STAT 701 HW5 Solutions, 4/28/23

(#6.3.1). The Poisson(λ_i) structure results in the exponential family log-likelihood

$$\log f(\mathbf{y}, \vartheta) \;=\; -\sum_{i=1}^n \log y_i! \,+\, \sum_{i=1}^n \Big(y_i\,(\vartheta_1 + \vartheta_2 z_i) \,-\, e^{\vartheta_1+\vartheta_2 z_i}\Big)$$

with natural parameters ϑ_1, ϑ_2, sufficient statistics $T_1 = \sum_{i=1}^n Y_i$, $T_2 \,=\, \sum_{i=1}^n z_i\, Y_i$, and $A(\vartheta) \,=\, \sum_{i=1}^n \exp(\vartheta_1 + \vartheta_2\, z_i)$. The condition $z_1 < \cdots < z_n$ implies the information

$$I(\vartheta) \;=\; \ddot{A}(\vartheta) \;=\; \sum_{i=1}^n e^{\vartheta_1+\vartheta_2 z_i} \begin{pmatrix} 1 \\ z_i \end{pmatrix}^{\otimes 2}$$

is nonsingular, so that the MLE $\hat{\vartheta}$ uniquely solves the generalized method of moment equation

$$\sum_{i=1}^n \Big(Y_i \,-\, e^{\hat\vartheta_1+\hat\vartheta_2 z_i}\Big) \begin{pmatrix} 1 \\ z_i \end{pmatrix} \;=\; \mathbf{0} \quad\Longrightarrow\qquad e^{\hat\vartheta_1} \;=\; n\bar{Y} \Big/ \sum_{i=1}^n e^{\hat\vartheta_2\, z_i} \tag{1}$$

At $\vartheta_2 = 0$, it is easy to check that $V_{22} \;=\; (I(\vartheta)^{-1})_{22} \;=\; n\,(e^{\vartheta_1}\,(n-1)\,s_z^2)^{-1}$, which is readily estimated in terms of $\hat\vartheta$ by substituting (1). Therefore the asymptotic form of the Wald test of $H:\,\vartheta_2 = 0$ rejects when

$$|\hat\vartheta_2|\, s_z \,\Big((n-1)\,\bar{Y} \Big/ \sum_{i=1}^n e^{\hat\vartheta_2\, z_i} \Big)^{1/2} \;\ge\; z_{\alpha/2}$$

Next, it is easy to see that the restricted ML estimator for ϑ_1 under H is $\hat\vartheta_1^{(r)} \;=\; \log \bar{Y}$, so that the Rao score statistic becomes

$$\nabla_{\vartheta_2}\, \log f(\mathbf{y}, \vartheta)\Big|_{\hat\vartheta^{(r)}} \;=\; \sum_{i=1}^n Y_i\, z_i \,-\, \sum_{i=1}^n z_i\, e^{\hat\vartheta_1^{(r)}} \;=\; \sum_{i=1}^n z_i\,(Y_i - \bar{Y})$$

and the asymptotic form of the score test rejects H when this statistic in absolute value exceeds $(\bar{Y}\,\sum_{i=1}^n z_i^2)^{1/2}\, z_{\alpha/2}$.

Finally, the (generalized) likelihood ratio test rejects when

$$2\Big[\sum_{i=1}^n Y_i\,(\hat\vartheta_1 + \hat\vartheta_2\, Z_i) \,-\, \sum_{i=1}^n e^{\hat\vartheta_1+\hat\vartheta_2 z_i} \,-\, \sum_{i=1}^n Y_i\, \log\bar{Y} \,+\, n\,\bar{Y}\Big] \;\ge\; \chi^2_{1,\alpha}$$

which simplifies after a little algebra involving the substitution (1) to

$$\hat\vartheta_2 \sum_{i=1}^n Y_i\, z_i \,-\, n\,\bar{Y}\, \log(\frac{1}{n}\sum_{i=1}^n e^{\hat\vartheta_2\, z_i}) \;\ge\; \frac{1}{2}\,\chi^2_{1,\alpha}$$

(# 6.3.5). Here the log-likelihood is a constant plus

$$-\,(1/2)\,\sum_{i=1}^n \Big\{ -\,(X_i-\vartheta_1)^2 \,-\, (Y_i-\vartheta_2)^2 \Big\}$$

