Statistics 582, Final Exam Solutions

Wellner; 3/27/98

1. (36 points) (More inequalities involving Hellinger distance and total variation distance.) Suppose that P and Q are two probability measures with densities p and q with respect to some dominating measure μ. So that the Hellinger distance $d_H(P,Q)$ is given by

$$d_H^2(P,Q) = (1/2)\int |\sqrt{p}-\sqrt{q}|^2 d\mu = 1-\rho(P,Q)$$

where $\rho \equiv \rho(P,Q) = \int \sqrt{pq}\,d\mu$, and the total variation distance $d_{TV}(P,Q)$ is given by

$$d_{TV}(P,Q) = (1/2)\int |p-q| d\mu = 1-\pi(P,Q)$$

where $\pi \equiv \pi(P,Q) = \int p\wedge q\, d\mu$.

A. Prove the following inequalities:

$$\pi^2 \le \rho^2 \le 1-(1-\pi)^2 = \pi(2-\pi) \le 2\pi\,.$$

Thus we have

$$(1-d_H^2(P,Q))^2 \le 1-d_{TV}(P,Q)^2 = 1-\left(1-\int p\wedge q\,d\mu\right)^2 \le 2\int p\wedge q d\mu\,.$$

[Hint: The first inequality is easy. To prove the second inequality, use Cauchy-Schwarz to show that $\rho(P,Q)^2 + d_{TV}(P,Q)^2 \le 1$.]

B. Show that $K(P,Q) \ge d_H^2(P,Q)$. [Hint: Use Jensen's inequality and $\log(1-x) \le -x$.]

Solution: A. The first inequality is equivalent to

$$\pi(P,Q) = \int p\wedge q d\mu \le \int \sqrt{pq} d\mu = \rho(P,Q)$$

and this follows easily since

$$\begin{aligned} p\wedge q &= p1_{[p\le q]} + q1_{[p>q]} \\ &= \sqrt{p}\sqrt{p}1_{[p\le q]} + \sqrt{q}\sqrt{q}1_{[p>q]} \\ &\le \sqrt{p}\sqrt{q}1_{[p\le q]} + \sqrt{p}\sqrt{q}1_{[p>q]} \\ &= \sqrt{p}\sqrt{q}\,. \end{aligned}$$

The second inequality follows via Cauchy-Schwarz from

$$\begin{aligned}
\rho^2(P,Q) + d_{TV}(P,Q)^2 &= \rho^2 + \left(\frac{1}{2}\int |p-q|\, d\mu\right)^2 \\
&= \rho^2 + \left(\frac{1}{2}\int |\sqrt{p}-\sqrt{q}||\sqrt{p}+\sqrt{q}|\, d\mu\right)^2 \\
&\le \rho^2 + \frac{1}{4}\int |\sqrt{p}-\sqrt{q}|^2\, d\mu \int |\sqrt{p}+\sqrt{q}|^2\, d\mu \\
&= \rho^2 + \frac{1}{4}(2-2\rho)(2+2\rho) \\
&= \rho^2 + (1-\rho^2) = 1\,.
\end{aligned}$$

Thus we have

$$\rho^2(P,Q) \le 1 - d_{TV}^2(P,Q) = 1 - (1-\pi(P,Q))^2 = \pi(2-\pi) \le 2\pi\,.$$

Another, perhaps more straightforward, way of organizing this is as follows:

$$\begin{aligned}
(1-\pi)^2 &= d_{TV}^2(P,Q) = \left(\frac{1}{2}\int |p-q| d\mu\right)^2 \\
&= \frac{1}{4}\left(\int |\sqrt{p}-\sqrt{q}||\sqrt{p}+\sqrt{q}| d\mu\right)^2 \\
&\le \frac{1}{4}\int |\sqrt{p}-\sqrt{q}|^2 d\mu \int |\sqrt{p}+\sqrt{q}|^2 d\mu \\
&= \frac{1}{4}(2-2\rho)(2+2\rho) = 1 - \rho^2(P,Q)\,.
\end{aligned}$$

This yields

$$\rho^2(P,Q) \le 1 - (1-\pi(P,Q))^2\,.$$

B. Note that by convexity of the function $g(x) = -\log x$ and Jensen's inequality we have

$$\begin{aligned}
K(P,Q) &= E_P \log \frac{p}{q}(X) = -\int p \log \frac{q}{p} d\mu \\
&= -2\int \log(\sqrt{q/p}) d\mu \\
&\ge -2\log \int p\sqrt{q/p} d\mu = -2\log \int \sqrt{pq} d\mu \\
&= -2\log \rho(P,Q) = -2\log(1 - d_H^2(P,Q)) \\
&\ge 2d_H^2(P,Q)
\end{aligned}$$

where the last inequality follows since $-\log(1-x) \ge x$. (Thus my statement of the inequality was crude – exactly by a factor of 2!).

2. (56 points) Suppose that $X \sim$ Hypergeometric(n, θ, M); that is, X is the number of red balls drawn in n draws without replacement from an urn containing θ red balls and $M - \theta$ black balls, and it has probability mass function

$$P_\theta(X = x) = \frac{\binom{\theta}{x}\binom{M-\theta}{n-x}}{\binom{M}{n}}, \qquad \text{for } \ x = 0 \vee (\theta + n - M), \cdots, \theta \wedge n\,.$$

Suppose that θ has a Beta-Binomial prior distribution $BetaBin(\alpha, \beta, M)$ with mass function

$$\lambda(\theta) = \binom{M}{\theta} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(M+\alpha+\beta)} \qquad \text{for} \qquad \theta = 0, 1, \ldots, M\,.$$

Here M is a positive integer and $\alpha > 0$, $\beta > 0$. Suppose that the loss function is squared error: $L(\theta, a) = (\theta - a)^2$.

A. What is the maximum likelihood estimator of θ? (Note that differentiation won't work here since θ is integer-valued.) What, if any, of our theory for MLE's applies in this case?

B. Show that the unconditional distribution of X is $BetaBin(\alpha, \beta, n)$.

C. Show that the conditional distribution of $\theta - x$ given $X = x$ is $BetaBin(x + \alpha, n - x + \beta, M - n)$.

D. Show that the Bayes estimate of θ is

$$d(X) = \frac{(M + \alpha + \beta)X + \alpha(M - n)}{n + \alpha + \beta}\,.$$

E. Find the risk function of the estimates $d_{a,b}(X) = aX + b$.

F. Show that the estimate $d_0(X) = a_0 X + b_0$ has a constant risk equal to b_0^2 when

$$a_0 = \frac{M}{n(1+\delta)}, \qquad b_0 = (M - a_0 n)/2$$

where $\delta = \sqrt{(M-n)/(n(M-1))}$.

G. Show that the estimate d_0 is minimax (Bayes wrt $BetaBin(\alpha, \beta, n)$ when $\alpha = \beta = b_0/(a_0 - 1)$). Illustrate with the special case $N = 10$ and $n = 4$.

[Hint: See Ferguson page 100 for properties of the Beta-Binomial distributions.]

H. Compare and contrast this problem with the corresponding sampling with replacement situation in which $X \sim$ Binomial(n, θ) and $\theta \sim$ Beta(α, β). To begin this comparison, what is the natural analogue, in the "without replacement" (Hypergeometric) problem, of $\theta \in (0, 1)$ in the "with replacement" problem?

