \begin{pmatrix} \frac{\sum_{i=1}^n Y_{i1}}{\theta_1} - \frac{\sum_{i=1}^n Y_{i3}}{(1-\theta_1-\theta_2)} \\ \\ \frac{\sum_{i=1}^n Y_{i2}}{\theta_2} - \frac{\sum_{i=1}^n Y_{i3}}{(1-\theta_1-\theta_2)} \end{pmatrix} \qquad \mathbf{J}(\vec{\theta}) = \begin{pmatrix} \frac{n(1-\theta_2)}{\theta_1(1-\theta_1-\theta_2)} & \frac{n}{(1-\theta_1-\theta_2)} \\ \\ \frac{n}{(1-\theta_1-\theta_2)} & \frac{n(1-\theta_1)}{\theta_2(1-\theta_1-\theta_2)} \end{pmatrix}
$$

We then find the MLE $\hat{\vec{\theta}}$ by solving $\vec{\mathcal{U}}(\hat{\vec{\theta}}) = \vec{0}$. One approach would be to make the reasonable guess based on the results for a Bernoulli random variable. So we guess that

$$
\hat{\theta}_k = \frac{1}{n}\sum_{i=1}^n Y_{ik},
$$

and then verify that such a choice does indeed solve the score equations. Alternatively, you explicitly solve the score equations in their reparameterized form (letting $\overline{Y}_{\cdot k} = \frac{1}{n}\sum_{i=1}^{n} Y_{ik}$)

$$\begin{aligned}
\mathcal{U}_1(\hat{\vec{\theta}}) &= \frac{n\overline{Y}_{\cdot 1}}{\hat{\theta}_1} - \frac{n(1-\overline{Y}_{\cdot 1}-\overline{Y}_{\cdot 2})}{(1-\hat{\theta}_1-\hat{\theta}_2)} = 0 \quad \Rightarrow \quad \overline{Y}_{\cdot 1} - \hat{\theta}_2\overline{Y}_{\cdot 1} - \hat{\theta}_1 + \hat{\theta}_1\overline{Y}_{\cdot 2} = 0 \\
&\Rightarrow \quad \hat{\theta}_1 = \left(\frac{1-\hat{\theta}_2}{1-\overline{Y}_{\cdot 2}}\right)\overline{Y}_{\cdot 1} \\
\mathcal{U}_2(\hat{\vec{\theta}}) &= \frac{n\overline{Y}_{\cdot 2}}{\hat{\theta}_2} - \frac{n(1-\overline{Y}_{\cdot 1}-\overline{Y}_{\cdot 2})}{(1-\hat{\theta}_1-\hat{\theta}_2)} = 0 \quad \Rightarrow \quad \overline{Y}_{\cdot 2} - \hat{\theta}_1\overline{Y}_{\cdot 2} - \hat{\theta}_2 + \hat{\theta}_2\overline{Y}_{\cdot 1} = 0 \\
&\Rightarrow \quad \overline{Y}_{\cdot 2} - \left(\frac{1-\hat{\theta}_2}{1-\overline{Y}_{\cdot 2}}\right)\overline{Y}_{\cdot 1}\overline{Y}_{\cdot 2} - \hat{\theta}_2 + \hat{\theta}_2\overline{Y}_{\cdot 1} = 0 \\
&\Rightarrow \quad \hat{\theta}_2 = \overline{Y}_{\cdot 2} \\
&\Rightarrow \quad \hat{\theta}_1 = \overline{Y}_{\cdot 1}
\end{aligned}$$

We could have also worked in the over-parameterized setting using Lagrange multipliers to model the constraint that $\sum_{k=1}^{3}\theta_k = 1$. For $\lambda > 0$, we maximize $h(\vec{\theta}, \lambda)$

$$h(\vec{\theta},\lambda) = \mathcal{L}(\vec{\theta}) + \lambda(1-\sum_{k=1}^{3}\theta_k) = n\overline{Y}_{\cdot 1}\log(\theta_1) + n\overline{Y}_{\cdot 2}\log(\theta_2) + n\overline{Y}_{\cdot 3}\log(\theta_3) + \lambda(1-\sum_{k=1}^{3}\theta_k).$$

