Solutions to Homework 4
BIOS 7731

1. **BD 1.6.12 Show directly using the definition of rank of an exponential family that the multinomial distribution** $M(n;\theta_1,\ldots,\theta_k), 0 < \theta_j < 1, 1 \le j \le k$**,** $\sum_{j=1}^{k}\theta_j = 1$**, is of rank** $k-1$**.**

   There are different ways to prove this, first by reparameterizing the model and then showing that one of the conditions of Thm 1.6.4 hold on the resulting $T(x)$. Here, we use the direct definition.

   For $T_j(x) = N_j$, where N_j are the counts of category j, suppose that this model is not of rank $k-1$, then $P(\sum_{j=1}^{k-1} a_j N_j = a_k) = 1$, where at least one $a_j \neq 0$. WLOG, assume $a_1 = 1$, then given $N_2 = n_2, \ldots N_{k-1} = n_{k-1}$

$$P(N_1 = a_k - \sum_{j=2}^{k-1} a_j n_j | N_2 = n_2, \ldots N_{k-1} = n_{k-1}) = 1,$$

   which implies that N_1 is a constant. However, we know that conditionally on $N_j = n_j$, $2 \le j \le k-1$,

$$N_1 \sim Binomial(n - \sum_{j=2}^{k-1} n_j, \theta_1).$$

   So N_1 cannot be a constant unless $\sum_{j=2}^{k-1} n_j = n$ with probability one. This is not the case since this sum is a random variable with,

$$\sum_{j=2}^{k-1} N_j \sim Binomial(n, \sum_{j=2}^{k-1} \theta_j).$$

2. **BD 1.6.18 Suppose $Y_1, \ldots, Y_n$ are independent with $Y_i \sim N(\beta_1 + \beta_2 z_i, \sigma^2)$, where $z_1, \ldots, z_n$ are covariate values not all equal. Assume $\sigma^2 = 1$ and show that this family has rank 2. Give the mean vector and the variance matrix of T.**

   (a) **Show that this family has rank 2.**

$$\begin{aligned} p(y,\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(y_i-\beta_1-\beta_2 z_i)^2}{2}\} \\ &= (\frac{1}{\sqrt{2\pi}})^n \exp\{-\frac{\sum_{i=1}^n y_i^2}{2}\} \exp\{\beta_1 \sum_{i=1}^n y_i + \beta_2 \sum_{i=1}^n y_i z_i - \frac{1}{2}\sum_{i=1}^n (\beta_1+\beta_2 z_i)^2\} \end{aligned}$$

This is an exponential family generated by $h(y) = (\frac{1}{\sqrt{2\pi}})^n \exp\{-\frac{\sum_{i=1}^n y_i^2}{2}\}$ and $T(y) = (\sum_{i=1}^n y_i, \sum_{i=1}^n y_i z_i)$, where $\eta = (\beta_1, \beta_2)$ and $A(\eta) = \frac{1}{2}\sum_{i=1}^n (\eta_1 + \eta_2 z_i)^2$. Assuming that $-\infty < \beta_1 < \infty$ and $-\infty < \beta_2 < \infty$, then $\mathcal{E}$ is open, $\eta \in R^2$.

$T(Y)$ is two dimensional but we need to show the exponential family is also rank 2. There are many different approaches for solving this problem based on Thm 1.6.4. This solution is based on showing that η is identifiable:

Suppose that this exponential family is not of rank 2, then β_1 and β_2 are not identifiable. So for $\theta_1 \neq \theta_2$, where $\theta_1 = (\beta_1, \beta_2)$ and $\theta_2 = (\beta_1^{'}, \beta_2^{'})$, then

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(y_i-\beta_1-\beta_2 z_i)}{2}\} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(y_i-\beta_1^{'}-\beta_2^{'} z_i)}{2}\}.$$

This implies that for all i,

$$\beta_1 + \beta_2 z_i = \beta_1^{'} + \beta_2^{'} z_i.$$

Therefore for all z_i,

$$z_i = \frac{\beta_1 - \beta_1^{'}}{\beta_2 - \beta_2^{'}}.$$

But this contradicts that condition that the z_i's are not all equal, thus β_1 and β_2 are identifiable and this family is of rank 2.

You can also prove this by showing that the $T_1(Y)$ and $T_2(Y)$ are linearly independent since the z_is are all not equal or by showing that $Var[T(Y)]$ is positive definite.