The special feature of this problem is the restriction to nonnegative $\vartheta_1,\,\vartheta_2$, so that the null-hypothesis values $\vartheta_1=\vartheta_2=0$ lie at one corner of the parameter space Θ, thus invalidating the conditions of Wilks' Theorem. The MLE's must by definition lie in the parameter space, so that $\hat\vartheta_1 = \bar X^+ \equiv \max(\bar X, 0)$, and similarly $\hat\vartheta_2 = \bar Y^+$. (a). The Generalized Likelihood Ratio Test-statistic for the bivariate-normal data (X_i, Y_i) with mean $\vartheta = (\vartheta_1,\vartheta_2)\,\in\,\Theta = \mathbf{R}_+^2$ and variance the 2×2 identity matrix is then

$$G \;=\; 2\,\log\lambda(\mathbf{X},\mathbf{Y}) \;=\; \sum_{i=1}^n \Big\{ -\,(X_i-\bar X^+)^2 \,-\, (Y_i-\bar Y^+)^2 \,+\, X_i^2 \,+\, Y_i^2)\Big\}$$

which is also given by

$$\sum_{i=1}^n \Big\{(2\bar X X_i - \bar X^2)\,I_{[\bar X>0]} \,+\, (2\bar Y Y_i - \bar Y^2)\,I_{[\bar Y>0]}\Big\} \;=\; n\bar X^2\,I_{[\bar X>0]} \,+\, n\bar Y^2\,I_{[\bar Y>0]}$$

By symmetry of the standard-normal random variables $\sqrt{n}\bar X,\,\sqrt{n}\bar Y$ (under H_0), the variables $I_{[\bar X>0]}$ and $\bar X^2$ are independent, and similarly for $\bar Y$. Therefore, given each of the four events (each of probability 1/4) defined by combinations $(0,0),\,(1,0),\,(0,1),\,(1,1)$ of $(I_{\bar X>0]},\,I_{\bar Y>0]})$, the conditional distribution of the GLRT statistic G is $0,\;\chi_1^2,\;\chi_1^2,\;\chi_2^2$, so

$$P(G\le t) \;=\; (1/4)\,F_{\chi_2^2}(t) \,+\, (1/2)\,F_{\chi_1^2}(t) \,+\, (1/4) \quad \forall \quad t\ge 0$$

(b). Now $\Theta = \{(\vartheta_1,\vartheta_2) : \,0\le\vartheta_2\le c\vartheta_1\} \;=\; \{(\vartheta_1,\,v\,\vartheta_1) : \,0\le v\le c,\,\vartheta_1\ge 0\}$, and the maximum of $\vartheta_1\bar X+\vartheta_2\bar Y \,-\, (\vartheta_1^2+\vartheta_2^2)/2$ is found over this Θ. The resulting statistic is

$$LR \;=\; n\bar X^2\,I_{[\bar X>0,\,\bar Y\le 0]} \,+\, n\,\frac{(\bar X+c\bar Y)^2}{1+c^2}\,I_{[\bar Y>0,\,\bar X+c\bar Y>0,\,c\bar X<\bar Y]} \,+\, n(\bar X^2+\bar Y^2)\,I_{[c\bar X>\bar Y>0]}$$

The method in this problem part is to note that we are maximizing $\vartheta_1\,(\bar X+v\bar Y) - \frac{1}{2}\,(1+v^2)\,\vartheta_1^2$, which occurs at $\vartheta_1 = (\bar X+v\bar Y)^+/(1+v^2)$, $\vartheta_2 = v\vartheta_1$ for fixed v, and leads to maximized value $((\bar X+v\bar Y)^+)^2/(2(1+v^2))$, and this value is maximized over $v\in[0,c]$ at $v = I_{[\bar Y>0,\,\bar X+c\bar Y>0]}\,c$.

In this problem part, the probability that $LR=0$ is for example

$$P(\bar Y<0,\,\bar X<0) \,+\, P(\bar Y\ge 0,\,\bar X+c\bar Y<0) \;=\; \frac{1}{4} \,+\, (\frac{1}{4} \,-\, \frac{1}{2\pi}\,\arcsin(\frac{c}{\sqrt{1+c^2}}))$$

However, the other probabilities, of $(2\pi)^{-1}\,\arcsin(c/\sqrt{1+c^2})$ that LR is equal to the χ^2_2 distributed variable $n(\bar{X}^2+\bar{Y}^2)$ or of $1/2$ that LR is equal to one of the χ^2_1 distributed statistics $n\bar{X}^2$ or $n(\bar{X}+c\bar{Y})^2/(1+c^2)$, do not in this problem part completely specify the distribution of LR since, for example, the event that $LR = n\bar{X}^2$ is dependent on the variable-value $n\bar{X}^2$.

