Homework 2: problems about exponential families

TA : Arjama Das

Due: February 14th at 11:59 PM

This homework set will cover problems concerning exponential family theory. All derivations must be typed. Screenshots of work done with pen and paper will not be accepted.

Problem 1 [10 points]: 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.

Solution: (a) $Y \sim \mathrm{Bin}(n, p)$

$$\begin{aligned} \implies f_Y(y) &= \binom{n}{y} p^y (1-p)^{n-y} \\ &= \binom{n}{y} e^{y \log p/(1-p) + n \log(1-p)} \end{aligned}$$

Thus,

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

We know that,

$$E_\theta(Y) = np = n\frac{e^\theta}{1+e^\theta}; \quad \mathrm{Var}_\theta(Y) = np(1-p) = n\frac{e^\theta}{\left(1+e^\theta\right)^2}$$

Now,

$$\begin{aligned} c'(\theta) &= n\frac{1}{1+e^\theta}e^\theta = np = E_\theta(Y) \\ c''(\theta) &= -n\frac{e^{2\theta}}{(1+e^\theta)^2} + n\frac{e^\theta}{1+e^\theta} \\ &= n\frac{e^\theta}{\left(1+e^\theta\right)^2} \\ &= np(1-p) \\ &= \mathrm{Var}_\theta(Y) \end{aligned}$$

Thus equation (7) holds.

(b) $Y \sim \mathrm{Poi}(\lambda)$

$$f_Y(y) = \frac{e^{-\lambda}\lambda^y}{y!} = e^{-\lambda + y\log\lambda - \log(x!)}$$

Thus,

$$\theta = \log(\lambda); \quad c(\theta) = e^{\theta}$$

We know that $E_\theta(Y) = \lambda = \mathrm{Var}_\theta(Y)$,

$$c'(\theta) = e^{\theta} = \lambda = E_\theta(Y); \quad c''(\theta) = e^{\theta} = \lambda = \mathrm{Var}_\theta(Y)$$

Thus equation (7) holds.

(c) $Y \sim N\left(\mu, \sigma^2\right)$

$$\begin{aligned} f_{\mu,\sigma^2}(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y^2 - 2y\mu + mu^2\right)}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \log\left(\sigma^2\right)\right)\right) \end{aligned}$$

Thus,

$$Y(y) = \left(y, -y^2\right)^T; \quad \theta = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right); \quad c(\theta) = \frac{1}{2}\left(\frac{\theta_1^2}{2\theta_2} - \log\left(2\theta_2\right)\right)$$

We know that $\mathrm{E}(Y) = \mu, \mathrm{E}\left(Y^2\right) = \sigma^2 + \mu^2, \mathrm{Var}(Y) = \sigma^2$ Let us now compute $\mathrm{Var}\left(Y^2\right)$. The mgf of normal $N\left(\mu, \sigma^2\right)$ is $\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$. Thus,

$$\begin{aligned} M_X'(t) &= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right) \\ \Longrightarrow M_X''(t) &= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)^2 + \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\sigma^2 \\ \Longrightarrow M_X^{(3)}(t) &= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)^3 + 3\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)\sigma^2 \\ \Longrightarrow M^{(4)}(t) &= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)^4 + 6\exp\left(\mu t + 3\frac{\sigma^2 t^2}{2}\right)\left(\mu + \sigma^2 t\right)^2\sigma^2 + 3\exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\sigma^4 \end{aligned}$$

Thus,

$$\begin{aligned} \mathrm{E}\left(Y^3\right) = \left. M_X^{(3)}(t)\right|_{t=0} &= \mu^3 + 3\mu\sigma^2 \\ \mathrm{E}\left(Y^4\right) &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \\ \Longrightarrow \mathrm{Var}\left(Y^2\right) &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 - \left(\sigma^2 + \mu^2\right)^2 \\ &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 - \sigma^4 - \mu^4 - 2\sigma^2\mu^2 \\ &= 4\mu^2\sigma^2 + 2\sigma^4 \end{aligned}$$

and,

$$\mathrm{Cov}\left(Y, Y^2\right) = \mathrm{E}\left(Y^3\right) - \mathrm{E}\left(Y^2\right)\mathrm{E}(Y) = \mu^3 + 3\mu\sigma^2 - \mu\left(\sigma^2 + \mu^2\right) = 2\mu\sigma^2$$