(b) To find the mean vector and variance matrix of T, note that $A(\eta) = \frac{1}{2}\sum_{i=1}^n (\eta_1 + \eta_2 z_i)^2$.
Then,

$$E[T(Y)] = A'(\eta) = (\sum_{i=1}^n (\eta_1 + \eta_2 z_i), \sum_{i=1}^n z_i(\eta_1 + \eta_2 z_i))$$

$$= (\sum_{i=1}^n (\beta_1 + \beta_2 z_i), \sum_{i=1}^n z_i(\beta_1 + \beta_2 z_i)),$$

and

$$Var[T(Y)] = \begin{bmatrix} n & \sum_{i=1}^n z_i \\ \sum_{i=1}^n z_i & \sum_{i=1}^n z_i^2 \end{bmatrix}.$$

3. **Suppose we have the following discrete distribution. Given one observation, X, obtain the maximum likelihood estimator $\hat{\theta}$ for θ and show that it is not unique.**

x	-2	-1	0	1	2	
$P(X = x\|\theta)$	$\frac{1-\theta}{4}$	$\frac{\theta}{12}$	$\frac{1}{2}$	$\frac{3-\theta}{12}$	$\frac{\theta}{4}$	$0 \leq \theta \leq 1$

Since there is only one observation, we can find the MLE $\hat{\theta}$ for each possible $X = x$:

x	$\max_\theta(L_x(\theta))$	$\arg\max_\theta(L_x(\theta))$
-2	$\frac{1}{4}$	0
-1	$\frac{1}{12}$	1
0	$\frac{1}{2}$	[0,1]
1	$\frac{1}{4}$	0
2	$\frac{1}{4}$	1

At $x \neq 0$, a unique maximum is achieved listed above. However, for $x = 0$ the maximum is achieved for multiple values of $\theta = [0, 1]$. Therefore, the MLE is not unique for $x = 0$.

4. **BD 1.6.27 (assumption in Thm 2.3.1 Proof). Let $\mathcal{P}$ denote the exponential family generated by T and h. For any $\eta_0 \in \mathcal{E}$, set $h_0(x) = q(x, \eta_0)$ where q is given by (1.6.9). Show that $\mathcal{P}$ is also the canonical family generated by $T(x)$ and h_0.**

   In canonical family form, for $\eta \in \mathcal{E}$,

$$(1) \quad \log q(x, \eta) = \eta T(x) - A(\eta) + \log h(x),$$

   where $A(\eta) = \log \int h(x) \exp\{\eta T(x)\} dx$.

   The density q' for the family generated by $h_0 = q(x, \eta_0) = h(x) \exp\{\eta_0 T(x) - A(\eta_0)\}$, and T is

$$(2) \quad \log q' = \eta T(x) - A^*(\eta) + \log h_0(x),$$

   where

$$A^*(\eta) = \log \int h_0(x) \exp\{\eta T(x)\} dx.$$

   Substituting h_0, we have

$$(3) \quad \log q' = (\eta + \eta_0) T(x) - (A^*(\eta) + A(\eta_0)) + \log h(x),$$

   Equation (3) is the expression for $q(x, \eta + \eta_0)$ if

$$A(\eta + \eta_0) = A^*(\eta) + A(\eta_0).$$

   Relating equations (1) and (3), to show that the exponential family generated by $T(X)$ and $h(X)$, is the same as the exponential family generated by $T(X)$ and $h_0(X)$, we need to show that this last statement is true.

   Solving for A^*,

$$A^*(\eta) = \log \int h(x) \exp\{(\eta + \eta_0) T(x) - A(\eta_0)\} dx.$$

$$A^*(\eta) = -A(\eta_0) + \log \int h(x) \exp\{(\eta + \eta_0) T(x)\} dx$$

$$A^*(\eta) = -A(\eta_0) + A(\eta + \eta_0)$$

   showing that $A(\eta + \eta_0) = A^*(\eta) + A(\eta_0)$.

   Therefore, the family generated by T and h is the same family generated by T and h_0.

5. **BD 2.3.1 Suppose $Y_1, \ldots, Y_n$ are independent**

$$P[Y_i = 1] = p(x_i, \alpha, \beta) = 1 - P[Y_i = 0], \;\; 1 \le i \le n, n \ge 2,$$

$$\log \frac{p}{1-p}(x, \alpha, \beta) = \alpha + \beta x, \;\; x_1 < \ldots < x_n$$

Show that the MLE of α, β exists iff $(Y_1, \ldots, Y_n)$ is not a sequence of 1's followed by all 0's or the reverse.

