Stat 700 HW 7 Solutions, F'09

Bickel-Doksum, #2.2.35. Differntiate the log-likelihood with respect to ϑ and multiply through by $(1+(x_1-\vartheta)^2)(1+(x_2-\vartheta)^2)$ to get the likelihood equation

$$0 \;=\; 2(x_1-\vartheta)\,(1+(x_2-\vartheta)^2)+2(x_2-\vartheta)\,(1+(x_1-\vartheta)^2) \;=\; 2(\bar{x}-\vartheta)\,(1+(x_1-\vartheta)(x_2-\vartheta))$$

$$=\; 2(\bar{x}-\vartheta)\,(1+(\bar{x}+\Delta-\vartheta)(\bar{x}-\Delta-\vartheta)) \;=\; 2(\bar{x}-\vartheta)\,(1+(\bar{x}-\vartheta)^2-\Delta^2)$$

(a). First, if $|\Delta| \;=\; |x_1-x_2|/2 \le 1$, then the second factor in the last expression is positive, and the only root of the likelihood equation is $\bar{x}=\vartheta$.

(b). Second, if $|\Delta|>1$, then in addition to the root $\vartheta=\bar{x}$ the likelihood equation has the roots $\vartheta=\bar{x}\pm\sqrt{\Delta^2-1}$.

Bickel-Doksum, #2.3.1. Here we have an exponential family with

$$f(\mathbf{x},\mathbf{y},\alpha,\beta) \;=\; \exp(\sum_{i=1}^{n}(\alpha+\beta x_i)\,y_i)\,/\,\prod_{i=1}^{n}(1+e^{\alpha+\beta x_i})\,, \quad T(\mathbf{X}) \;=\; \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$$

By inspection, the exponential family has open parameter space $\vartheta=(\alpha,\beta)\in\mathbf{R}^2$, and is identifiable with sufficient statistic T of rank 2. So we are only checking the condition (2.3.2) to verify whether, in terms of the specific observed values $\mathbf{y}=(y_i,\ i=1,\ldots,n)$, the MLE exists or not. The Hint, which is very easy to prove, is intended to help in that verification.

Assume the hint, and let $m=\sum_i y_i$, $z=\sum_i y_i\,x_i$ be the sufficient statistic values for the observed data, and note that there is no loss in generality in assuming all $x_i>0$. (Otherwise, by adding $1-x_1$ to all of the terms x_i, note that precisely the same model holds with α replaced by $\alpha-\beta(1-x_1)$.) We have in (2.3.2) the necessary and sufficient condition for existence of the MLE that for all constants c_1, c_2 , $P(\sum_i Y_i(c_1+c_2x_i) > c_1m+c_2z) > 0$. When c_1, c_2 are both positive, this says that $\mathbf{y}$ must not consist only of 1's, and when both $c_1, c_2<0$, (2.3.2) says that $\mathbf{y}$ must not consist only of 0's. With $\mathbf{y}\neq\mathbf{0},\mathbf{1}$ and $c_2>0>c_1$, the only way for the probability to be positive is for $\sum_i Y_i$ to be m or smaller, but for the indices with $Y_i=1$ to be located at x_i's with larger values, and this has positive probability only if $\mathbf{y}$ does not consist of a block of 0's followed by a block of 1's. (With such $\mathbf{y}$, there is no way for there to be a larger sum of x_i's among a fixed number of indices where $y_i=1$.) Similarly, when $c_1>0>c_2$, the only way for the probability to be positive is for $\mathbf{y}$ not to be a block of 1's followed by a block of 0's.

