**Statistics 582, Problem Set 2, Solutions**

Wellner; 1/23/98

1. This is a further continuation of problem 10.4 from 581:
A. With the given data, carry out the Wald, likelihood ratio, and Rao tests of the (composite) null hypothesis $H_0 : \beta = 1$ versus the alternative hypothesis $H_1 : \beta \neq 1$. Briefly compare the results.
B. Find what the statistics you used in part A, appropriately normalized, converge to in probability. [That is, prove analogues for this special case of the general results proved in theorem 4.2.1 for the simple null hypothesis case.]
**Solution:** A. The likelihood ratio statistic is, for the given data with $n = 19$, $2\log\lambda_n = 2.474$; the Wald statistic is $W_n = 2.7627$, and the Rao (score) statistic is $R_n = 3.2741$. [The Wald and Rao statistic were computed using the theoretical information matrix $I(\theta)$ with $\theta$ estimated by $\widehat{\theta}$ and $\widehat{\theta}_0$ respectively; see the Mathematica program for values of these statistics with the empirical (second derivative of log-likelihood) estimator instead.] Since $\chi^2_{1,.05} = 3.84$, none of the tests rejects the null hypothesis at level .05. The approximate P - values are: LR – .116, Wald – .096 ; Rao – .070. [The somewhat larger value of the Rao statistic makes me wonder if I've made a computational error in computing $R_n$; I need to check this one!]
B. First, the likelihood ratio statistic. The numerator of the likelihood ratio is

$$
\begin{aligned}
L(\widehat{\theta}_n|\underline{X}) &= \prod_{i=1}^n \frac{\widehat{\beta}}{\widehat{\alpha}}(\frac{X_i}{\widehat{\alpha}})^{\widehat{\beta}-1}\exp(-(\frac{X_i}{\widehat{\alpha}})^{\widehat{\beta}}) \\
&= (\frac{\widehat{\beta}}{\widehat{\alpha}})^n \frac{\prod_1^n X_i^{\widehat{\beta}-1}}{\widehat{\alpha}^{n(\widehat{\beta}-1)}}\exp(-(\sum_1^n X_i^{\widehat{\beta}})/\widehat{\alpha}^{\widehat{\beta}})
\end{aligned}
$$

and hence the logarithm of the numerator is

$$
\begin{aligned}
l(\widehat{\theta}_n|\underline{X}) = n\log(\widehat{\beta}) \quad &- \quad n\log(\widehat{\alpha}) + (\widehat{\beta}-1)\sum_1^n \log X_i \\
&- n(\widehat{\beta}-1)\log(\widehat{\alpha}) - \frac{\sum_1^n X_i^{\widehat{\beta}}}{\widehat{\alpha}^{\widehat{\beta}}}.
\end{aligned}
$$

The logarithm of the denominator of the statistic is

$$
l(\widehat{\theta}_n^0|\underline{X}) = \log\left(\prod_{i=1}^n \frac{1}{\overline{X}}\exp(-X_i/\overline{X})\right) = -n\log(\overline{X}) - n\,.
$$

Thus, using $\widehat{\alpha}^{\widehat{\beta}} = n^{-1}\sum_1^n X_i^{\widehat{\beta}}$,

$$
\begin{aligned}
\frac{2}{n}\log\lambda_n &= \frac{2}{n}\left\{ l(\widehat{\theta}_n|\underline{X}) - l(\widehat{\theta}_n^0|\underline{X})\right\} \\
&= 2\{\log(\widehat{\beta}) - \log(\widehat{\alpha}) + (\widehat{\beta}-1)\frac{1}{n}\sum_{i=1}^n \log(X_i) - (\widehat{\beta}-1)\log(\widehat{\alpha}) + \log(\overline{X})\} \\
&\to_p 2\{\log\beta - \log\alpha + (\beta-1)E(\log X) - (\beta-1)\log\alpha + \log(\alpha\Gamma(1+1/\beta))\} \\
&= 2\{\log\beta - (1-\frac{1}{\beta})\gamma + \log\Gamma(1+\frac{1}{\beta})\} \\
&\qquad \text{using} \;\; E(-\log X) = \frac{\gamma}{\beta} - \log(\alpha) \\
(0.1) &= 2\inf_{\theta_0\in\Theta_0} K(P_\theta, P_{\theta_0}) \equiv \mathrm{lr}(\beta).
\end{aligned}
$$