Solution: A. Consider the ratio $P_\theta(X = x)/P_{\theta-1}(X = x)$; we compute

$$\frac{P_\theta(X = x)}{P_{\theta-1}(X = x)} = \frac{\binom{\theta}{x}\binom{M-\theta}{n-x}/\binom{M}{n}}{\binom{\theta-1}{x}\binom{M-(\theta-1)}{n-x}/\binom{M}{n}}$$

$$
\begin{aligned}
&= \frac{\frac{\theta!}{x!(\theta-x)!}\cdot\frac{(M-\theta)!}{(n-x)!(M-\theta-n+x)!}}{\frac{(\theta-1)!}{x!(\theta-1-x)!}\cdot\frac{(M-\theta+1)!}{(n-x)!(M-\theta+1-n+x)!}}\\
&= \frac{\theta}{\theta-x}\cdot\frac{M-\theta+1-n+x}{M-\theta+1}\\
&\geq 1
\end{aligned}
$$

if and only if

$$\theta(M-\theta+1-n+x) \geq (\theta-x)(M-\theta+1)\,,$$

or, equivalently, if

$$n\theta \leq x(M+1)\,.$$

Thus the likelihood $L(\theta) \equiv P_\theta(X=x)|_X$ increases as a function of θ as long as θ is smaller than $(M+1)X/n$, and then it decreases. Hence the MLE $\widehat{\theta}_n$ of θ is given by

$$\widehat{\theta}_n = \lfloor (M+1)X/n \rfloor$$

where $\lfloor\cdot\rfloor$ is the greatest integer function. Note that this is quite sensible since

$$E_\theta(X) = n\frac{\theta}{M}\,;$$

see e.g. Ferguson MS page 100 (with $m=\theta$). Because the "without replacement sampling" scheme involves dependence between the successive draws, and because the parameter θ is integer valued, unfortunately very little of our theory from Chapter 4 applies. [Nevertheless, "with replacement sampling" yields i.i.d. data which provides a valuable comparison model in which the theory of Chapter 4 does apply!]

B. If we draw x red balls in n draws without replacement from the urn containing θ red balls and $M-\theta$ black balls, then we have $\theta \geq x$ and $n-x \leq M-\theta$ or $\theta \leq M-n+x$. Thus the marginal distribution of X with the BetaBinomial prior given by $\lambda(\theta)$ is

$$
\begin{aligned}
p(x) &= \sum_{\theta=x}^{M-n+x} P_\theta(X=x)\lambda(\theta)\\
&= \sum_{\theta=x}^{M-n+x} \frac{\binom{\theta}{x}\binom{M-\theta}{n-x}}{\binom{M}{n}}\binom{M}{\theta}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(M+\alpha+\beta)}\\
&= \sum_{\theta=x}^{M-n+x} \frac{\theta!}{x!(\theta-x)!}\cdot\frac{(M-\theta)!}{(n-x)!(M-\theta-n+x)!}\cdot\frac{n!(M-n)!}{M!}\\
&\quad \cdot\frac{M!}{\theta!(M-\theta)!}\cdot\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(M+\alpha+\beta)}
\end{aligned}
$$

$$
\begin{aligned}
&= \binom{n}{x}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)} \\
&\quad \cdot \sum_{\theta=x}^{M-n+x}\binom{M-n}{\theta-x}\frac{\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)\Gamma(M+\alpha+\beta)} \\
&= \binom{n}{x}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)} \\
&\quad \cdot \sum_{\theta'=}^{M'}\binom{M'}{\theta'}\frac{\Gamma(\alpha'+\beta')\Gamma(M'+\beta'-\theta')\Gamma(\alpha'+\theta')}{\Gamma(\alpha')\Gamma(\beta')\Gamma(M'+\alpha'+\beta')} \\
&\quad \text{by setting} \qquad \theta'=\theta-x,\ M'=M-n,\ \alpha'=\alpha+x,\ \beta'=n+\beta-x \\
&= \binom{n}{x}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}
\end{aligned}
$$

since the sum is over the BetaBinomial(α', β', M') probability mass function and hence equals 1. Thus $X \sim \text{BetaBinomial}(\alpha, \beta, n)$ marginally.

C. It follows from B that the posterior distribution of θ is given by

$$
\begin{aligned}
\lambda(\theta|x) &= \frac{P_\theta(X=x)\lambda(\theta)}{p(x)} \\
&= \frac{\frac{\binom{\theta}{x}\binom{M-\theta}{n-x}}{\binom{M}{n}}\binom{M}{\theta}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(M+\alpha+\beta)}}{\binom{n}{x}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}} \\
&= \binom{M-n}{\theta-x}\frac{\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)\Gamma(M+\alpha+\beta)}\,.
\end{aligned}
$$

Thus $\theta' \equiv \theta - x$ has the probability mass function

$$
\lambda(\theta'|x) = \binom{M'}{\theta'}\frac{\Gamma(\alpha'+\beta')\Gamma(\alpha'+\theta')\Gamma(M'+\beta'-\theta')}{\Gamma(\alpha')\Gamma(\beta')\Gamma(M'+\alpha'+\beta')}\,.
$$

Therefore $(\theta - x|X = x) \sim \text{BetaBinomial}(\alpha', \beta', M') = \text{BetaBinomial}(x+\alpha, n-x+\beta, M-n)$.

D. For squared-error loss, the Bayes estimate of θ is the posterior mean of θ, $E(\theta|X = x)$. Here we can calculate this using $\theta' = \theta - x$ as follows:

$$
\begin{aligned}
E(\theta|X=x) &= E(X+\theta-X|X=x) = x + E(\theta-X|X=x) \\
&= x + \frac{M'\alpha'}{\alpha'+\beta'} = x + \frac{(M-n)(x+\alpha)}{n+\alpha+\beta} \\
&= \frac{(M+\alpha+\beta)x + \alpha(M-n)}{n+\alpha+\beta}\,.
\end{aligned}
$$

Hence the Bayes estimator of θ for this prior is

$$d_\Lambda(X) = \frac{(M+\alpha+\beta)X + \alpha(M-n)}{n+\alpha+\beta}\,.$$

E. The linear estimators $d_{a,b}(X) = aX + b$ have risk functions $R(\theta, d_{a,b})$ given by

$$\begin{aligned}
R(\theta, d_{a,b}) &= E_\theta(\theta - d_{a,b}(X))^2 \\
&= E_\theta(a(X - E_\theta X) + (aE_\theta X + b - \theta))^2 \\
&= a^2 Var_\theta(X) + (aE_\theta X + b - \theta)^2 \\
&= a^2(1 - \frac{n-1}{M-1})\frac{\theta}{M}(1 - \frac{\theta}{M}) + (an\frac{\theta}{M} + b - \theta)^2 \\
&= \left(\frac{(M-na)^2}{M^2} - \frac{n(M-n)a^2}{M^2(M-1)}\right)\theta^2 + \left(\frac{n(M-n)a^2}{M(M-1)} - \frac{2(M-na)b}{M}\right)\theta + b^2\,.
\end{aligned}$$

F. To find a rule with constant risk among the linear rules, we need to equate the coefficients of θ^2 and θ in the last display above to 0: i.e.

$$\left(\frac{(M-na)^2}{M^2} - \frac{n(M-n)a^2}{M^2(M-1)}\right) = 0$$

and

$$\left(\frac{n(M-n)a^2}{M(M-1)} - \frac{2(M-na)b}{M}\right) = 0\,,$$

and solve these for $a \equiv a_0$ and $b \equiv b_0$. The first equation yields

$$(M - na_0)^2 = \frac{n(M-n)}{M-1}a_0^2$$

or equivalently (since ???)