Bickel-Doksum, #2.3.7. The exponential family form for the density of each Y_i does not help much here. However, the log-likelihood is explicitly given and strictly concave:

$$\text{logLik}(\mathbf{Y},\alpha,\beta) \;=\; n\alpha \,+\, n\beta\,\bar{z} \,-\, \sum_{i=1}^{n} e^{\alpha+\beta z_i}\,Y_i$$

Concavity implies that the unique MLE can be found as the solution of the likelihood equations defined by setting partial derivatives equal to 0. The two equations are

$$n \;=\; \sum_{i=1}^{n} e^{\alpha+\beta z_i}\, Y_i \quad , \qquad n\,\bar{z} \;=\; \sum_{i=1}^{n} e^{\alpha+\beta z_i}\, z_i\, Y_i$$

These two equations can be written equivalently as:

$$\sum_{i=1}^{n} (z_i - \bar{z})\, Y_i\, e^{\beta(z_i-\bar{z})} \;=\; 0 \;\; , \quad e^{\alpha} \;=\; n/\sum_{i=1}^{n} Y_i\, e^{\beta z_i}$$

The first of these equations has a solution by the intermediate value theorem, since as $\beta \to -\infty$, the left-hand side converges to $-\infty$, while as $\beta \to \infty$, the left-hand side converges to $+\infty$. In terms of the solution $\hat{\beta}$ of the first equation, $\hat{\alpha}$ is evidently determined as $\log(n/\sum_{i=1}^{n} Y_i \exp(\hat{\beta} z_i))$. Strict concavity of the log-likelihood implies the solution $(\hat{\alpha},\, \hat{\beta})$ determined in this way is unique.

Bickel-Doksum, #2.4.4. (a). First,

$$f_{\mathbf{X}}(\mathbf{j},\mathbf{y}) \;=\; \lambda^{\sum j_i}\, (1-\lambda)^{\sum (1-j_i)}\, (2\pi)^{-n/2}\, \sigma_1^{-\sum j_i}\, \sigma_0^{-\sum (1-j_i)} \,\cdot$$

$$\cdot \exp\Big(-\, \frac{1}{2\sigma_1^2} \sum_{i=1}^{n} j_i\, (y_i-\mu)^2 \;-\; \frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (1-j_i)\, (y_i-\mu)^2 \Big)$$

which by inspection gives the sufficient statistics

$$\frac{1}{\sigma_1^2} \sum_{i=1}^{n} Y_i\, I_i \;+\; \frac{1}{\sigma_1^2} \sum_{i=1}^{n} Y_i\, (1-I_i) \qquad \text{for} \quad \mu$$

$$\sum I_i \qquad \text{for} \qquad \log \frac{\lambda}{1-\lambda} \;-\; \frac{\mu^2}{2\sigma_1^2} \;+\; \frac{\mu^2}{2\sigma_0^2}$$

(b). These sufficient statistics are minimal because complete and rank 2 (i.e., parameters are identifiable in open region).

(c). For ML, solve $\hat{\lambda} = \sum_i I_i/n$ and

$$0 \;=\; \frac{1}{\sigma_1^2} \sum_i I_i(Y_i - \hat{\mu}) \;+\; \frac{1}{\sigma_0^2} \sum_i (1-I_i)(Y_i - \hat{\mu})$$

which implies the existence of MLE always:

$$\hat{\mu} \;=\; \Big(\sum_i I_i Y_i/\sigma_1^2 \;+\; \sum_i (1-I_i) Y_i/\sigma_0^2 \Big) \Big/ \Big(\sum_i I_i/\sigma_1^2 \;+\; \sum_i (1-I_i)/\sigma_0^2 \Big)$$

Bickel-Doksum, #2.4.9. In this problem, $X \;=\; (U,V,W)$ is a discrete random variable with values in $\{1,\ldots,A\} \times \{1,\ldots,B\} \times \{1,\ldots,C\}$. The cell probabilities are $p_{abc} = P(X = (a,b,c)) = P(U = a, V = b, W = c)$. First, if $\log p_{abc} = \mu_{ac} + \nu_{bc}$, then

$$P(U=a, V=b|W=c) = p_{abc} / \sum_{j,k} p_{j,k,c} = exp(\mu_{ac}+\nu_{bc}) / \Big(\sum_{j=1}^{A} e^{\mu_{jc}} \sum_{k=1}^{B} e^{\nu_{kc}} \Big)$$

$$= \frac{e^{\mu_{ac}}}{\sum_{j=1}^{A} e^{\mu_{jc}}} \cdot \frac{e^{\nu_{bc}}}{\sum_{k=1}^{B} e^{\nu_{kc}}} = P(U=a|W=c) \cdot P(V=b|W=c)$$