(c) By simple re-scaling (since $\sigma^2_{1,0}$, $\sigma^2_{2,0}$ are assumed known, there is no loss of generality in assuming that the known variances are equal to 1. Then the transformation $Z_i = (\rho_0 X_i - Y_i)/\sqrt{1-\rho^2}$ makes the transformed data $(X_i,\, Z_i)$ into *iid* independent normally distributed paris with means ϑ_1 and $\vartheta^*_2 = E(Z_i)$ satisfying $\vartheta_1 \ge 0$, $(\rho_0/\sqrt{1-\rho_0^2})\,\vartheta_1 - \vartheta^*_2 \ge 0$, and the hypothesis $H:\, \vartheta_1 = \vartheta_2 = 0$ that we want to test is the same as the hypothesis that $\vartheta_1 = \vartheta^*_2 = 0$. Therefore, with $c = \rho_0/\sqrt{1-\rho_0^2}$, and with the former (X_i, Y_i) pairs in (b) replaced by (X_i, Z_i), this problem part is identical to part (b).

(#6.3.8). In the slightly different notation adopted in class, with $H_0: \theta_1 = \cdots = \theta_q = 0$, the problem is to show that $\sqrt{n}\,\Psi_n(\hat{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(\theta_0))$ under regularity conditions A0–A6, with $\hat{\theta}^{(1)} = \hat{\beta}$ denoting the estimating equation estimator for the initial q coordinates that we are testing for equality to $\underline{0}$ and $\hat{\theta}_0$ denoting the vector $\hat{\theta}_{(r)} = (\underline{0}, \hat{\theta}^{(2)}) = (0, \hat{\eta}_{(r)})$, and $\Psi_n(\theta) = n^{-1}\nabla_\beta \log L(\theta)$. Then by the Taylor's expansion given in the problem's hint,

$$\begin{aligned}\sqrt{n}\,\Psi_n(\hat{\theta}_{(r)}) \;&=\; n^{-1/2}\,\nabla_\beta \log L(\theta_0) \,+\, n^{-1}\nabla_\beta \nabla^{tr}_\eta \log L(\theta^*) n^{1/2}(\hat{\eta}((r) - \eta_0) \\ &=\; n^{-1/2}\,\nabla_\beta \log L(\theta_0) \,-\, I_{\beta,\eta}(I_{\eta,\eta})^{-1}\, n^{-1/2}\,\nabla_\eta \log L(\theta_0) \,+\, o_P(1)\end{aligned}$$

and since the score statistic $n^{-1/2}\nabla_\theta \log L(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\underline{0},\, I(\theta_0))$, we conclude

$$\sqrt{n}\,\Psi_n(\hat{\theta}_{(r)}) \,\approx\, (I_{q\times q}\,|\, I_{\beta,\eta}(I_{\eta,\eta})^{-1})\, n^{-1/2}\nabla_\theta \log L(\theta_0) \;\xrightarrow{\mathcal{D}} \mathcal{N}(\underline{0},\, I_{\beta,\beta} \,-\, I_{\beta,\eta}(I_{\eta,\eta})^{-1} I^{tr}_{\beta,\eta})$$

as desired.

(#6.4.5). Problem 6.4.5 on HW5 refers to "the result from Problem 6.2.4" but obviously means Problem 6.4.4 because 6.4.4(c) is the one where the Hypergeometric(r,c,n) distribution is derived as the conditional probability distribution of N_{11} given $R_1 = N_{11} + N_{12} = r, C_1 = N_{11} + N_{2,1} = c$, where $(N_{11}, N_{12}, N_{21} N_{22})$ is distributed Multinomial$(n, (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}))$, where $\theta_{11}/(\theta_{11}+\theta_{12}) = \theta_{21}/(\theta_{21}+\theta_{22})$.

You probably already knew that fact about the hypergeometric from undergraduate probability, but in any case that is the only way that problem 6.4.4 is relevant.

Although problem 6.4.5 states this badly, the null hypothesis being tested is the independence of $R_1 m C_1$, which problem 6.4.4(b) tells us is equivalent to the relation $\theta_{11}/(\theta_{11}+\theta_{12}) = \theta_{21}/(\theta_{21}+\theta_{22})$.

Then in 6.4.5, "exact level alpha" means that the null hypothesis is rejected with probability less than or equal to α, but the actual significance level is (because of the discreteness of the random variables) in general not exactly α.