Here's a proof of this last equality:

$$K(P_\theta, P_{\theta_0}) = E_\theta \log\frac{p_\theta}{p_{\theta_0}}(X)$$

where

$$p_\theta(x) = \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1}\exp(-(\frac{x}{\alpha})^\beta)1_{(0,\infty)}(x), \qquad p_{\theta_0}(x) = \frac{1}{\alpha_0}\exp(-x/\alpha_0)1_{(0,\infty)}(x).$$

Therefore

$$
\begin{aligned}
\log\frac{p_\theta}{p_{\theta_0}}(X) &= \log\beta + \log(\frac{\alpha}{\alpha_0}) + \frac{\beta-1}{\beta}\log(\frac{X}{\alpha})^\beta - \{(\frac{X}{\alpha})^\beta - \frac{X}{\alpha_0}\} \\
&=_d \log\beta + \log(\frac{\alpha}{\alpha_0}) + \frac{\beta-1}{\beta}\log(Y) - \{Y - \frac{\alpha}{\alpha_0}Y^{1/\beta}\}
\end{aligned}
$$

where $Y \sim \exp(1)$. Thus we compute, using $E\{-\log(Y)\} = \gamma$, Euler's constant, and $E(Y^{1/\beta}) = \Gamma(1+1/\beta)$,

$$
\begin{aligned}
K(P_\theta, P_{\theta_0}) &= \log\beta + \log(\frac{\alpha}{\alpha_0}) - (1-\frac{1}{\beta})\gamma - \{1 - \frac{\alpha}{\alpha_0}\Gamma(1+1/\beta)\} \\
&\equiv g(\alpha_0)
\end{aligned}
$$

Hence

$$g'(\alpha_0) = \frac{1}{\alpha_0} - \frac{\alpha}{\alpha_0^2}\Gamma(1+\frac{1}{\beta})$$

which equals 0 if $\alpha_0 = \alpha_{0min} \equiv \alpha\Gamma(1+1/\beta)$. Furthermore,

$$g''(\alpha_{0min}) = 1/(\alpha^2\Gamma^2(1+1/\beta)) > 0\,,$$

, so

$$\begin{aligned}\inf_{\theta_0\in\Theta_0} K(P_\theta, P_{\theta_0}) &= \inf_{\alpha_0} g(\alpha_0) = g(\alpha_{0min}) \\ &= \log\beta - (1-\frac{1}{\beta})\gamma + \log\Gamma(1+\frac{1}{\beta}),\end{aligned}$$

proving (0.1) The equality (0.1) is a special case of a general result due to Bahadur (1965); see Bahadur (1971), *Some Limit Theorems in Statistics*, pages 37 - 40.

The behavior of the Wald statistic under fixed alternatives is much easier: by consistency of $\widehat{\beta}$ and the continuous mapping theorem

$$\begin{aligned}\frac{1}{n}W_n &= (\widehat{\theta}_{n2}-\theta_{20})\hat{I}_{22\cdot 1}(\widehat{\theta}_{n2}-\theta_{20}) \\ &= (\widehat{\beta}-1)^2\widehat{I}_{22\cdot 1} = (\widehat{\beta}-1)^2\frac{\pi^2}{6\widehat{\beta}^2} \\ &\to_p \frac{\pi^2}{6}(1-\frac{1}{\beta})^2 \equiv w(\beta).\end{aligned}$$

[Question: What is the general form of the limit – analogous to (0.1)?]