For the converse, if we assume that $p_{abc}/P(W=c)$ factors as in the final expression into a function $P(U=a|W=c)$ depending only on (a,c) and another function $P(V=b|W=c)$ dependeing only on (b,c), then define $\mu_{ac} = \log(P(U=a, W=c))$, $\nu_{bc} = \log(P(V=b|W=c))$, making $p_{abc} = \exp(\mu_{ac} + \nu_{bc})$.

(b) Let N_{abc} denote the count $\sum_{i=1}^{n} I[U_i = a, V_i = b, W_i = c]$, and use subscript +'s to denote summation over unwanted indices, e.g., $N_{a+c} = \sum_{b=1}^{B} N_{abc} = \sum_{i=1}^{n} I_{[U_i=a,\, W_i=c]}$. Then we obtain the exponential family form of the data $X_1, \ldots, X_n$ by expressing the likelihood as $\prod_{i=1}^{n} p_{U_i V_i W_i} =$

$$\exp \Big(\sum_{i=1}^{n} (\mu_{U_i W_i} + \nu_{V_i W_i}) \Big) = \exp \Big(\sum_{a=1}^{A} \sum_{c=1}^{C} N_{a+c}\, \mu_{ac} + \sum_{b=1}^{B} \sum_{c=1}^{C} N_{+bc}\, \nu_{bc} \Big)$$

From this, we can see that the statistics $\{N_{a+c}, N_{+bc} \;:\; a = 1, \ldots, A,\; b = 1, \ldots, B,\; c = 1, \ldots, C\}$ are sufficient. But to reduce these to a minimal set, we remove the linear degeneracies by defining the sufficient statistics as

$$\{N_{++c},\; c = 1, \ldots, C-1\}\,, \quad \{N_{a+c},\; a = 1, \ldots, A-1,\; c = 1, \ldots, C\}$$

$$\{N_{+bc},\; b = 1, \ldots, B-1,\; c = 1, \ldots, C\}$$

These sufficient statistics have no further linear degeneracies, since they have positive probabilities of taking on any nonnegative integer-value combinations such that within the first of the three curly-brackets the sum is $\le n$, and within the second and third the respective sums over a and b are $\le N_{++c}$. The corresponding parameters are reduced by defining as at the end of part (a):

$$\gamma_c = \log P(W=c), \quad \mu_{a|c} = \log P(U=a|W=c), \quad \nu_{bc} = \log P(V=b|W=c)$$

resulting in the constraints (for all fixed c in the second and third equalities)

$$\sum_{c=1}^{C} e^{\gamma_c} = 1\,, \quad \sum_{a=1}^{A} e^{\mu_{a|c}} = 1\,, \quad \sum_{b=1}^{B} e^{\nu_{bc}} = 1 \qquad (*)$$

(c) The maximization could be directly by using Lagrange multipliers, but using the parameters defined in (b) above, with the constraints $\gamma_+ = \mu_{+|c} = \nu_{+c} \equiv 0$, the log-likelihood becomes

$$\sum_{a,c} N_{a+c} \big(\frac{1}{A} \gamma_c + \mu_{a|c}\big) + \sum_{b,c} N_{+bc}\, \nu_{bc}$$

$$= \sum_c N_{++c}\,\gamma_c/A \,+\, \sum_{a,c} N_{a+c}\,\mu_{a|c} \,+\, \sum_{b,c} N_{+bc}\,\nu_{bc}$$

