Homework 2: Exponential of Families

Solution Set

Problem 1: Verify that displayed equation 7 in the exponential family notes holds for the binomial distribution, the Poisson distribution, and the normal distribution with both μ and σ^2 unknown.

Binomial:

For Binomial, $\theta = \log(\frac{p}{1-p})$ and $c(\theta) = n\log(1+e^\theta)$.

$$\nabla c(\theta) = n\frac{e^\theta}{1+e^\theta} = n\frac{1}{1+\frac{p}{1-p}} \cdot \frac{p}{1-p} = np = \mathrm{E}(Y)$$

$$\nabla^2 c(\theta) = n\frac{e^\theta}{(1+e^\theta)^2} = n(1-p)^2\frac{p}{1-p} = np(1-p) = \mathrm{Var}(Y)$$

Poisson:

For Poisson distribution, $\theta = \log(\lambda)$ and $c(\theta) = e^\theta$

$$\nabla c(\theta) = e^\theta = e^{\log(\lambda)} = \lambda = \mathrm{E}(Y)$$

$$\nabla^2 c(\theta) = e^\theta = e^{\log(\lambda)} = \lambda = \mathrm{Var}(Y)$$

Normal:

For normal distribution, $\theta = (\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2})^T$ and $c(\theta) = \frac{1}{2}\left(\frac{\theta_1^2}{2\theta_2} - \log(2\theta_2)\right)$ and $Y = (w, -w^2)^T$

$$\frac{\partial c(\theta)}{\partial \theta_1} = \frac{\theta_1}{2\theta_2} = \mu \qquad \frac{\partial c(\theta)}{\partial \theta_2} = -\frac{\theta_1^2}{4} - \frac{1}{\theta_2^2} - \frac{1}{2\theta_2} = -\mu^2 - \sigma^2$$

so $\nabla c(\theta) = (\mu, -\mu^2 - \sigma^2) = \mathrm{E}(Y)$.

$$\frac{\partial^2 c(\theta)}{\partial \theta_1^2} = \frac{1}{2\theta_2} = \sigma^2 = \mathrm{Var}(Y_1) \qquad \frac{\partial^2 c(\theta)}{\partial \theta_2^2} = \frac{\theta_1^2\theta_2}{2\theta_2^4} + \frac{1}{2\theta_2^2} = 4\mu^2\sigma^2 + 2\sigma^4 = \mathrm{Var}(Y_2)$$

$$\frac{\partial^2 c(\theta)}{\partial \theta_1 \partial \theta_2} = \frac{\partial^2 c(\theta)}{\partial \theta_2 \partial \theta_1} = -\frac{\theta_1}{2\theta_2^2} = -2\mu\sigma^2 = \mathrm{Cov}(Y_1, Y_2)$$

so $\nabla^2 c(\theta) = \mathrm{Var}(Y)$

Problem 2: Show that the second derivative of the map h (displayed equation 11 in the exponential family notes) is equal to $-\nabla^2 c(\theta)$ and justify that this matrix is negative definite when the exponential family model is identifiable.

$$\frac{\partial h(\theta)}{\partial \theta_i} = \mu_i - \frac{\partial c(\theta)}{\partial \theta_i} \qquad \frac{\partial^2 h(\theta)}{\partial \theta_i^2} = -\frac{\partial^2 c(\theta)}{\partial \theta_i^2} \qquad \frac{\partial^2 h(\theta)}{\partial \theta_i \partial \theta_j} = -\frac{\partial^2 c(\theta)}{\partial \theta_i \partial \theta_j}$$

So $h''(\theta) = -\nabla^2 c(\theta) = -\mathrm{Var}(Y)$

$$u^T \mathrm{Var}(Y) u = u^T \mathrm{E}[(Y - \mathrm{E}(Y))(Y - \mathrm{E}(Y))^T]u = \mathrm{E}[(u^T Y - u^T \mathrm{E}(Y))^2]$$

This expectation will be 0 only when there exists u such that $u^T Y = u^T \mathrm{E}(Y) = c$. For identifiable exponential family, such u does not exist, so $u^T \mathrm{Var}(Y) u > 0 \ \forall u \in \mathbb{R}^d$, therefore $h''(\theta) = -\mathrm{Var}(Y)$ is negative definite.

Problem 3: The above problem is one of the steps needed to finish the proof of Theorem 2 in the exponential family notes. Finish the proof of Theorem 2.

(i) Since moment generating functions are infinitely differentiable at zero. By definition, the cumulant generating function is the log of mgf, because log is infinitely differentiable so are cumulant generating functions. Hence the change-of-parameter $g(\theta) = \nabla c(\theta)$ is infinitely differentiable and therefore continuously differentiable.

(ii) It is suffices to show that matrix inversion is infinitely differentiable. The formula for the derivative of a matrix inverse can be shown by differentiating $AA^{-1} = I$

$$I' = (AA^{-1})' = A'A^{-1} + A(A^{-1})' = 0 \ \Rightarrow A(A^{-1})' = -A'A^{-1} \ \Rightarrow (A^{-1})' = -A^{-1}A'A^{-1}$$

Thus, if A is infinitely differentiable, then so is A^{-1}. This proves that g^{-1} is infinitely differentiable.

Problem 4: Let Y be a regular full exponential family with canonical parameter θ. Verify that Y is sub-exponential.