Finally the Rao statistic:

$$\frac{1}{n}R_n = \frac{1}{n}Z_n^T(\widehat{\theta}_n^0)\hat{I}^{-1}(\widehat{\theta}_n^0)Z_n(\widehat{\theta}_n^0)$$

where

$$\frac{1}{\sqrt{n}}Z_n(\widehat{\theta}_n^0) = \begin{pmatrix} 0 \\ 1-\frac{1}{n}\sum_{i=1}^n(\frac{X_i}{\overline{X}}-1)\log(\frac{X_i}{\overline{X}}) \end{pmatrix}.$$

Since $I_{22\cdot 1}^{-1}(\widehat{\theta}_n^0) = \frac{6}{\pi^2}$, we find that

$$\begin{aligned}\frac{1}{n}R_n &= \left\{1-\frac{1}{n}\frac{1}{\overline{X}}\sum_{i=1}^n(X_i-\overline{X})(\log X_i - \log\overline{X})\right\}^2\frac{6}{\pi^2} \\ &\to_p \left\{1-\frac{1}{E(X)}\{E(X\log(X)) - E(X)E(\log X)\}\right\}^2\frac{6}{\pi^2}. \qquad (0.2)\end{aligned}$$

But

$$E\{-\log X\} = \frac{\gamma}{\beta} - \log\alpha, \qquad EX = \alpha\Gamma(1+1/\beta),$$

and

$$\begin{aligned}E(X\log X) &= \frac{\alpha}{\beta}E\{Y^{1/\beta}\log Y\} + \alpha\Gamma(1+1/\beta)\log\alpha \\ &= \frac{\alpha}{\beta}\psi(1+1/\beta)\Gamma(1+1/\beta) + \alpha\Gamma(1+1/\beta)\log\alpha\end{aligned}$$

where $\psi$ is the digamma function, $\psi(z) = (d/dz)\log\Gamma(z)$. Hence we find that

$$\frac{1}{n}R_n \rightarrow_p \{1 - \frac{\gamma}{\beta} - \frac{1}{\beta}\psi(1+\frac{1}{\beta})\}^2\frac{6}{\pi^2} \equiv r(\beta)\,.$$

[Question: What is the general form of the limit analogous to (0.1)?] See the plots of the three functions $lr(\beta)$, $w(\beta)$, and $r(\beta)$ at the end of this solution set.

2. Consider the following parametric model: suppose that

$$N = (N_{ij} : 1 \le i \le r,\ 1 \le j \le s) \sim \text{Multinomial}_{rs}(n, \underline{p} = (p_{ij} : 1 \le i \le r,\ 1 \le j \le s))\,.$$

Here $\sum_{i=1}^r \sum_{j=1}^s p_{ij} = 1$.
A. What is the maximum likelihood estimator $\underline{\widehat{p}}$ of $\underline{p}$? What is the limiting distribution of $\sqrt{n}(\widehat{p} - p)$?
B. Suppose we consider estimation of $\underline{p}$ in the restricted model in which the marginal probabilities are known: $p_{i\cdot} \equiv \sum_{j=1}^s p_{ij} = p_{i\cdot}^0$, $i = 1, \ldots, r$ and $p_{\cdot j} \equiv \sum_{i=1}^r p_{ij} = p_{\cdot j}^0$, $j = 1, \ldots, s$ where the $p_{i\cdot}^0$'s and $p_{\cdot j}^0$'s are known. Show that the maximum likelihood estimator $\widehat{p}$ has the form $\widehat{p}_{ij} = (N_{ij}/n)/(a_i + b_j)$. What can you say about the limiting distribution of $\sqrt{n}(\widehat{p} - p)$ in this (sub)model?

**Solution:** A. In the full model, the MLE $\underline{\widehat{p}}$ is just the usual vector of sample proportions, $\widehat{p}_{ij} = N_{ij}/n$. If we think of this as one long vector of length $rs$, then the asymptotic distribution is, by the multivariate CLT, $N_{rs}(0, \Sigma)$ where $\Sigma = \text{diag}(\underline{p}) - \underline{p}\underline{p}^T$. Here $\Sigma$ is a matrix with $rs$ rows and columns.