Maximizing this subject to the constraints (*) is possible if and only if $N_{++c} = 0$, yielding the equations

$$N_{++c}/A + \lambda e^{\gamma_c} = 0 \quad \forall\ c \qquad \Rightarrow \qquad \lambda = -n/A,\ \hat{\gamma}_c \;=\; \log(N_{++c}/n)$$

$$N_{a+c} + \rho_c e^{\mu_{a|c}} = 0 \quad \forall\ a, c \qquad \Rightarrow \qquad \rho_c = -N_{++c},\ \hat{\mu}_{a|c} \;=\; \log(N_{a+c}/N_{++c})$$

$$N_{+bc} + \tau_c e^{\nu_{bc}} = 0 \quad \forall\ b, c \qquad \Rightarrow \qquad \tau_c = -N_{++c},\ \hat{\nu}_{bc} \;=\; \log(N_{+bc}/N_{++c})$$

Putting these estimates in for the parameters $\gamma_c,\, \mu_{a|c},\, \nu_{bc}$ immediately leads to the desired equivalent form $\hat{p}_{abc} \;=\; N_{a+c}\,N_{+bc}/(n\,N_{++c})$.

Bickel-Doksum, #2.4.11. (a) Here S_i and Δ_i are jointly distributed with

$$\Delta_i \;\sim\; \text{Binom}(1, \lambda) \;\;, \quad \text{and given } \; \Delta_i \;\;, S_i \;\sim\; \mathcal{N}(\mu_{2-\Delta_i},\, \sigma^2_{2-\Delta_i})$$

(The subscript $2 - \Delta_i$ is 1 for $\Delta_i = 1$ and is 2 for $\Delta_i = 0$. Thus the marginal density of S_i is, as desired,

$$P(\Delta)\, f_{S|\Delta}(s|1) + (1 - P(\Delta = 1))\, f_{S|\Delta}(s|0) \;=\; \lambda\, \varphi_{\sigma_1}(s - \mu_1) + (1 - \lambda)\, \varphi_{\sigma_2}(s - \mu_2)$$

As defined in this problem, the observed data are $\mathbf{X}_{obs} = (S_i,\ i = 1, \ldots, n)$, the missing data are $\mathbf{Y}_{mis} = (S_i,\ i = 1, \ldots, n)$, and the unknown parameters are $\vartheta = (\lambda,\, \mu_1,\, \mu_2,\, \sigma_1,\, \sigma_2)$.

(b) The joint complete (including missing) data log-likelihood is a constant plus

$$-\frac{1}{2} \sum_{i=1}^{n} \Big(\Delta_i \Big[-2\log\lambda \,+\, \log\sigma_1^2 \,+\, \frac{(S_i - \mu_1)^2}{\sigma_1^2} \Big]$$

$$+\, (1 - \Delta_i) \Big[-2\log(1-\lambda) \,+\, \log\sigma_2^2 \,+\, \frac{(S_i - \mu_2)^2}{\sigma_2^2} \Big] \Big) \qquad (\dagger)$$

To find the E- and M- steps, starting from an initial guess ϑ_1, we must find

$$p_1(S_i) \;=\; P_{\vartheta_1}(\Delta_i = 1 | S_i) \;=\; \frac{\lambda_1\, \varphi_{\sigma_{1,1}}(S_i - \mu_{1,1})}{\lambda_1\, \varphi_{\sigma_{1,1}}(S_i - \mu_{1,1}) \,+\, (1 - \lambda_1)\, \varphi_{\sigma_{2,1}}(S_i - \mu_{2,1})}$$

