STAT 528 HW2: Exponential 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.

Solution 1:

(a) $Y \sim Bin(n, p)$

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

Thus

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

We know that

$$E_\theta(Y) = np = n\frac{e^\theta}{1+e^\theta}; \quad Var_\theta(Y) = np(1-p) = n\frac{e^\theta}{(1+e^\theta)^2}$$

Now,

$$c'(\theta) = n\frac{1}{1+e^\theta}e^\theta = np = E_\theta(Y)$$

$$\begin{aligned}
c''(\theta) &= -n\frac{e^{2\theta}}{(1+e^\theta)^2} + n\frac{e^\theta}{1+e^\theta} \\
&= n\frac{e^\theta}{(1+e^\theta)^2} \\
&= np(1-p) \\
&= Var_\theta(Y)
\end{aligned}$$

Thus equation (7) holds.

(b) $Y \sim Poi(\lambda)$

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

Thus,

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

We know that $E_\theta(Y) = \lambda = Var_\theta(Y)$

$$c'(\theta) = e^\theta = \lambda = E_\theta(Y); \quad c''(\theta) = e^\theta = \lambda = Var_\theta(Y)$$

Thus equation (7) holds.

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

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

Thus,

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

We know that $\mathrm{E}(Y) = \mu, \mathrm{E}(Y^2) = \sigma^2 + \mu^2, \mathrm{Var}(Y) = \sigma^2$ Let us now compute $Var(Y^2)$

The mgf of normal $N(\mu, \sigma^2)$ is $exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

Thus

$$M'_X(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)(\mu + \sigma^2 t)$$

$$\implies M''_X(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)(\mu + \sigma^2 t)^2 + exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\sigma^2$$

$$\implies M_X^{(3)}(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)(\mu + \sigma^2 t)^3 + 3exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)(\mu + \sigma^2 t)\sigma^2$$

$$\implies M^{(4)}(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)(\mu + \sigma^2 t)^4 + 6exp\left(\mu t + 3\frac{\sigma^2 t^2}{2}\right)(\mu + \sigma^2 t)^2\sigma^2 + 3exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\sigma^4$$

Thus

$$\mathrm{E}(Y^3) = M_X^{(3)}(t)|_{t=0} = \mu^3 + 3\mu\sigma^2$$

$$\mathrm{E}(Y^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

$$\begin{aligned} \implies Var(Y^2) &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 - (\sigma^2 + \mu^2)^2 \\ &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 - \sigma^4 - \mu^4 - 2\sigma^2\mu^2 \end{aligned}$$

$$= 4\mu^2\sigma^2 + 2\sigma^4$$

and,

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

Now,

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

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

Thus equation (7) holds.

Problem 2

(a) 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 2(a): We know that

$$h(\theta) = \langle \mu, \theta \rangle - c(\theta)$$
$$\implies h'(\theta) = \mu - \nabla c(\theta)$$

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

We know that $\mathrm{Var}_\theta(Y)$ is semi positive definite.

Consider a vector $x \neq 0$

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

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

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

(b) 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.

Solution 2(b): 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 $[\nabla^2 c(\theta)]^{-1}$ 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.

Problem 3

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

Solution 3:

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

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

Similarly

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

Choosing a λ such that $\mathrm{Var}_\theta(Y) < \lambda^2$ and b large enough such that $O(\phi^2)$ 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 subexponential.

Problem 4:

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

Solution 4:

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 subexponential.

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

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

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

Thus $\sum_{i=1}^n a_i\{y_i - \nabla c(\theta)\}$ is also subexponential.

Problem 5:

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

Solution 5:

We have the observed equals expected property for the canonical parameters of an exponential family. Using the froms of the exponential family derived above:

(a) $Y \sim Bin(n, p)$

$$y = n\frac{e^{\hat{\theta}}}{1+e^{\hat{\theta}}}$$

$$\hat{\theta} = \log\left(\frac{Y}{n-Y}\right)$$

(b) $Y_1, Y_2 \dots Y_n \sim N(\mu, \sigma^2)$

$$\left(\bar{y}_n, -\frac{\sum y_i^2}{n}\right) = \left(\frac{\hat{\theta}_1}{2\hat{\theta}_2}, -\frac{1}{2}\left(\frac{\hat{\theta}_1^2}{2\hat{\theta}_2^2} + \frac{1}{\hat{\theta}_2}\right)\right)$$

$$\implies \bar{y}_n = \frac{\hat{\theta}_1}{2\hat{\theta}_2}; \quad -\frac{\sum y_i^2}{n} = -\bar{y}_n^2 - \frac{1}{2\hat{\theta}_2}$$

$$\hat{\theta}_1 = \frac{\bar{y}_n}{\frac{\sum y_i^2}{n} - \bar{y}_n^2}; \quad \hat{\theta}_2 = \frac{1}{2\left(\frac{\sum y_i^2}{n} - \bar{y}_n^2\right)}$$

Problem 6:

Derive the asymptotic distribution for the MLE of the submodel canonical parameter vector $\hat{\beta}$ and the submodel mean-value parameter vector $\hat{\tau}$. Problem 8:

Solution 6:

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 $c^*(\theta)$

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

and

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

Using this we get

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

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

and,

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

Problem 7:

Prove Lemma 1 in the exponential family notes.

Solution 7

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.