B. First we consider the model with just one marginal distribution known, say $p_{i\cdot} = p_{i\cdot}^0$, $i = 1, \ldots, r$. First note that $\sum_i p_{i\cdot}^0 = 1$ implies that $\sum_{i,j} p_{ij} = 1$ is forced by the marginal constraints. Thus we will introduce Lagrange multipliers $a_i$, $i = 1, \ldots, r$, one for each constraint, and maximize

$$\begin{aligned} l(p, a, ) &= \sum_{i,j} N_{ij}\log p_{ij} + \log(\frac{n!}{\prod_{i,j} N_{ij}!}) \\ &\quad + \sum_i a_i(\sum_j p_{ij} - p_{i\cdot}^0)\,. \end{aligned}$$

Here $a \in R^r$. Thus the (constrained) likelihood equations become

$$\frac{\partial}{\partial p_{ij}} l(p, a, b) = \frac{N_{ij}}{p_{ij}} + a_i + = 0, \qquad i = 1, \ldots, r, \quad j = 1, \ldots, s, \tag{0.3}$$

$$\frac{\partial}{\partial a_i} l(p, a, b) = \sum_j p_{ij} - p_{i\cdot}^0 = 0, \quad i = 1, \ldots, r, \tag{0.4}$$

Solving (0.3) yields, with $\widehat{p}_{ij} \equiv N_{ij}/n$,

$$\widehat{p}_{ij}^{ML} = -\frac{N_{ij}}{a_i}. \tag{0.5}$$

Plugging (0.5) into (0.4) yields

$$-\sum_j \frac{\widehat{p}_{ij}}{\tilde{a}_i} = p_{i\cdot}^0, \quad i = 1, \ldots, r, \tag{0.6}$$

or

$$-a_i = \frac{n\widehat{p}_{i\cdot}}{p_{i\cdot}^0}. \tag{0.7}$$

Thus we find that

$$\widehat{p}_{ij}^{ML} = N_{ij}\frac{p_{i\cdot}^0}{n\widehat{p}_{i\cdot}} = \widehat{p}_{ij}\frac{p_{i\cdot}^0}{\widehat{p}_{i\cdot}}. \tag{0.8}$$

To study the asymptotic behavior of this estimator, it is helpful to consider the tangent space of the model $\mathcal{P}_0$ first in order to understand what we expect to obtain as the influence function of an efficient estimator. With one known marginal distribution, i.e. the constraints $p_{i\cdot} = p_{i\cdot}^0$, $i = 1, \ldots, r$, $\dot{\mathcal{P}}_0$ of the model $\mathcal{P}_0$ Suppose that $p_\eta = (p_{\eta,ij})$ is a one parameter submodel of $\mathcal{P}_0$. Then since the marginals $\{p_{i\cdot}^0\}$ are known, for any function $g : \{1, \ldots, r\} \to R$,

$$E_\eta g(I) = \sum_{i,j} g(i)p_{\eta,ij} = \sum_i g(i)p_{i\cdot}^0 = \text{constant in } \eta\,.$$

Thus, differentiation wrt $\eta$ yields

$$\sum_{i,j} g(i)\dot{l}_\eta(i,j)p_{ij} = Eg(I)\dot{l}_\eta(I,J) = 0\,,$$

Since this is true for all the score functions $\dot{l}_\eta$ in the model $\mathcal{P}_0$, we deduce that

$$\dot{\mathcal{P}}_0 \perp \mathbf{H}_I \equiv \{g(i) : Eg(I) = 0\}\,,$$

In fact

$$\dot{\mathcal{P}}_0 = \mathbf{H}_I^\perp \equiv \{h(I,J) \in L_2^0(P) : E\{a(I)h(I,J)\} = 0 \qquad \text{for all } \ a \in \mathbf{H}_I\}.$$