Since Y is a regular full exponential familiy, the moment generating function exists

$$\mathrm{E}[e^{\phi(Y-\mu)}] = \mathrm{E}[e^{\phi Y - \phi\mu}] = \mathrm{E}[e^{\phi Y}]e^{-\phi\mu} = e^{c(\theta+\phi) - c(\theta) - \phi\mu}$$

We want to show

$$c(\theta + \phi) - c(\theta) - \phi\mu \leq \frac{\lambda^2 \phi^2}{2}, |\phi| < 1/b$$

The Taylor expansion of $c(\theta + \phi)$ at θ is

$$c(\theta + \phi) = c(\theta) + c'(\theta)\phi + c''(\theta)\phi^2/2 + c^{(3)}(\theta)\phi^3/6 + o(\phi^3)$$

therefore,

$$c(\theta + \phi) - c(\theta) - \phi\mu = \mathrm{Var}(Y)\phi^2/2 + c^{(3)}(\theta)\phi^3/6 + o(\phi^3)$$

Take $\lambda^2 = 2\mathrm{Var}(Y)$ and $b \geq \frac{c^{(3)}(\theta)}{3\mathrm{Var}(Y)}$, when ϕ is small enough, we have

$$\mathrm{Var}(Y)\phi^2/2 + c^{(3)}(\theta)\phi^3/6 + o(\phi^3) \leq \frac{\lambda^2\phi^2}{2}$$

Problem 5: In the notes it was claimed that the scalar products of $\sum_{i=1}^n \{y_i - \nabla c(\theta)\}$ are also sub-exponential (page 15). Show that this is in fact true when the observations y_i are iid from a regular full exponential family.

Consider $\sum_{i=1}^{n} a_i\{y_i - \nabla c(\theta)\}$. Let $z_i = a_i y_i$. We know that y_i is generated from an exponential family and thus $f_{Y_i}(y) = h(y)exp\{\langle y, \theta\rangle - c(\theta)\}$

We can transform: $f_{Z_i}(z) = \frac{h(z)}{a_i} exp\{\langle z/a_i, \theta\rangle - c(\theta)\}$ which shows that Z_i are also independent random variables from an exponential family with mean paramter $a_i\mu$, and thus Z_i are also sub-exponential (problem 4)

$$\implies \mathrm{E}(\exp\{\phi_i(z_i - a_i\mu)\}) \le exp\left\{\frac{\lambda_i^2\phi_i^2}{2}\right\} \text{ for some } \lambda_i \text{ and } b_i \text{ for all } |\phi_i| < 1/b_i$$

Since n is finite we can consider a b large enough such that $|\phi| < \frac{1}{b} \le \frac{1}{b_i}\ \forall\ i$

$$\begin{aligned}
\mathrm{E}(\exp\{\phi(\sum_{i=1}^{n} z_i - a_i\mu - 0)\}) &= \prod_{i=1}^{n} \mathrm{E}(\exp\{\phi(z_i - a_i\mu)\}) \text{ since independent} \\
&\le exp\left\{\frac{\phi^2\sum\lambda_i^2}{2}\right\} \\
&\text{Choosing a } \lambda \text{ s.t. } \lambda^2 = \sum_i \lambda_i^2 \\
&= exp\left\{\frac{\phi^2\lambda^2}{2}\right\}
\end{aligned}$$

Which completes the proof.

Problem 6: Derive the MLEs of the canonical parameters of the binomial distribution, the Poisson distribution, and the normal distribution with both μ and σ^2 unknown.

Binomial:

For Binomial distribution

$$g(\theta) = \nabla c(\theta) = n\frac{e^\theta}{1+e^\theta} \Rightarrow g^{-1}(\theta) = \log\left(\frac{\theta}{n-\theta}\right)$$

so $\hat{\theta} = g^{-1}(y) = \log\left(\frac{y}{n-y}\right)$

Poisson:

For Binomial distribution

$$g(\theta) = \nabla c(\theta) = e^\theta \Rightarrow g^{-1}(\theta) = \log(\theta)$$

so $\hat{\theta} = g^{-1}(y) = \log(y)$

Normal:

For Normal distribution

$$l(\theta) = y_1\theta_1 + y_2\theta_2 - \frac{1}{2}[\frac{\theta_1^2}{2\theta_2} - \log(2\theta_2)]$$

$$\frac{\partial l(\theta)}{\partial \theta_1} = y_1 - \frac{\theta_1}{2\theta_2} = 0 \qquad \frac{\partial l(\theta)}{\partial \theta_2} = y_2 + \frac{1}{2\theta_2} = 0$$

so we have $\hat{\theta}_1 = -\frac{y_1}{y_2}$ and $\hat{\theta}_2 = -\frac{1}{2y_2}$

Problem 7: Derive the asymptotic distribution for the MLE of the submodel mean value parameter vector $\hat{\tau}$.

We have,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I(\theta)^{-1})$$

Further since: $\tau = M^T \mu = M^T \nabla c(\theta)$, we have using multivariate delta method:

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, M^T \nabla^2 c(\theta) I(\theta)^{-1} \nabla^2 c(\theta) M^T)$$

And since, $I(\theta) = \nabla^2 c(\theta)$, we have

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, I(\beta))$$

Where, $I(\beta) = M^T I(\theta) M$

Problem 8: Prove Lemma 1 in the exponential family notes.

Since the log-likelihood of exponential family takes the form

$$l(\theta) = \langle y, \theta \rangle - c(\theta)$$

We can see that $c(\theta)$ only depends on the canonical parameter θ and $\langle y, \theta \rangle$ is a function of θ and statistic, by factorization theorem the canonical statistic vector is sufficient statistic.