Then the E-step, conditional expectation under ϑ_0 of the complete log-likelihood ($\dagger$), gives as a function of ϑ :

$$-\frac{1}{2} \sum_{i=1}^{n} \Big(p_1(S_i) \Big[-2\,\log\lambda \,+\, \log\sigma_1^2 \,+\, \frac{(S_i - \mu_1)^2}{\sigma_1^2} \Big]$$

$$+\, (1 - p_1(S_i)) \Big[-2\,\log(1-\lambda) \,+\, \log\sigma_2^2 \,+\, \frac{(S_i - \mu_2)^2}{\sigma_2^2} \Big] \Big)$$

and the M-step is to maximize this (at next iterative guess $\vartheta = \vartheta_2$), which is easily seen to yield:

$$\hat{\lambda}_2 = \frac{1}{n} \sum_{i=1}^n p_1(S_i) \quad , \quad \hat{\mu}_{1,2} = \frac{\sum_{i=1}^n p_1(S_i)\, S_i}{\sum_{i=1}^n p_1(S_i)} \quad , \quad \hat{\mu}_{2,2} = \frac{\sum_{i=1}^n (1-p_1(S_i))\, S_i}{\sum_{i=1}^n (1-p_1(S_i))}$$

$$\hat{\sigma}^2_{1,2} = \frac{\sum_{i=1}^n p_1(S_i)\,(S_i - \hat{\mu}_{1,2})^2}{\sum_{i=1}^n p_1(S_i)} \quad , \quad \hat{\sigma}^2_{2,2} = \frac{\sum_{i=1}^n (1-p_1(S_i))\,(S_i - \hat{\mu}_{2,2})^2}{\sum_{i=1}^n (1-p_1(S_i))}$$

Bickel-Doksum, #3.4.11 (a). The probability mass function for each Y_i is

$$p(k, \alpha, \beta) = \frac{1}{k!}\, e^{(\alpha+\beta z_i)k} \exp(-e^{\alpha+\beta z_i}) \quad , \qquad k = 0, 1, \ldots$$

This is of exponetial-family form for each i, and the joint probability mass function for $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is

$$\exp(\alpha \sum_{i=1}^n y_i \,+\, \beta \sum_{i=1}^n y_i\, z_i \,-\, \sum_{i=1}^n \exp(\alpha + \beta z_i)) \prod_{i=1}^n \frac{1}{(y_i)!}$$

The sufficient statistic for the natural parameter (α, β) is $\sum_{i=1}^n y_i\,(1, z_i)'$.

(b) $I(\alpha, \beta)$ is obtained by taking the negative Hessian log-likelihood, which no longer involves the data:

$$I(\alpha, \beta) = \sum_{i=1}^n e^{\alpha+\beta z_i} \begin{pmatrix} 1 \\ z_i \end{pmatrix}^{\otimes 2}$$

(c) The limiting information per observation with $z_i = \log(i/(n+1))$ is

$$\lim_n \frac{1}{n} \sum_{i=1}^n e^{\alpha} \Big(\frac{i}{n+1}\Big)^{\beta} \begin{pmatrix} 1 \\ \log(i/(n+1)) \end{pmatrix}^{\otimes 2} = e^{\alpha} \int_0^1 x^{\beta} \begin{pmatrix} 1 \\ \log x \end{pmatrix}^{\otimes 2} dx$$

and since, by the substitution $x = \exp(-z)$, for $j = 0,\, 1,\, 2$

$$\int_0 x^{\beta}\, (\log x)^j\, dx = (-1)^j \int_0^{\infty} z^j\, e^{-(1+\beta)z}\, dz = (-1)^j\, (j!)\, (1+\beta)^{-j-1}$$

we obtain

$$\lim_n \frac{1}{n}\, I(\alpha, \beta) = \frac{1}{1+\beta} \begin{pmatrix} 1 & -1/(\beta+1) \\ -1/(\beta+1) & 2/(\beta+1)^2 \end{pmatrix}$$

The limit on n times the lower bounds of variances of the estimators are the diagonal elements of the inverse of the last-displaued matrix, which is

$$a.var = (\beta+1) \begin{pmatrix} 2 & (\beta+1) \\ (\beta+1) & (\beta+1)^2 \end{pmatrix}$$