Soln. There is very little to do here. When the θ's satisfy the H_0 relation, the definition of $j(\alpha)$ says that it is the smallest integer such that $P(N_{11} \geq j(\alpha)\,|\,R_1 = r_1, C_1 = c_1) \leq \alpha)$, and this inequality expresses that the significance level (without any approximation, or any dependence on the exact values of θ_{ij} subject to $\theta_{11}/(\theta_{11}+\theta_{12}) = \theta_{21}/(\theta_{21}+\theta_{22})$.

(#6.4.6). (a). The multinomial log-likelihood is $\sum_{i,j} N_{ij} \log \vartheta_{ij}$. Under the null hypothesis H_0 that $\vartheta_{ij} = \eta_{1i}\,\eta_{2j}$, the log-likelihood takes the decoupled form

$$\sum_{i=1}^{a} N_{i\cdot} \log \eta_{1i} + \sum_{j=1}^{b} N_{\cdot j} \log \eta_{2j} \quad , \quad R_i = N_{i\cdot} \equiv \sum_{j=1}^{b} N_{ij} \quad , \quad C_j = N_{\cdot j} \equiv \sum_{i=1}^{a} N_{ij}$$

Maximizing separately over $\{\eta_{1i}\}_i$ and $\{\eta_{2j}\}_j$ respectively under the parameter constraints that $\sum_{i=1}^{a} \eta_{1i} = 1 = \sum_{j=1}^{b} \eta_{2j}$, yields the MLE's $\hat{\eta}_{1i} = R_i/n$, $\hat{\eta}_{2j} = C_j/n$.

(b). The LRT degrees of freedom are $(ab-1)-(a-1+b-1) = (a-1)(b-1)$. Wilks' theorem gives the asymptotic $\chi^2_{(a-1)(b-1)}$ distribution under H_0 for the LR statistic

$$LR \;=\; 2\Big[-\sum_i R_i \log\frac{R_i}{n} - \sum_j C_j \log\frac{C_j}{n} + \sum_{i,j} N_{ij} \log\frac{N_{ij}}{n}\Big]$$

$$=\; 2\sum_{i,j} \frac{R_i\,C_j}{n}\Big\{\frac{n\,N_{ij}}{R_i\,C_j}\,\log\Big(\frac{n\,N_{ij}}{R_i\,C_j}\Big)\Big\}$$

By expanding the summand $g(x) = x\log x$ (with $g(1)=0,\, g'(1) = g''(1) = 1$) at $x = n\,N_{ij}/(R_i\,C_j)$ around $x=1$, this statistic becomes

$$=\; 2\sum_{i,j} \frac{R_i\,C_j}{n}\Big\{0 + \Big(\frac{n\,N_{ij}}{R_i\,C_j} - 1\Big) + \frac{1}{2}\Big(\frac{n\,N_{ij}}{R_i\,C_j} - 1\Big)^2 + o_P(1)$$

and apart from the negligible remainder, the last expression is identical to the Pearson χ^2 test statistic for row-column independence.

Extra Problem (A). Unintentionally, I made parts (b), (c) of this problem exactly repeat the corresponding problem parts of #6.3.1. The difference here is that the constant regression predcitors z_i from #6.3.1 are replaced here by random X_i with (X_i, Y_i) *iid*, and the X_i random variables themselves have distribution governed by unk nown parameter λ. So here we can calculate the per-observation information matrix from the (mixed-type) one-observation log-likelihood

$$I(a,b,\lambda) \;=\; -E\Big(\nabla^{\otimes 2}_{a,b,\lambda}\Big\{\log\lambda - \lambda X - e^{a+bX} + (a+bX)Y\Big\}\Big) \;=\; \begin{pmatrix} m_0(a,b) & m_1(a,b) & 0 \\ m_1(a,b) & m_2(a,b) & 0 \\ 0 & 0 & \lambda^{-2} \end{pmatrix}$$

where

$$m_0(a,b) \;=\; \frac{\lambda\, e^a}{\lambda - b}\;, \quad m_1(a,b) \;=\; \frac{\lambda\, e^a}{(\lambda - b)^2}\;, \quad m_2(a,b) \;=\; \frac{2\,\lambda\, e^a}{(\lambda - b)^3}$$

But note that for the upper-left block entries of the information matrix to be finite, it is necessary that $b < \lambda$. In any case, the 0's in the off-diagonal positions in the third row and column imply that (when $b < \lambda$), ML estimation of (a,b) leads to the same asymptotic variances whether or not λ is known.