$$M - na_0 = \sqrt{\frac{n(M-n)}{M-1}}\,a_0\,.$$

Thus

$$a_0 = \frac{M}{n(1 + \sqrt{(M-n)/n(M-1)})} \equiv \frac{M}{n(1+\delta)}$$

where

$$\delta \equiv \sqrt{\frac{M-n}{n(M-1)}}\,.$$

Then the second equation yields

$$b_0 = \frac{n(M-n)a_0^2}{2(M-na_0)(M-1)} = \frac{\delta M}{2(1+\delta)} = \frac{M-a_0 n}{2} = \frac{M}{2}\frac{\delta}{1+\delta}\,.$$

Thus with $d_0 \equiv d_{a_0,b_0}$ it follows from part E that

$$R(\theta, d_0) = b_0^2 = \frac{M^2}{4}\left(\frac{\delta}{1+\delta}\right)^2 = \frac{M^2}{4}\frac{1}{(1+1/\delta)^2} <= \frac{M^2}{4}\frac{1}{(1+\sqrt{n})^2}$$

since

$$1/\delta = \sqrt{n\frac{M-1}{M-n}} = \frac{\sqrt{n}}{\sqrt{1-\frac{n-1}{M-1}}} \geq \sqrt{n}\,.$$

G. The Bayes rule found in part E is a linear rule with

$$a = \frac{M+\alpha+\beta}{n+\alpha+\beta} \qquad \text{and} \qquad b = \frac{\alpha(M-n)}{n+\alpha+\beta}\,.$$

By setting these equal to a_0 and b_0 respectively and solving for α_0 and β_0, we find a prior distribution $\lambda(\theta; \alpha_0, \beta_0, M) \equiv \lambda_0(\theta)$ for which the Bayes rule d_0 has constant risk and hence is minimax. The equations then become:

$$a_0 = \frac{M+\alpha+\beta}{n+\alpha+\beta} \qquad \text{and} \qquad b_0 = \frac{\alpha(M-n)}{n+\alpha+\beta}\,.$$

The first equation yields

$$M+\alpha+\beta = na_0 + (\alpha+\beta)a_0$$

so that

$$\alpha+\beta = \frac{M-na_0}{a_0-1}\,.$$

Substituting this into the second equation yields

$$\alpha(M-n) = (n + \frac{M-na_0}{a_0-1})b_0 = \frac{M-n}{a_0-1}b_0\,,$$

and solving for α gives

$$\alpha = \frac{b_0}{a_0-1}\,.$$

From the second equation and the formula for b_0 in part F we obtain

$$\alpha+\beta = \frac{2b_0}{a_0-1}$$

and therefore it follows that

$$\beta = \alpha = \frac{b_0}{a_0 - 1}\,.$$

For this prior distribution the Bayes rule has constant risk and hence is minimax. When $M = 10$ and $n = 4$ we compute

$$\begin{aligned}
\delta &= \sqrt{(10-4)/4(10-1)} = \sqrt{6/36} = 1/\sqrt{6}\,, \\
a_0 &= \frac{10}{4(1+1/\sqrt{6})} = \frac{10\sqrt{6}}{4(\sqrt{6}+1)} = 3 - \frac{1}{2}\sqrt{6}\,, \\
b_0 &= \frac{10 - 4(3-\sqrt{6}/2)}{2} = \sqrt{6} - 1\,,
\end{aligned}$$

so that

$$\alpha = \beta = \frac{b_0}{a_0 - 1} = \frac{\sqrt{6}-1}{2-\sqrt{6}/2} = \frac{2+3\sqrt{6}}{5}\,,$$

and hence the estimator

$$d_0(X) = (3 - \frac{1}{2}\sqrt{6})X + (\sqrt{6}-1)$$

is minimax with constant risk

$$R(\theta, d_0) = b_0^2 = (\sqrt{6}-1)^2 = 7 - 2\sqrt{6} \doteq 2.101\ldots\,.$$

Note that the comparable minimax risk in sampling with replacement is

$$R_{wr}(\theta, d_0) = \frac{M^2}{4}\frac{1}{1+\sqrt{n})^2} = \frac{25}{(1+\sqrt{4})^2} = \frac{25}{9} \doteq 2.778\ldots\,.$$

H. The first basic thing to note is that the probability p of getting a red ball in one draw in the sampling "with replacement" situation is θ/M since the urn contains M balls, θ of which are red. Thus while the natural parameter to estimate in the "without replacement" situation is θ, the natural parameter to estimate in the "with replacement" situation is $p = \theta/M$, and hence we need to multiply the "with replacement" estimators of p by M to get estimators of θ. Here is a comparison of the sampling "with replacement" and "without replacement" situations:

quantity	with replacement	without replacement
$P_\theta(X=x)$ $\lambda(\theta)$	$\binom{n}{x}p^x(1-p)^{n-x}$ $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$	$\frac{\binom{\theta}{x}\binom{M-\theta}{n-x}}{\binom{M}{n}}$ $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\binom{M}{\theta}\frac{\Gamma(\alpha+\theta)\Gamma(M+\beta-\theta)}{\Gamma(M+\alpha+\beta)}$
$d_{MLE}(X)$	X/n	$\lfloor(M+1)X/n\rfloor$
d_B	$\frac{X+\alpha}{n+\alpha+\beta}$	$\frac{(M+\alpha+\beta)X+\alpha(M-n)}{n+\alpha+\beta}$
$R(\theta,d_M)$ $R(p,d_M)$	$\frac{M^2}{4}\frac{1}{(1+\sqrt{n})^2}$ $\frac{1}{4}\frac{1}{(1+\sqrt{n})^2}$	$\frac{M^2}{4}\frac{1}{(1+1/\delta)^2}$ $\frac{1}{4}\frac{1}{(1+1/\delta)^2}$
 $E(X)$	$\sqrt{n}<$ $np=n\theta/M$	$1/\delta=\sqrt{n\frac{M-1}{M-n}}$ $n\theta/M$
$Var(X)$	$np(1-p)$	$np(1-p)(1-\frac{n-1}{M-1})$

3. (48 points) (Testing in a 2×2 contingency table). Consider a two-dimensional contingency table with two rows and two columns. Let θ_{jk} be the cell probability for the jth row and kth column, so that $\sum_{j=1}^{2}\sum_{k=1}^{2}\theta_{jk} = 1$. Suppose that $\underline{X} = (X_{11}, X_{12}, X_{21}, X_{22}) \sim \text{Mult}_4(n; \underline{\theta} = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}))$. The following questions concern testing hypotheses about the cross-ratio $\psi \equiv \theta_{11}\theta_{22}/(\theta_{12}\theta_{21})$. Note that $\psi = 1$ if and only if $\theta_{jk} = \theta_{j\cdot}\theta_{\cdot k}$ for $j = 1, 2$ and $k = 1, 2$, i.e. independence holds; here $\theta_{j\cdot}$ and $\theta_{\cdot k}$ are the marginal cell probabilities.
A. What is the likelihood ratio test of $H_0: \ \psi = 1$ versus $H_1: \ \psi \neq 1$? What are the large-sample properties of this test under:
(a) the null hypthesis;
(b) local alternatives of the form $\psi_n = 1 + tn^{-1/2}$;
(c) fixed alternatives?
B. Describe at least one other test for testing the hypotheses in A. Compare and contrast it to the LR test.
C. How would you construct an approximate $1 - \alpha$ confidence interval for the parameter ψ?
D. Let $\Theta_0 \equiv \{\underline{\theta}: \ \psi \leq 1\}$, and $\Theta_1 \equiv \Theta_0^c = \{\underline{\theta}: \ \psi > 1\}$, Now suppose that we assume the Dirichlet($\underline{\alpha}$) prior distribution for $\underline{\theta}$; here $\underline{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$ and the prior density is of the form

$$\lambda(\underline{\theta}) = c(\underline{\alpha}) \prod_j \prod_k \theta_{jk}^{\alpha_{jk}-1} 1_{[\sum_j \sum_k \theta_{jk}=1]}(\underline{\theta}) \,.$$