Now,

$$c'(\theta) = \left(\frac{\theta_1}{2\theta_2}, -\left(\frac{\theta_1^2}{4\theta_2^2} + \frac{1}{2\theta_2}\right)\right) = \left(\mu, -\mu^2 - \sigma^2\right) = \left(\mathrm{E}(Y), \mathrm{E}\left(-Y^2\right)\right) = \mathrm{E}(Y(y))$$

$$c''(\theta) = \begin{bmatrix} \frac{1}{2\theta_2} & -\frac{\theta_1}{2\theta_2^2} \\ -\frac{\theta_1}{2\theta_2^2} & \frac{\theta_1^2}{2\theta^3} + \frac{1}{2\theta^2} \end{bmatrix} = \begin{bmatrix} \sigma^2 & -2\mu\sigma^2 \\ -2\mu\sigma^2 & 4\mu^2\sigma^2 + 2\sigma^4 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(Y) & \mathrm{Cov}\left(Y, -Y^2\right) \\ \mathrm{Cov}\left(Y, -Y^2\right) & \mathrm{Var}\left(Y^2\right) \end{bmatrix} = \mathrm{Var}(Y(y))$$

Thus equation (7) holds.

Problem 2 [20 points]: This problem concerns the proof of Theorem 3 in the exponential family notes. Do the following:

- **part a** [10 points]: Show that the second derivative of the map h is equal to $-\nabla^2 c(\theta)$ and justify that this matrix is negative definite when the exponential family model is identifiable.

Solution: We know that,

$$\begin{aligned} h(\theta) &= \langle \mu, \theta \rangle - c(\theta) \\ \implies h'(\theta) &= \mu - \nabla c(\theta) \\ h''(\theta) &= -\nabla^2 c(\theta) = -\operatorname{Var}_\theta(Y) \end{aligned}$$

We know that $\operatorname{Var}_\theta(Y)$ is semi positive definite. Consider a vector $x \neq 0$,

$$\begin{aligned} x^T \operatorname{Var}_\theta(Y) x &= x^T \left[\operatorname{E}_\theta \left(Y Y^T \right) - \operatorname{E}_\theta(Y) \operatorname{E}_\theta(Y)^T \right] x \\ &= \operatorname{E}_\theta \left(x^T Y Y^T x \right) - \operatorname{E}_\theta \left(x^T Y \right) \operatorname{E}_\theta \left(x^T Y \right)^T \\ &= \operatorname{Var}_\theta \left(x^T Y \right) \end{aligned}$$

$$\therefore x^T \operatorname{Var}_\theta(Y) x = 0 \implies x^T Y = c \text{ for any vector } x \neq 0 \text{ and some const c}$$

But since the model is identifiable $x^T \operatorname{Var}_\theta(Y) x \neq 0$ for any non zero vector x . Thus $\operatorname{Var}_\theta(Y)$ is positive definite. Which implies that $-\nabla^2 c(\theta)$ is negative definite.

- **part b** [10 points]: Finish the proof of Theorem 3.

Solution: The two steps that need to proof to complete the proof of the theorem are the following,

(a) The cumulant functions are infinitely differentiable and are therefore continuously differentiable.

Proof: We know that the moment generating function is infinitely differentiable. Since log is also an infinitely differentiable function and the cumulant generating function is the log of the moment generating function, the cumulant generating function is also infinitely differentiable. Consider $k_\theta(t)$ to be the cumulant generating function corresponding to the canonical statistic θ and $c(\theta)$ be the cumulant function, then we know that

$$k_\theta(t) = c(t + \theta) - c(\theta)$$

The derivatives of $k_\theta(t)$ evaluated at 0 are the same as the cumulant function c evaluated at θ. Thus since the cumulant generating functions are infinitely differentiable, the cumulant functions are also infinitely differentiable and this further implies that the cumulant functions are also continuously differentiable.