Thus

$$\tilde{l} = \prod(\psi|\dot{\mathcal{P}}_0) = \psi - E(\psi|I)$$

and since $\psi = \psi - E(\psi|I) + E(\psi|I) = \tilde{l} + E(\psi|I)$,

$$\begin{aligned}\text{diag}(\underline{p}) - \underline{pp}^T &= E\psi(I,J)\psi^T(I,J) \\ &= E\{Cov(\psi|I)\} + Cov\{E(\psi|I)\}.\end{aligned}$$

A little calculation shows that

$$E(\psi_{ij}|I=i') = \frac{p_{i'j}}{p_{i'\cdot}} 1_{[i'=i]} - p_{ij}\,,$$

so that

$$\begin{aligned} \widetilde{l}_{ij}(I,J) &= 1_{[I=i,J=j]} - p_{ij} - \frac{p_{ij}}{p_{i\cdot}}(1_{[I=i]} - p_{i\cdot}) \\ &= 1_{[I=i,J=j]} - \frac{p_{ij}}{p_{i\cdot}} 1_{[I=i]}. \end{aligned}$$

Note that the (constrained) ML estimator $\widehat{p}^{ML}$, is, in this case,

$$\widehat{p}_{ij}^{ML} = N_{ij}\frac{p_{i\cdot}^0}{N_{i\cdot}} = \widehat{p}_{ij}\frac{p_{i\cdot}^0}{\widehat{p}_{i\cdot}}\,.$$

This estimator is asymptotically linear with influence function $\widetilde{l}$ since we can write:

$$\begin{aligned} \sqrt{n}(\widehat{p}_{ij}^{ML} - p_{ij}) &= \frac{p_{i\cdot}^0}{\widehat{p}_{i\cdot}}\sqrt{n}(\widehat{p}_{ij} - p_{ij}) - \frac{p_{ij}}{\widehat{p}_{i\cdot}}\sqrt{n}(\widehat{p}_{i\cdot} - p_{i\cdot}^0) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{l}_{ij}(I_k, J_k) + o_p(1)\,. \end{aligned}$$

Now we consider the model $\mathcal{P}_{00}$ with two known marginal distributions. In the constrained model $\mathcal{P}_{00}$, the (raw) log - likelihood is the same as above, but the constraints need to be imposed with additional Lagrange multipliers, one for each constraint. Note, however, that $\sum_i p_{i\cdot}^0 = 1 = \sum_j p_{\cdot j}^0$ implies that $\sum_{i,j} p_{ij} = 1$ is forced by both of the marginal constraints, and is therefore redundant. Thus we will maximize

$$\begin{aligned} l(p,a,b) &= \sum_{i,j} N_{ij}\log p_{ij} + \log\left(\frac{n!}{\prod_{i,j} N_{ij}!}\right) \\ &\quad + \sum_i a_i\left(\sum_j p_{ij} - p_{i\cdot}^0\right) + \sum_j b_j\left(\sum_i p_{ij} - p_{\cdot j}^0\right). \end{aligned}$$

Here $a \in R^r$, $b \in R^s$. Thus the (constrained) likelihood equations become

(0.9) $$\frac{\partial}{\partial p_{ij}} l(p,a,b) = \frac{N_{ij}}{p_{ij}} + a_i + b_j = 0, \qquad i = 1,\ldots,r, \quad j = 1,\ldots,s,$$

(0.10) $$\frac{\partial}{\partial a_i} l(p,a,b) = \sum_j p_{ij} - p_{i\cdot}^0 = 0, \quad i = 1,\ldots,r,$$

$$(0.11) \qquad \frac{\partial}{\partial b_j} l(p,a,b) = \sum_i p_{ij} - p^0_{\cdot j} = 0, \quad i = 1, \ldots, s.$$