Derive the Bayes test for $0 - 1$ loss of Θ_0 versus Θ_1.
[Hint: Find the posterior density of $\underline{\theta}$, and use this to show that the two variables $\theta_{11}/\theta_{1\cdot}$ and $\theta_{21}/\theta_{2\cdot}$ are independent (aposteriori) with posterior densities which are Beta$(N_{11} + \alpha_{11} - 1, N_{12} + \alpha_{12} - 1)$ and Beta$(N_{21} + \alpha_{21} - 1, N_{22} + \alpha_{22} - 1)$. Use this to show that the posterior probability that $\psi < 1$ is

$$P(\psi < 1 \Big| \underline{N}) = P(\frac{\theta_{11}}{\theta_{1\cdot}} < \frac{\theta_{21}}{\theta_{2\cdot}} \Big| \underline{N}) = \sum_{k=k_0}^{N_{12}+\alpha_{12}-1} \binom{N_{\cdot 1} + \alpha_{\cdot 1} - 1}{k} \binom{N_{\cdot 2} + \alpha_{\cdot 2} - 1}{N_{2\cdot} + \alpha_{2\cdot} - k} \Big/ \binom{N_{\cdot\cdot} + \alpha_{\cdot\cdot} - 2}{N_{1\cdot} + \alpha_{1\cdot} - 1}$$

where $k_0 \equiv 0 \vee (N_{21} + \alpha_{21} - N_{12} - \alpha_{12})$.]

Solution: A. Under the general hypothesis the MLE of θ is just the usual vector of sample proportions

$$\widehat{\theta}_n = n^{-1}\underline{X} \,.$$

Under the null hypothesis $H_0 : \psi = 1$, we use the fact that $\psi = 1$ if and only if $\theta_{jk} = \theta_{j\cdot}\theta_{\cdot k}$ to write the distribution as

$$P_\theta(\underline{X} = \underline{x}) \quad = \quad \frac{n!}{\prod_{j,k} x_{jk}} \prod_{j,k} \theta_{jk}^{x_{jk}} 1_{[\sum_{j,k} x_{j,k}=n]}$$

$$
\begin{aligned}
&= \frac{n!}{\prod_{j,k} x_{jk}} \prod_{j,k} (\theta_{j\cdot}\theta_{\cdot k})^{x_{jk}} \\
&= \frac{n!}{\prod_{j,k} x_{jk}} \prod_{j,k} \theta_{j\cdot}^{x_{jk}} \theta_{\cdot k}^{x_{jk}} \\
&= \frac{n!}{\prod_{j,k} x_{jk}} \prod_{j} \theta_{j\cdot}^{x_{j\cdot}} \prod_{k} \theta_{\cdot k}^{x_{\cdot k}} \\
&= \frac{n!}{\prod_{j,k} x_{jk}} \theta_{1\cdot}^{x_{1\cdot}} (1-\theta_{1\cdot})^{n-x_{1\cdot}} \theta_{\cdot 1}^{x_{\cdot 1}} (1-\theta_{\cdot 1})^{n-x_{\cdot 1}}
\end{aligned}
$$

since $x_{1\cdot} + x_{2\cdot} = x_{\cdot 1} + x_{\cdot 2} = n$ and $\theta_{1\cdot} + \theta_{2\cdot} = \theta_{\cdot 1} + \theta_{\cdot 2} = n$. Thus it follows easily that the MLE's of $\theta_{1\cdot}$ and $\theta_{\cdot 1}$ are given by

$$
\widehat{\theta}_{1\cdot} = \frac{X_{1\cdot}}{n} = \frac{X_{11} + X_{12}}{n} \qquad \widehat{\theta}_{\cdot 1} = \frac{X_{\cdot 1}}{n} = \frac{X_{11} + X_{21}}{n} .
$$

It follows that the MLE of θ under the null hypothesis $\psi = 1$ is given by $\widehat{\theta}^0 = (\widehat{\theta}_{11}^0, \widehat{\theta}_{12}^0, \widehat{\theta}_{21}^0, \widehat{\theta}_{22}^0)$ where

$$
\widehat{\theta}_{jk}^0 = \widehat{\theta}_{j\cdot} \cdot \widehat{\theta}_{\cdot k} = n^{-1} X_{j\cdot} n^{-1} X_{\cdot k} .
$$

The likelihood ratio is just

$$
\lambda_n = \frac{\sup_{\theta \in \Theta} L_n(\theta)}{\sup_{\theta \in \Theta_0} L_n(\theta)} = \frac{L_n(\widehat{\theta}_n)}{L_n(\widehat{\theta}_n^0)} = \prod_{j,k=1}^{2} \left(\frac{\widehat{\theta}_n}{\widehat{\theta}_n^0} \right)^{X_{jk}} ,
$$

and the likelihood ratio statistic is

$$
2 \log \lambda_n = 2 \sum_{j,k=1}^{2} X_{jk} \log \left(\frac{\widehat{\theta}_n}{\widehat{\theta}_n^0} \right) = 2 \sum_{j,k=1}^{2} X_{jk} \log (\frac{n X_{j,k}}{X_{j\cdot} X_{\cdot k}}) .
$$

(a) Our theory of likelihood based tests applies in this situation: under the null hypothesis we have $2 \log \lambda_n \rightarrow_d \chi_1^2$ since the dimension of the parameter space under the general hypothesis is $d = 3$ and the dimension of the parameter space under the null hypothesis is $d - m = 3 - 1 = 2$.
(c) Under a fixed alternative θ with $\psi \neq 1$ we have

$$
\frac{1}{n} 2 \log \lambda_n \rightarrow_p 2 \sum_{j,k=1}^{2} \theta_{jk} \log \left(\frac{\theta_{jk}}{\theta_{j\cdot}\theta_{\cdot k}} \right) = 2 \inf_{\theta_0 \in \Theta_0} K(P_\theta, P_{\theta_0}) ,
$$

where the equality can be proved directly by solving the minimization problem.
(b) Under local alternatives of the form $\psi_n = 1 + t n^{-1/2}$, it is probably easiest to

examine the limiting distribution of the Wald statistic, and then use the equivalence of the limiting distributions of the three types of statistics under local alternatives. Thus we will answer this question in the course of the solution of part C.