(b) $g^{-1}(\theta)$ is infinitely differentiable

Proof: From the inverse function theorem we know that

$$\nabla g^{-1}(\theta) = [\nabla g(\theta)]^{-1}$$

Now note that $\nabla g(\theta) = \nabla^2 c(\theta)$ and as shown above $\nabla^2 c(\theta)$ is positive definite. Also $\frac{1}{x^2}$ is infinitely differentiable when $x \neq 0$. Thus $\left[\nabla^2 c(\theta)\right]^{-1}$ is infinitely differentiable since $c(\theta)$ is infinitely differentiable (which would imply that $\nabla^2 c(\theta)$ is also infinitely differentiable). Combining the results: $\nabla g^{-1}(\theta) = [\nabla g(\theta)]^{-1}$ exists and is infinitely differentiable we get that $g^{-1}(\theta)$ is infinitely differentiable.

Note that part a will be referenced later in this course. Hence, it is treated as its own sub-problem.

Problem 3 [20 points]: Let $y \in \mathbb{R}$ be a regular full exponential family with canonical parameter $\theta \in \mathbb{R}$. Do the following:

- **part a** [15 points]: Verify that y is sub-exponential.

Solution: Need to show that Y is sub-exponential i.e. $\mathrm{E}\left(e^{\phi(Y-\mu)}\right) \leq e^{\lambda^2\phi^2/2} \forall |\phi| < 1/b$ Using taylor expansion of e^x we get,

$$\begin{aligned}
\mathrm{E}(\exp\{\phi(Y-\mu)\}) &\approx E\left(1 + \phi(Y-\mu) + \frac{\phi^2(Y-\mu)^2}{2} + O\left(\phi^2\right)\right) \\
&= 1 + \frac{\phi^2 \operatorname{Var}_\theta(Y)}{2} + O\left(\phi^2\right)
\end{aligned}$$

Similarly,

$$e^{\lambda^2\phi^2/2} \approx 1 + \frac{\lambda^2\phi^2}{2} + O\left(\phi^2\right)$$

Choosing a λ such that $\operatorname{Var}_\theta(Y) < \lambda^2$ and b large enough such that $O\left(\phi^2\right)$ is negligible, for all $|\phi| < 1/b$ we get,

$$\mathrm{E}(\exp\{\phi(Y-\mu)\}) \leq e^{\lambda^2\phi^2/2}$$

Thus, Y is sub-exponential.

- **part b** [5 points]: In the notes it was claimed that the scalar products of $\sum_{i=1}^n \{y_i - \nabla c(\theta)\}$ are sub-exponential (see the "Finite sample concentration of MLE" section in the exponential family notes). Verify that this is true when y_i are iid realizations from a regular full exponential family.

Solution: 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 $f_{Y_i}(y) = h(y)\exp\{\langle y, \theta\rangle - c(\theta)\}$ Thus $f_{Z_i}(z) = \frac{h(z)}{a_i}\exp\{\langle z/a_i, \theta\rangle - c(\theta)\}$ which shows that Z_i are also independent random variables from an exponential family with mean paramter $a_i\mu$ We have showed in the above problem that if Z is generated from a regular full exponential family then Z is sub-exponential.

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

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

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

Thus, $\sum_{i=1}^{n} a_i \left\{y_i - \nabla c(\theta)\right\}$ is also sub-exponential.

Problem 4 [10 points]: Derive the MLEs of the canonical parameters of the binomial distribution, and the normal distribution with both μ and σ^2 unknown.

Solution: (a)A binomial random variable $Y \sim \mathrm{Bin}(n, p)$ has the probability mass function:

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, \ldots, n.$$

Given m independent observations $y_1, y_2, \ldots, y_m$ from $\mathrm{Bin}(n, p)$, the likelihood function is:

$$L(p) = \prod_{i=1}^{m} \binom{n}{y_i} p^{y_i} (1-p)^{n-y_i}.$$

Taking the log-likelihood:

$$\ell(p) = \sum_{i=1}^{m} \log\binom{n}{y_i} + \sum_{i=1}^{m} y_i \log p + \sum_{i=1}^{m} (n - y_i)\log(1-p).$$

Ignoring constants:

$$\ell(p) = \sum_{i=1}^{m} y_i \log p + \sum_{i=1}^{m} (n - y_i)\log(1-p).$$

Taking the derivative:

$$\frac{d}{dp}\ell(p) = \sum_{i=1}^{m} \frac{y_i}{p} - \sum_{i=1}^{m} \frac{n - y_i}{1-p}.$$