Solving (0.9) yields, with $\widehat{p}_{ij} \equiv N_{ij}/n$,

$$(0.12) \qquad \widehat{p}^{ML}_{ij} = -\frac{N_{ij}}{a_i + b_j} = \frac{\widehat{p}_{ij}}{\tilde{a}_i + \tilde{b}_j}$$

with $\tilde{a}_i \equiv -a_i/n$, $\tilde{b}_j \equiv -b_j/n$. Plugging (0.12) into (0.10) and (0.11) yields

$$(0.13) \qquad \sum_j \frac{\widehat{p}_{ij}}{\tilde{a}_i + \tilde{b}_j} = p^0_{i\cdot}, \quad i = 1, \ldots, r,$$

and

$$(0.14) \qquad \sum_i \frac{\widehat{p}_{ij}}{\tilde{a}_i + \tilde{b}_j} = p^0_{\cdot j}, \quad j = 1, \ldots, s.$$

This gives a set of $r+s$ nonlinear equations in the $r+s$ variables $\tilde{a}_i$, $\tilde{b}_j$ which apparently cannot be solved in closed form. [Could these be solved by Newton - Raphson iteration? What do these equations become when they are linearized?]

To describe the behavior of this estimator (assuming we can compute it), we need to consider the tangent space $\dot{\mathcal{P}}_{00}$ of the model $\mathcal{P}_{00}$ Suppose that $p_\eta = (p_{\eta,ij})$ is a one parameter submodel of $\mathcal{P}_{00}$. Then since the marginals $\{p^0_{i\cdot}\}$ and $\{p^0_{\cdot j}\}$ are known, for any functions $g : \{1, \ldots, r\} \to R$, $h : \{1, \ldots, s\} \to R$,

$$E_\eta g(I) = \sum_{i,j} g(i) p_{ij} = \sum_i g(i) p^0_{i\cdot} = \text{constant in } \eta\,,$$

$$E_\eta h(J) = \sum_{i,j} h(j) p_{ij} = \sum_j h(j) p^0_{\cdot j} = \text{constant in } \eta\,,$$

Thus, differentiation wrt $\eta$ yields

$$\sum_{i,j} g(i) \dot{l}_\eta(i,j) p_{ij} = E g(I) \dot{l}_\eta(I,J) = 0\,,$$

and

$$\sum_{i,j} h(j) \dot{l}_\eta(i,j) p_{ij} = E h(J) \dot{l}_\eta(I,J) = 0\,,$$

Since this is true for all the score functions $\dot{l}_\eta$ in the model $\mathcal{P}_{00}$, we deduce that

$$\dot{\mathcal{P}}_{00} \perp \mathbf{H}_I \equiv \{g(i) : E g(I) = 0\}\,,$$

and

$$\dot{\mathcal{P}}_{00} \perp \mathbf{H}_J \equiv \{h(j) : E h(J) = 0\}\,.$$

In fact

$$\dot{\mathcal{P}}_0 = (\mathbf{H}_I + \mathbf{H}_J)^{\perp}\,.$$

Now $\hat{p} = N_n/n$ is a linear estimator with

$$\sqrt{n}(\widehat{p} - p) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\psi(I_k, J_k) \tag{0.15}$$

where $\psi(I,J) = (\psi_{ij}(I,J)) = (1[I=i, J=j] - p_{ij})$; both sides of (0.15) have $rs$ coordinates. Note that $\psi \notin \dot{\mathcal{P}}_0$. Hence to find the efficient influence function $\tilde{l}$ for estimating $\underline{p}$ in $\mathcal{P}_{00}$ we want to project $\psi$ onto $\dot{\mathcal{P}}_{00}$. Now projection onto $\mathbf{H}_I$ and $\mathbf{H}_J$ separately is easy:

$$\prod(\psi|\mathbf{H}_I) = E(\psi(I,J)|I) \equiv \prod_I \psi,$$

and

$$\prod(\psi|\mathbf{H}_J) = E(\psi(I,J)|J) \equiv \prod_J \psi\,.$$

It is easy to see that $a + b \equiv a(I) + b(J)$ is the projection of $\psi$ onto $\mathbf{H}_I + \mathbf{H}_J$ if and only if