B. Another standard (large sample) test statistic for this testing problem is the Pearson chi-square statistic, which is essentially the Rao or score statistic for the problem. It is given by

$$R_n = \sum_{j,k=1}^{2} \frac{(X_{jk} - n\widehat{\theta}_{j\cdot}\widehat{\theta}_{\cdot k})^2}{n\widehat{\theta}_{j\cdot}\widehat{\theta}_{\cdot k}} = \sum_{j,k=1}^{2} \frac{(X_{jk} - X_{j\cdot}X_{\cdot k}/n)^2}{X_{j\cdot}X_{\cdot k}/n}\,.$$

Of course $R_n \to_d \chi_1^2$ under the null hypothesis, and $R_n \to \chi_1^2(\delta)$ under local alternatives with the same noncentrality parameter δ as for the likelihood ratio test. The behavior under fixed alternatives is different: from the weak law of large numbers and the continuous mapping theorem it follows that

$$n^{-1}R_n \to_p \sum_{j,k=1}^{2} \frac{(\theta_{jk} - \theta_{j\cdot}\theta_{\cdot k})^2}{\theta_{j\cdot}\theta_{\cdot k}} > 0$$

for $\theta \in \Theta_1 = \Theta \setminus \Theta_0$.

C. Several different strategies for finding confidence intervals for ψ will work. The most straightforward of these is via the MLE $\widehat{\psi}$ of ψ. The MLE is given by

$$\widehat{\psi} = \psi(\widehat{\theta}) = \widehat{\theta}_{11}\widehat{\theta}_{22}/\widehat{\theta}_{12}\widehat{\theta}_{21}\,.$$

Now

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to_d Z \sim N_4(0, \Sigma)$$

with $\Sigma = \mathrm{diag}(\theta) - \theta\theta^T$, and therefore it follows by the delta-method that

$$\sqrt{n}(\widehat{\psi}_n - \psi) \to_d \nabla\psi(\theta) Z \sim N(0, \sigma_\theta^2)$$

where $\sigma_\theta^2 = \nabla\psi^T \Sigma \nabla\psi$ and

$$\nabla\psi(\theta)^T = (\theta_{11}^{-1}, -\theta_{12}^{-1}, -\theta_{21}^{-1}, \theta_{22}^{-1})\psi\,.$$

Thus we compute

$$\begin{aligned}
\sigma_\theta^2 &= \nabla\psi^T \Sigma \nabla\psi \\
&= \nabla\psi^T \mathrm{diag}(\theta) \nabla\psi - \nabla\psi^T \theta\theta^T \nabla\psi \\
&= \psi^2 \sum_{j,k} \theta_{jk}^{-1} - (\nabla\psi^T\theta)^2 \\
&= \psi^2 \sum_{j,k} \theta_{jk}^{-1}\,.
\end{aligned}$$

We could form an approximate $1-\alpha$ confidence interval based on this limiting distribution for $\widehat{\psi}_n$. But the appearance of ψ^2 in the form of the variance (and the fact that ψ is a ratio of parameters) suggests that it might be better to transform to $\log\psi$: then by the delta-method again we find that

$$\sqrt{n}(\log\widehat{\psi}_n - \log\psi) \to_d \frac{1}{\psi}N(0,\sigma_\theta^2) = N(0,\psi^{-2}\sigma_\theta^2) = N(0,\tilde{\sigma}_\theta^2)$$

where $\tilde{\sigma}_\theta^2 = \sum_{j,k}\theta_{j,k}^{-1}$. Therefore an approximate $1-\alpha$ confidence interval for $\log\psi$ is given by

$$\log\widehat{\psi}_n \pm z_{\alpha/2}\frac{\sum_{j,k}\widehat{\theta}_{j,k}^{-1}}{\sqrt{n}}$$

and this yields an approximate $1-\alpha$ confidence interval for ψ by exponentiation:

$$\exp\left(\log\widehat{\psi}_n \pm z_{\alpha/2}\frac{\sum_{j,k}\widehat{\theta}_{j,k}^{-1}}{\sqrt{n}}\right) = \widehat{\psi}_n\cdot\exp(\pm z_{\alpha/2}\frac{\sum_{j,k}\widehat{\theta}_{j,k}^{-1}}{\sqrt{n}}).$$

The above arguments lead naturally to a solution for subproblem C(b) which we by-passed before: under local alternatives of the form $\psi_n = 1 + n^{-1/2}t$ we have

$$\sqrt{n}(\widehat{\psi}_n - \psi_n) \to_d N(0,\sigma_{\theta^0}^2)$$

where $\theta^0 \in \Theta_0$, and hence $\sigma_{\theta^0}^2 = \tilde{\sigma}_{\theta^0}^2 = \sum_{j,k}(\theta_{j,k}^0)^{-1}$. Therefore

$$\sqrt{n}(\widehat{\psi}_n - 1) \to_d N(t,\sigma_{\theta^0}^2),$$

and

$$\sqrt{n}(\widehat{\psi}_n - 1)/\sigma_{\theta^0} \to_d N(t/\sigma_{\theta^0},1).$$

We therefore see that under these local alternatives the Wald statistic

$$W_n = \left(\sqrt{n}(\widehat{\psi}_n - 1)/\widehat{\sigma}_\theta\right)^2 \to_d \chi_1^2(\delta)$$

where $\delta = t^2/\sigma_{\theta^0}^2 = t^2/\sum_{j,k}(\theta_{j,k}^0)^{-1}$. By equivalence of the powers of likelihood ratio and Wald statistics under local alternatives we conclude that under these local alternatives we have $2\log\lambda_n \to_d \chi_1^2(\delta)$ with this same δ.

D. When $(\underline{X}|\underline{\theta}) \sim \text{Mult}_4(n,\underline{\theta})$ and $\underline{\theta} \sim \text{Dirichlet}(\underline{\alpha})$, then we know that $(\underline{\theta}|\underline{X}) \sim \text{Dirichlet}(\underline{\alpha}+\underline{X})$; i.e.

$$\begin{aligned}
\lambda(\underline{\theta}|\underline{X}) &= \frac{\Gamma(\sum_{j,k}\alpha_{j,k}+n)}{\prod_{j,k}\Gamma(\alpha_{j,k}+X_{j,k})}\prod_{j,k}\theta_{jk}^{\alpha_{jk}+X_{jk}-1}1_{[\sum_{j,k}\theta_{jk}=1]} \\
&= c(\underline{\alpha},\underline{X})\theta_{11}^{\alpha_{11}+X_{11}-1}\theta_{12}^{\alpha_{12}+X_{12}-1}\theta_{21}^{\alpha_{21}+X_{21}-1}\theta_{22}^{\alpha_{22}+X_{22}-1} \\
&= c(\underline{\alpha},\underline{X})\left(\frac{\theta_{11}}{\theta_{1\cdot}}\right)^{\alpha_{11}+X_{11}-1}\left(1-\frac{\theta_{11}}{\theta_{1\cdot}}\right)^{\alpha_{12}+X_{12}-1} \\
&\quad\cdot\left(\frac{\theta_{21}}{\theta_{2\cdot}}\right)^{\alpha_{21}+X_{21}-1}\left(1-\frac{\theta_{21}}{\theta_{2\cdot}}\right)^{\alpha_{22}+X_{22}-1}\cdot\theta_{1\cdot}^{X_{1\cdot}+\alpha_{1\cdot}-2}\cdot(1-\theta_{1\cdot})^{X_{2\cdot}+\alpha_{2\cdot}-2};
\end{aligned}$$

Here $\theta_{j\cdot} = \theta_{j1} + \theta_{j2}$, so that $\theta_{2\cdot} = 1 - \theta_{1\cdot}$, $X_{j\cdot} = X_{j1} + X_{j2}$, and $\alpha_{j\cdot} = \alpha_{j1} + \alpha_{j2}$. We now change variables to

$$r \equiv \theta_{11}/\theta_{1\cdot}\,, \qquad s \equiv \theta_{21}/\theta_{2\cdot}\,, \qquad \text{and} \qquad t \equiv \theta_{1\cdot}\,.$$