The likelihood is

$$L_y(\theta) = \prod_{i=1}^{n} [p(x_i, \alpha, \beta]^{Y_i} [1 - p(x_i, \alpha, \beta]^{1-Y_i} = \exp[\eta^t T(Y) - A(\eta)],$$

where $\eta = (\alpha, \beta)$, $T(Y) = (T_1, T_2) = (\sum_{i=1}^n Y_i, \sum_{i=1}^n Y_i x_i)$ and $A(\eta) = \sum_{i=1}^n \log(1 - p(x_i, \eta_1, \eta_2))$. Assuming that $-\infty < \alpha < \infty$ and $-\infty < \beta < \infty$, then $\mathcal{E}$ is open, $\eta \in R^2$.

This is a canonical exponential family with rank 2. T_1 and T_2 are linearly independent since the x_i's all are not equal (otherwise $T_2 = zT_1$).

Assuming that the parameter space is open, then by Thm 2.3.1, the MLE of α and β exists if $P[c^t T(Y) > c^t T(y)] > 0 \;\forall c \neq 0$, where y is the observed data.

For observed y,

$$c^t t(y) = c_1 \sum_{i=1}^n y_i + c_2 \sum_{i=1}^n y_i x_i = \sum_{i=1}^n c_1 y_i + c_2 y_i x_i$$

$$\le \sum_{i=1}^n c_1 y_i + c_2 y_i x_i 1(c_1 y_i + c_2 y_i x_i > 0)$$

To maximize this bound take $y_i = 1 \;\forall i \ge k$ and $y_i = 0 \;\forall i < k$, where k is the smallest index s.t. $(c_1 + c_2 x_i > 0)$. The x_i's are ordered so this is possible and this observed y' corresponds to a sequence of 0's followed by 1's.

For this y', the bound above is equal to $\sum_{i=k}^n c_1 + c_2 x_i$, which is the maximum possible value of $c^t T(Y)$. Therefore, there exists $c \neq 0$ (i.e., $c_2 > 0, c_1$), where $-\frac{c_1}{c_2} \le x_i \;\forall i \ge k$, s.t. $P[c^t T(Y) > c^t T(y)] = 0$ and the MLE does not exist by Thm. 2.3.1

A similar arguement follows for the reverse case where $y_i = 1 \;\forall i < k$ and $P[c^t T(Y) > c^t T(y)] = 1$, since the minimum value of $c^t T(Y)$ is attained.

6. **BD 2.3.8b Show that if $a = 1$ and $p = 1$ for the density below, the MLE $\hat{\theta}$ exists but is not unique if n is even.**

   Let $X_1, \ldots, X_n \in R^p$ be i.i.d. with density

$$f_X(\theta) = c(\alpha) \exp\{-|x - \theta|^\alpha\}, \theta \in R^p, \alpha \geq 1,$$

   where $c^{-1}(\alpha) = \int_{R^p} \exp\{-|x|^\alpha\} dx$ and $|.|$ is the Euclidean norm.

   Assume that $n = 2m$ is even and without loss of generality order the data $x_1 \leq \ldots \leq x_n$. The log likelihood function is given by

$$l_x(\theta) = n \log c(\alpha) - \sum_{i=1}^{n} |x_i - \theta|.$$

   Here $\partial\Theta = \{-\infty, \infty\}$ and this is a continuous function, so

$$\lim_{\theta \to \infty} l_x(\theta) = -\infty$$

   and

$$\lim_{\theta \to -\infty} l_x(\theta) = -\infty$$

   Thus, by Lemma 2.3.1, a maximum of $l_x(\theta)$ (the MLE) exists.

   $l_x(\theta)$ is a continuous function whose derivative does not exist at the points $\theta \in \{x_i\}$ and elsewhere satisfies

$$\frac{d}{d\theta} l_x(\theta) = -\sum_{i=1}^{n} \frac{d}{d\theta} |x_i - \theta|$$

$$= \sum_{x_i < \theta} (-1) + \sum_{x_i > \theta} (+1).$$

   The derivative equals the number of $x_i > \theta$ minus the number of $x_i < \theta$. For the derivative to be zero, these numbers need to equal each other. For the n even case, half the x_i's are indexed from $i = 1, \ldots, m$ and the other half of the x_i's are indexed from $i = m + 1, \ldots n$. Thus, in the interval $x_m < \theta < x_{m+1}$ the derivative is 0. Therefore, the MLE exists, but it is not unique since multiple values are possible (x_m, x_{m+1}).