$$a(I) = \prod_I(\psi - b), \tag{0.16}$$

and

$$b(J) = \prod_J(\psi - b)\,. \tag{0.17}$$

Then

$$\prod_I(\psi - a - b), = \prod_I(\psi - b) - a = 0$$

from (0.16) and

$$\prod_J(\psi - a - b) = \prod_J(\psi - a) - b = 0$$

by (0.17). Then for any $c \in \mathbf{H}_I$, $d \in \mathbf{H}_J$,

$$\begin{aligned}
&E\{(\psi - a - b)(c+d)\}\\
&= E\{c(I)(\psi(I,J) - a(I) - b(J))\} + E\{d(J)(\psi(I,J) - a(I) - b(J))\}\\
&= E\{c(I)E(\psi(I,J) - a(I) - b(J)|I)\} + E\{d(J)E(\psi(I,J) - a(I) - b(J)|J)\}\\
&= E\{c(I)\prod_I(\psi(I,J) - a(I) - b(J))\} + E\{d(J)\prod_J(\psi(I,J) - a(I) - b(J))\}\\
&= 0\,.
\end{aligned}$$

Note that we can rewrite (0.16) and (0.17) as

$$a(i) = \sum_j p_{ij}(\psi(i,j) - b(j))/p^0_{i\cdot}, \quad i = 1, \ldots, r, \tag{0.18}$$

and

(0.19) $$b(j) = \sum_j p_{ij}(\psi(i,j) - a(i))/p^0_{\cdot j}, \quad j = 1, \ldots, s\,.$$

For fixed $p_{ij}$ this is a system of $r+s$ linear equations in the $r+s$ unknowns $a(i), b(j)$.

3. Suppose, as in Example 4.3.10, that $\underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $\text{Mult}_k(1, \underline{p})$, so that $\underline{N}_n \equiv \sum_{j=1}^n \underline{X}_j \sim \text{Mult}_k(n, \underline{p})$. Use Jensen's inequality to show that the log-likelihood
$$l_n(\underline{p}|\underline{X}) = \sum_{j=1}^k N_j \log p_j + \sum_{i=1}^n \log(\frac{1!}{X_{i1}! \cdots X_{ik}!})$$
is maximized by $\widehat{\underline{p}} = \underline{N}_n/n$. [Hint: write the first term of $l_n(\underline{p}|\underline{X})$ as $n\sum_{j=1}^k \widehat{p}_j \log p_j$.]

**Solution:** The loglikelihood is clearly as claimed, and hence, to show that it is maximized by $\widehat{\underline{p}} = \underline{N}_n/n$, it suffices to show that

$$\sum_{j=1}^k \widehat{p}_j \log p_j \le \sum_{j=1}^k \widehat{p}_j \log \widehat{p}_j$$

with equality if and only if $\underline{p} = \widehat{\underline{p}}$. But this is equivalent to showing that

(0.20) $$\sum_{j=1}^k \widehat{p}_j \log\left(\frac{p_j}{\widehat{p}_j}\right) \le 0$$

with equality if and only if $\underline{p} = \widehat{\underline{p}}$. Consider a random variable $Y$ with

$$P(Y = \frac{p_j}{\widehat{p}_j}) = \widehat{p}_j$$

for $j = 1, \ldots, k$. Then $E(Y) = \sum_j p_j = 1$ and the above sum is just $E \log Y$. By strict convexity of the function $-\log(y)$, it follows from Jensen's inequality that

$$E \log Y \le \log(E(Y)) = \log(1) = 0$$

with equality if and only if $\underline{p} = \widehat{\underline{p}}$. [Note that (0.20) is equivalent to $K(\widehat{\underline{p}}, \underline{p}) \ge 0$.]