Thus

$$\theta_{11} = rt\,, \qquad \theta_{21} = s(1-t)\,, \qquad \text{and} \qquad \theta_{12} = (1-r)t\,.$$

Then the density of (R, S, T) is given by

$$f_{R,S,T}(r,s,t|\underline{X}) = \lambda(\underline{\theta}(r,s,t)|\underline{X}) \cdot |J|$$

where

$$J = \begin{pmatrix} \frac{\partial\theta_{11}}{\partial r} & \frac{\partial\theta_{11}}{\partial s} & \frac{\partial\theta_{11}}{\partial t} \\ \frac{\partial\theta_{21}}{\partial r} & \frac{\partial\theta_{21}}{\partial s} & \frac{\partial\theta_{21}}{\partial t} \\ \frac{\partial\theta_{12}}{\partial r} & \frac{\partial\theta_{12}}{\partial s} & \frac{\partial\theta_{12}}{\partial t} \end{pmatrix} = \begin{pmatrix} t & 0 & r \\ 0 & 1-t & -s \\ -t & 0 & 1-r \end{pmatrix}\,,$$

so that

$$|J| = t(1-t)(1-r) + t(1-t)r = t(1-t)\,.$$

Therefore we find that

$$\begin{aligned} f_{R,S,T}(r,s,t) &= c(\underline{\alpha},\underline{X}) r^{\alpha_{11}+X_{11}-1}(1-r)^{\alpha_{12}+X_{12}-1} s^{\alpha_{21}+X_{21}-1}(1-s)^{\alpha_{22}+X_{22}-1} \\ &\qquad t^{X_{1\cdot}+\alpha_{1\cdot}-1}(1-t)^{X_{2\cdot}+\alpha_{2\cdot}-2} \\ &= f_R(r) f_S(s) f_T(t) \end{aligned}$$

where f_R, f_S, and f_T are all Beta densities. Thus R, S, and T are all conditionally independent given $\underline{X}$ (or independent "a posteriori"), with $R \sim \text{Beta}(\alpha_{11} + X_{11}, \alpha_{12}+X_{12})$, $S \sim \text{Beta}(\alpha_{21}+X_{21}, \alpha_{22}+X_{22})$, and $T \sim \text{Beta}(\alpha_{1\cdot}+X_{1\cdot}, \alpha_{2\cdot}+X_{2\cdot})$. Now the event

$$\begin{aligned} [\theta \in \Theta_0] &= [\psi \le 1] = [\theta_{11}\theta_{22}/\theta_{12}\theta_{21} \le 1] \\ &= [\frac{\theta_{11}}{\theta_{12}} \le \frac{\theta_{21}}{\theta_{22}}] \\ &= [\frac{\theta_{12}}{\theta_{11}} \ge \frac{\theta_{22}}{\theta_{21}}] \\ &= [\frac{\theta_{12}}{\theta_{11}} + 1 \ge \frac{\theta_{22}}{\theta_{21}} + 1] \\ &= [\frac{\theta_{12}+\theta_{11}}{\theta_{11}} \ge \frac{\theta_{22}+\theta_{21}}{\theta_{21}}] \\ &= [\frac{\theta_{11}}{\theta_{1\cdot}} \le \frac{\theta_{21}}{\theta_{2\cdot}}] = [R \le S]\,. \end{aligned}$$

Note that $(R|\underline{X}) \sim \text{Beta}(\alpha_{11}+X_{11}, \alpha_{12}+X_{12})$ has the same density as the kth order statistic of a sample of size n Uniform$(0,1)$ random variables with $k = \alpha_{11}+X_{11}$ and $n = \alpha_{1\cdot}+X_{1\cdot}-1$. Thus, with $\mathbb{G}_n(t) = n^{-1}\sum_1^n 1_{[0,t]}(U_i)$ and $U_1,\ldots,U_n$ i.i.d. Uniform$(0,1)$, we have

$$\begin{aligned}
P(R \le s|\underline{X}) &= P(U_{[n:k]} \le s) \\
&= P(n\mathbb{G}_n(s) \ge k) \\
&= \sum_{j=k}^{n} \binom{n}{j} s^j (1-s)^{n-j} \\
&= \sum_{j=\alpha_{11}+X_{11}}^{\alpha_{1\cdot}+X_{1\cdot}-1} \binom{\alpha_{1\cdot}+X_{1\cdot}-1}{j} s^j (1-s)^{\alpha_{1\cdot}+X_{1\cdot}-1-j} .
\end{aligned}$$

This yields, with

$$B(\alpha,\beta) \equiv \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \,,$$

$$\begin{aligned}
P(\theta \in \Theta_0|\underline{X}) &= P(R \le S|\underline{X}) \\
&= \int_0^1 P(R \le s|\underline{X}) f_S(s|\underline{X}) ds \\
&= \int_0^1 \sum_{j=\alpha_{11}+X_{11}}^{\alpha_{1\cdot}+X_{1\cdot}-1} \binom{\alpha_{1\cdot}+X_{1\cdot}-1}{j} s^j (1-s)^{\alpha_{1\cdot}+X_{1\cdot}-1-j} f_S(s|\underline{X}) ds \\
&= \sum_{j=\alpha_{11}+X_{11}}^{\alpha_{1\cdot}+X_{1\cdot}-1} \binom{\alpha_{1\cdot}+X_{1\cdot}-1}{j} \\
&\qquad \cdot \int_0^1 s^j (1-s)^{\alpha_{1\cdot}+X_{1\cdot}-1-j} \frac{\Gamma(\alpha_{2\cdot}+X_{2\cdot})}{\Gamma(\alpha_{21}+X_{21})\Gamma(\alpha_{22}+X_{22})} s^{\alpha_{21}+X_{21}}(1-s)^{\alpha_{22}+X_{22}-1}\, ds \\
&= \sum_{j=\alpha_{11}+X_{11}}^{\alpha_{1\cdot}+X_{1\cdot}-1} \binom{\alpha_{1\cdot}+X_{1\cdot}-1}{j} \frac{B(j+\alpha_{21}+X_{21}, \alpha_{22}+X_{22}+\alpha_{1\cdot}+X_{1\cdot}-1-j)}{B(\alpha_{21}+X_{21},\alpha_{22}+X_{22})} \\
&= \sum_{k=0}^{X_{12}+\alpha_{12}-1} \binom{\alpha_{1\cdot}+X_{1\cdot}-1}{k+\alpha_{11}+X_{11}} \frac{B(k+\alpha_{\cdot 1}+X_{\cdot 1}, \alpha_{\cdot 2}+X_{\cdot 2}-k-1)}{B(\alpha_{21}+X_{21},\alpha_{22}+X_{22})} \\
&= \sum_{k=0}^{X_{12}+\alpha_{12}-1} \binom{\alpha_{1\cdot}+X_{1\cdot}-1}{k+\alpha_{11}+X_{11}} \binom{\alpha_{\cdot 2}+X_{\cdot 2}-k}{\alpha_{12}+X_{12}-k-1} \Big/ \binom{\alpha_{\cdot\cdot}+X_{\cdot\cdot}-2}{\alpha_{1\cdot}+X_{1\cdot}-1} .
\end{aligned}$$

This is as close as I have been able to get to the claimed solution (which is from Cox and Hinkley, *Theoretical Statistics*, page 394).