We find the partial derivatives

$$\left.\begin{aligned}
\tfrac{\partial}{\partial\theta_1}h(\vec{\theta},\lambda) &= \tfrac{n\overline{Y}_{\cdot 1}}{\theta_1} - \lambda \\
\tfrac{\partial}{\partial\theta_2}h(\vec{\theta},\lambda) &= \tfrac{n\overline{Y}_{\cdot 2}}{\theta_2} - \lambda \\
\tfrac{\partial}{\partial\theta_3}h(\vec{\theta},\lambda) &= \tfrac{n\overline{Y}_{\cdot 3}}{\theta_3} - \lambda \\
\tfrac{\partial}{\partial\lambda}h(\vec{\theta},\lambda) &= 1 - \textstyle\sum_{k=1}^{3}\theta_k
\end{aligned}\right\} \quad \Rightarrow \quad \begin{cases}
\frac{n\overline{Y}_{\cdot 1}}{\hat{\theta}_1} - \hat{\lambda} = 0 & \Rightarrow \quad \hat{\theta}_1 = \frac{n\overline{Y}_{\cdot 1}}{\hat{\lambda}} \\
\frac{n\overline{Y}_{\cdot 2}}{\hat{\theta}_2} - \hat{\lambda} = 0 & \Rightarrow \quad \hat{\theta}_2 = \frac{n\overline{Y}_{\cdot 2}}{\hat{\lambda}} \\
\frac{n\overline{Y}_{\cdot 3}}{\hat{\theta}_3} - \hat{\lambda} = 0 & \Rightarrow \quad \hat{\theta}_3 = \frac{n\overline{Y}_{\cdot 3}}{\hat{\lambda}} \\
1 - \sum_{k=1}^{3}\hat{\theta}_k = 0 &
\end{cases}$$

Combining the above, we find

$$\sum_{k=1}^{3}\hat{\theta}_k = 1 \quad \Rightarrow \quad \frac{\sum_{k=1}^{n} n\overline{Y}_{\cdot k}}{\hat{\lambda}} = 1 \quad \Rightarrow \quad \hat{\lambda} = n.$$

So by any of these methods, $\hat{\vec{\theta}}$ is the mean $\frac{1}{n}\sum_{i=1}^{n}\vec{Y}_i$, and thus by properties of expectation,

$$\begin{aligned}
E\left[\hat{\vec{\theta}}\right] &= \vec{\theta} \\
Var\left(\hat{\vec{\theta}}\right) &= \frac{1}{n}Var(\vec{Y}_1) = \frac{1}{n}\begin{pmatrix}\theta_1(1-\theta_1) & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_2(1-\theta_2)\end{pmatrix}
\end{aligned}$$

(c) What is the Cramér-Rao lower bound for variance-covariance matrix of an unbiased estimator of $\vec{\theta}$?

Ans: For an unbiased estimator $\tilde{\vec{\theta}}$ of $g(\vec{\theta}) = \vec{\theta}$, we have gradient $\nabla(g(\vec{\theta})+b(\vec{\theta})) = \vec{1}$. This is a regular probability model, so by the multidimensional Cramér-Rao theorem, the optimal variance-covariance matrix for $\tilde{\vec{\theta}}$ would be

$$\nabla^T(g(\vec{\theta}) + b(\vec{\theta}))\,\mathbf{J}^{-1}(\vec{\theta})\,\nabla(g(\vec{\theta}) + b(\vec{\theta})) = \mathbf{J}^{-1}(\vec{\theta}),$$

which from the above form of $\mathbf{J}(\vec{\theta})$ can be shown to be

$$\mathbf{J}^{-1}(\vec{\theta}) = \begin{pmatrix} \frac{\theta_1(1-\theta_1)}{n} & -\frac{\theta_1\theta_2}{n} \\ -\frac{\theta_1\theta_2}{n} & \frac{\theta_2(1-\theta_2)}{n} \end{pmatrix}$$

Note that we could have also invoked that the asymptotic variance of the MLE in this regular probability model must attain the C-R bound, hence owing to the multivariate CLT (see below), the variance of our MLE as derived above is the C-R bound.

(d) Derive an asymptotic distribution for $\hat{\vec{\theta}}$.

Ans: Because $\hat{\vec{\theta}} = \overline{\vec{Y}}$, the sample mean, using the multivariate CLT in the reparameterized model, we know

$$\sqrt{n}\left(\overline{\vec{Y}} - \vec{\theta}\right) \to_d \mathcal{N}\left(\vec{0},\, Var(\vec{Y}_1) = \begin{pmatrix} \theta_1(1-\theta_1) & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_2(1-\theta_2) \end{pmatrix}\right).$$

(We could also have invoked the results we have for asymptotic normality of MLE's in regular probability models. Note that $Var(\vec{Y}_1) = \mathbf{J}^{-1}(\vec{\theta})$.)

(e) Suppose we are interested in inference about $g(\vec{\theta}) = \theta_1 - (\theta_2 + \theta_3)$. Find the asymptotic distribution for the MLE of $g(\vec{\theta})$.

Ans: In the reparameterized model, $g(\vec{\theta}) = \theta_1 - (\theta_2 + 1 - \theta_1 - \theta_2) = 2\theta_1 - 1$, a linear transformation of θ_1. By the invariance of MLEs, we have that $g(\hat{\vec{\theta}}) = 2\hat{\theta}_1 - 1$. By the multivariate delta method, with $\nabla g(\vec{\theta}) = (2\ 0)^T$, we thus find

$$\begin{aligned} \sqrt{n}\left(g(\hat{\vec{\theta}}) - g(\vec{\theta})\right) &\to_d \mathcal{N}\left(\vec{0},\, (2\quad 0)\begin{pmatrix} \theta_1(1-\theta_1) & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_2(1-\theta_2) \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right) \\ &= \mathcal{N}\left(0,\, 4\theta_1(1-\theta_1)\right). \end{aligned}$$