Setting this to zero:

$$\sum_{i=1}^{m} \frac{y_i}{p} = \sum_{i=1}^{m} \frac{n - y_i}{1-p}.$$

Solving for p:

$$\hat{p} = \frac{\sum_{i=1}^{m} y_i}{mn} = \frac{\bar{y}}{n}.$$

Rewriting in exponential family form:

$$P(Y = y) = \exp\left(y \log \frac{p}{1-p} + n \log(1-p) + \log \binom{n}{y} \right).$$

where the canonical parameter is:

$$\theta = \log \frac{p}{1-p}.$$

Thus, the MLE for θ is:

$$\hat{\theta} = \log \frac{\hat{p}}{1-\hat{p}} = \log \frac{\bar{y}}{n - \bar{y}}.$$

(b)A normal random variable $Y \sim \mathcal{N}(\mu, \sigma^2)$ has the probability density function:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2} \right).$$

Given m independent observations $y_1, y_2, \ldots, y_m$, the likelihood function is:

$$L(\mu, \sigma^2) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2} \right).$$

Taking the log-likelihood:

$$\ell(\mu, \sigma^2) = -\frac{m}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - \mu)^2.$$

Differentiating with respect to μ:

$$\frac{d}{d\mu} \ell(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^{m} (y_i - \mu).$$

Setting to zero:

$$\sum_{i=1}^{m} (y_i - \hat{\mu}) = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} y_i = \bar{y}.$$

Differentiating with respect to σ^2:

$$\frac{d}{d\sigma^2}\ell(\mu,\sigma^2) = -\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{m}(y_i-\mu)^2.$$

Setting to zero:

$$\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(y_i-\bar{y})^2.$$

Rewriting the normal density function in exponential family form:

$$f(y) = \exp\left(\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right).$$

From this, the canonical parameters are:

$$\theta_1 = \frac{\mu}{\sigma^2}, \quad \theta_2 = -\frac{1}{2\sigma^2}.$$

Thus, the MLEs for the canonical parameters are:

$$\hat{\theta}_1 = \frac{\hat{\mu}}{\hat{\sigma}^2} = \frac{\bar{y}}{\hat{\sigma}^2}, \quad \hat{\theta}_2 = -\frac{1}{2\hat{\sigma}^2} = -\frac{1}{2}\left(\frac{1}{m}\sum_{i=1}^{m}(y_i-\bar{y})^2\right)^{-1}.$$

Problem 5 [10 points]: Derive the asymptotic distribution for $\hat{\tau}$, the MLE of the submodel mean value parameter vector. Hint: use the Delta method.

Solution: We know that the canonical linear submodel also has an exponential family with - canonical statistic $M^T y$, cumulant function $\beta \to c(M\beta)$, and • submodel canonical parameter vector β.

And we know that for a exponential family with canonical statistic Y^*, canonical parameter θ, mean parameter μ and cumulant function $c^*(\theta)$,

$$\sqrt{n}\left(\bar{y}_n - \mu\right) \xrightarrow{d} N(0, I(\theta))$$

and

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

Using this we get,

$$\sqrt{n}\left(M^T\left(\bar{y}_n\right) - \tau\right) \xrightarrow{d} N(0, I(\beta))$$

where $I(\beta) = \nabla^2_\beta c(M\beta) = M^T\left[\nabla^2_\theta c(\theta)\big|_{\theta=M\beta}\right]M$ and,

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N\left(0, I(\beta)^{-1}\right)$$

Problem 6 [20 points]: Do the following:

- **part a** [10 points]: Prove Lemma 1 in the exponential family notes.

Solution: The general form of an exponential family with canonical statistic X is

$$f_X(x) = \underbrace{h(x)}_{f(x)} \underbrace{\exp\{\langle X, \theta \rangle - c(\theta)\}}_{g(X|\theta)}$$

Thus, by the Neymann factorization Lemma, X is the sufficient statistic.

- **part b** [10 points]: Provide a brief explanation of why this Lemma is important without using any mathematical symbols.

Solution: This lemma demonstrates a method for dimension reduction without any loss of information. In this case, we had a n-dimensional data reduced to a single value of $T = \sum Y_i$ while still preserving sufficient information according to Sufficiency Principle.

The above six problems are worth 90 points in total. 10 points will be allocated for presentation and correct submission of the homework.