4. (48 points) Suppose that $X_i = (Y_i, Z_i)$, $i = 1, \ldots, n$ are i.i.d. with

$$(Y|Z = z) \sim Binomial(m, p(z, \alpha, \beta))$$

where $\theta \equiv (\alpha, \beta) \in R \times R$ and

$$p(z, \alpha, \beta) = (1 + \exp(-(\alpha + \beta z)))^{-1},$$

suppose that the distribution H of Z is known and that Z is not degenerate at a single point. You may assume that Z is bounded: $|Z| \le c$ with probability 1 for some $c < \infty$. Consider two different procedures for estimating the parameters α and β as follows:
(i) Maximum Likelihood;
(ii) Minimum Logit-χ^2 defined as follows: Define $\text{logit}(p) = \log(p/(1-p))$, and $\widehat{p}_i = Y_i/m$, $i = 1, \ldots, n$. Choose $\theta = (\alpha, \beta)$ to minimize Berkson's "minimum logit-χ^2 $B_n(\alpha, \beta)$ defined by

$$B_n(\alpha, \beta) \equiv \sum_{i=1}^{n} \widehat{p}_i(1 - \widehat{p}_i) \left\{\alpha + \beta Z_i - \text{logit}(\widehat{p}_i)\right\}^2 .$$

Call the resulting "minimum logit-χ^2" estimator $\widetilde{\theta}_n = (\widetilde{\alpha}_n, \widetilde{\beta}_n)$. [This is a reformulation of the problem discussed on pages 123 and 124 of Ferguson, MS.]
A. Find the scores for (α, β) based on one observation $X = (Y, Z)$.
B. What is the information matrix for θ? Hint: compute conditionally on Z, and leave the information matrix in terms of expectations of functions of Z.
C. Compute the "minimum logit-χ^2" estimator $\widetilde{\theta}_n$, and show that it is consistent and asymptotically normal.
D. Prove or disprove the claim in Ferguson, page 124, lines 7 - 10: "... these estimates have all the large sample optimal properties of the maximum likelihood and minimum χ^2 estimates."
[Hint: one way to prove the claim would be to show that the maximum likelihood estimator $\widehat{\theta}_n = (\widehat{\alpha}_n, \widehat{\beta}_n)$ is asymptotically equivalent to the "minimum logit-χ^2" estimator $\widetilde{\theta}_n$ in the sense that $\sqrt{n}(\widehat{\theta}_n - \widetilde{\theta}_n) = o_p(1)$.]

Solution: A. Here the distribution of one observation (Y, Z) is given by

$$p_\theta(y, z) = \binom{m}{y} p(z; \theta)^y (1 - p(z; \theta))^{m-y} h(z)$$

for $y \in \{0, 1\}$ where $\theta = (\alpha, \beta)$ and

$$p(z; \theta) = \frac{1}{1 + \exp(-\alpha - \beta z)} .$$

Note that

$$\text{logit}(p(z;\theta)) = \log\left(\frac{p(z;\theta)}{1-p(z;\theta)}\right) = \alpha + \beta z\,.$$

Thus we have

$$\log p_\theta(y,z) = y\log p(z;\theta) + (m-y)\log(1-p(z;\theta) + \text{constant in } \theta\,,$$

and therefore

$$\dot{l}_\alpha(y,z) = \left\{\frac{y}{p(z;\theta)} - \frac{m-y}{1-p(z;\theta)}\right\}\frac{\partial}{\partial\alpha}p(z;\alpha,\beta)$$

$$\dot{l}_\beta(y,z) = \left\{\frac{y}{p(z;\theta)} - \frac{m-y}{1-p(z;\theta)}\right\}\frac{\partial}{\partial\beta}p(z;\alpha,\beta)$$

Now

$$\nabla_\theta \text{logit}(p(z;\theta)) = \frac{\partial}{\partial p}(\text{logit}(p))\nabla_\theta p(z;\theta)$$

where $\nabla_\theta \text{logit}(p(z;\theta)) = (1,z)^T$ and

$$\frac{\partial}{\partial p}(\text{logit}(p)) = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}\,.$$

It follows that

$$\nabla_\theta p(z;\theta) = \nabla_\theta \text{logit}(p(z;\theta))/\frac{\partial}{\partial p}(\text{logit}(p)) = p(z;\theta)(1-p(z;\theta))(1,z)^T\,,$$

and therefore the scores for $\theta = (\alpha,\beta)$ are given by

$$\dot{l}_\alpha(y,z) = (y - mp(z;\theta))\,,$$

$$\dot{l}_\beta(y,z) = z(y - mp(z;\theta))\,.$$

Computing second deratives yields

$$\ddot{l}_{\alpha\alpha}(y,z) = -m\frac{\partial}{\partial\alpha}p(z;\theta) = -mp(z;\theta)(1-p(z;\theta))\,,$$

$$\ddot{l}_{\alpha\beta}(y,z) = -m\frac{\partial}{\partial\beta}p(z;\theta) = -zmp(z;\theta)(1-p(z;\theta))\,,$$

and

$$\ddot{l}_{\beta\beta}(y,z) = -zm\frac{\partial}{\partial\beta}p(z;\theta) = -z^2mp(z;\theta)(1-p(z;\theta))\,.$$

Hence the information matrix $I(\theta)$ is given by

$$I(\theta) = m \begin{pmatrix} E_\theta[p(Z)(1-p(Z))] & E_\theta[Zp(Z)(1-p(Z))] \\ E_\theta[Zp(Z)(1-p(Z))] & E_\theta[Z^2 p(Z)(1-p(Z))] \end{pmatrix}$$

C. The minimum-logit-chisquare estimators $\widetilde{\theta}_n = (\widetilde{\alpha}_n, \widetilde{\beta}_n)$ minimize

$$B_n(\alpha,\beta) \equiv \sum_{i=1}^n \widehat{p}_i(1-\widehat{p}_i)\left\{\alpha + \beta Z_i - \text{logit}(\widehat{p}_i)\right\}^2 .$$

Differentiating B_n with respect to α and β yields

$$\frac{\partial}{\partial\alpha} B_n(\alpha,\beta) = 2\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i)\left(\alpha + \beta Z_i - \text{logit}(\hat{p}_i)\right)$$

and

$$\frac{\partial}{\partial\beta} B_n(\alpha,\beta) = 2\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i)\left(\alpha + \beta Z_i - \text{logit}(\hat{p}_i)\right) Z_i .$$

Setting these equal to 0, we find that the estimators $(\widetilde{\alpha}_n, \widetilde{\beta}_n)$ satisfy

$$\widetilde{\alpha}_n \sum_{i=1}^n \hat{p}_i(1-\hat{p}_i) + \widetilde{\beta}_n \sum_{i=1}^n Z_i \hat{p}_i(1-\hat{p}_i) = \sum_{i=1}^n \text{logit}(\hat{p}_i)\hat{p}_i(1-\hat{p}_i)$$

and

$$\widetilde{\alpha}_n \sum_{i=1}^n Z_i \hat{p}_i(1-\hat{p}_i) + \widetilde{\beta}_n \sum_{i=1}^n Z_i^2 \hat{p}_i(1-\hat{p}_i) = \sum_{i=1}^n Z_i \text{logit}(\hat{p}_i)\hat{p}_i(1-\hat{p}_i) .$$

Equivalently,

$$\begin{pmatrix} n^{-1}\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i) & n^{-1}\sum_{i=1}^n Z_i\hat{p}_i(1-\hat{p}_i) \\ n^{-1}\sum_{i=1}^n Z_i\hat{p}_i(1-\hat{p}_i) & n^{-1}\sum_{i=1}^n Z_i^2\hat{p}_i(1-\hat{p}_i) \end{pmatrix} \begin{pmatrix} \widetilde{\alpha}_n \\ \widetilde{\beta}_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \text{logit}(\hat{p}_i)\hat{p}_i(1-\hat{p}_i) \\ \sum_{i=1}^n Z_i \text{logit}(\hat{p}_i)\hat{p}_i(1-\hat{p}_i) \end{pmatrix},$$

so that

$$\begin{pmatrix} \widetilde{\alpha}_n \\ \widetilde{\beta}_n \end{pmatrix} = \begin{pmatrix} n^{-1}\sum_{i=1}^n \hat{p}_i(1-\hat{p}_i) & n^{-1}\sum_{i=1}^n Z_i\hat{p}_i(1-\hat{p}_i) \\ n^{-1}\sum_{i=1}^n Z_i\hat{p}_i(1-\hat{p}_i) & n^{-1}\sum_{i=1}^n Z_i^2\hat{p}_i(1-\hat{p}_i) \end{pmatrix}^{-1} \cdot \begin{pmatrix} n^{-1}\sum_{i=1}^n \text{logit}(\hat{p}_i)\hat{p}_i(1-\hat{p}_i) \\ n^{-1}\sum_{i=1}^n Z_i \text{logit}(\hat{p}_i)\hat{p}_i(1-\hat{p}_i) \end{pmatrix} .$$

At this point the statement of the problem is somewhat incomplete (and possibly misleading; my apologies for this). To obtain consistency of the minimum-logit-chisquare estimators, we need $m \to \infty$, and perhaps also $n \to \infty$. If we consider the problem with only $n \to \infty$, then the minimum-logit-chisquare estimators are apparently biased. Here is a start at examining this.

When $m \to \infty$ and n is fixed, then conditionally on Z_i,

$$\hat{p}_i = m^{-1}Y_i \to_p p(Z_i;\theta) \equiv p_i \,,$$

and, by continuous mapping,

$$\text{logit}(\hat{p}_i) = \text{logit}(m^{-1}Y_i) \to_p \text{logit}(p(Z_i;\theta) = \text{logit}(p_i) = \alpha + \beta Z_i \,.$$

Then it follows easily from the expression for $\tilde{\alpha}_n, \tilde{\beta}_n)$ given above that

$$\begin{aligned}
\begin{pmatrix} \tilde{\alpha}_n \\ \tilde{\beta}_n \end{pmatrix} &\to_p \begin{pmatrix} n^{-1}\sum_{i=1}^n p_i(1-p_i) & n^{-1}\sum_{i=1}^n Z_i p_i(1-p_i) \\ n^{-1}\sum_{i=1}^n Z_i p_i(1-p_i) & n^{-1}\sum_{i=1}^n Z_i^2 p_i(1-p_i) \end{pmatrix}^{-1} \\
&\qquad \cdot \begin{pmatrix} n^{-1}\sum_{i=1}^n \text{logit}(p_i)p_i(1-p_i) \\ n^{-1}\sum_{i=1}^n Z_i\text{logit}(p_i)p_i(1-p_i) \end{pmatrix} \\
&= \begin{pmatrix} n^{-1}\sum_{i=1}^n p_i(1-p_i) & n^{-1}\sum_{i=1}^n Z_i p_i(1-p_i) \\ n^{-1}\sum_{i=1}^n Z_i p_i(1-p_i) & n^{-1}\sum_{i=1}^n Z_i^2 p_i(1-p_i) \end{pmatrix}^{-1} \\
&\qquad \cdot \begin{pmatrix} n^{-1}\sum_{i=1}^n p_i(1-p_i) & n^{-1}\sum_{i=1}^n Z_i p_i(1-p_i) \\ n^{-1}\sum_{i=1}^n Z_i p_i(1-p_i) & n^{-1}\sum_{i=1}^n Z_i^2 p_i(1-p_i) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\
&= \begin{pmatrix} \alpha \\ \beta \end{pmatrix} .
\end{aligned}$$

Here's a start on the case of m fixed. First, note that the function

$$g(p) = p(1-p)\text{logit}(p) = p(1-p)\log\frac{p}{1-p}$$

converges to 0 as $p \to 0$ or $p \to 1$. Thus the functions $p(1-p)$ and $g(p)$ are bounded on $[0,1]$. It follows that, with $\hat{p} \equiv Y/m$, the expectations

$$E(\hat{p}(1-\hat{p})) \qquad E(|Z|\hat{p}(1-\hat{p})) \qquad E(Z^2\hat{p}(1-\hat{p}))$$

$$E(|\text{logit}|(\hat{p})\hat{p}(1-\hat{p})) \qquad E(Z|\text{logit}(\hat{p})|\hat{p}(1-\hat{p}))$$

are all finite if Z is bounded. Thus the weak (and strong) laws of large numbers apply to all the averages in the expressions for $(\tilde{\alpha}_n\tilde{\beta}_n)$ and we find, using the continuous mapping theorem, that

$$\begin{pmatrix} \tilde{\alpha}_n \\ \tilde{\beta}_n \end{pmatrix} \to_p \begin{pmatrix} E\hat{p}(1-\hat{p}) & EZ\hat{p}(1-\hat{p}) \\ EZ\hat{p}(1-\hat{p}) & EZ^2\hat{p}(1-\hat{p}) \end{pmatrix}^{-1} \begin{pmatrix} E\text{logit}(\hat{p})\hat{p}(1-\hat{p}) \\ EZ\text{logit}(\hat{p})\hat{p}(1-\hat{p}) \end{pmatrix} .$$

Now we can compute

$$E(\hat{p}(1-\hat{p})) = \frac{m}{m-1} E\left(p(Z;\theta)(1-p(Z;\theta))\right) ,$$

$$EZ\hat{p}(1-\hat{p}) = E\left\{E\left(Z\hat{p}(1-\hat{p})|Z\right)\right\} = E\left\{ZE\left(\hat{p}(1-\hat{p})|Z\right)\right\} = \frac{m}{m-1} E\left(Zp(Z;\theta)(1-p(Z;\theta))\right)$$

and similarly,

$$EZ^2\hat{p}(1-\hat{p}) = \frac{m}{m-1} E\left(Z^2 p(Z;\theta)(1-p(Z;\theta))\right) ,$$

However,

$$E\left(\text{logit}(\hat{p})\hat{p}(1-\hat{p})\right) \qquad \text{and} \qquad E\left(Z\text{logit}(\hat{p})\hat{p}(1-\hat{p})\right)$$

do not seem to be easily computable in terms of expecatations of functions of $p(Z;\theta)$. Thus we have

$$\begin{pmatrix} \tilde{\alpha}_n \\ \tilde{\beta}_n \end{pmatrix} \to_p \frac{m-1}{m} \begin{pmatrix} Ep(Z;\theta)(1-p(Z;\theta)) & EZp(Z;\theta)(1-p(Z;\theta)) \\ EZp(Z;\theta)(1-p(Z;\theta)) & EZ^2p(Z;\theta)(1-p(Z;\theta)) \end{pmatrix}^{-1} \cdot \begin{pmatrix} E\text{logit}(\hat{p})\hat{p}(1-\hat{p}) \\ EZ\text{logit}(\hat{p})\hat{p}(1-\hat{p}) \end{pmatrix}$$

and this is apparently not equal to $(\alpha,\beta)^T$ in general.

D. Not graded; and the problem was reduced from 48 to 36 points.