Statistica Sinica **11**(2001), 409-418

CHARACTERIZATION OF CONJUGATE PRIORS FOR DISCRETE EXPONENTIAL FAMILIES

Jine-Phone Chou

Academia Sinica, Taipei

Abstract: Let X be a nonnegative discrete random variable distributed according to an exponential family with natural parameter $\theta \in \Theta$. Subject to some regularity we characterize conjugate prior measures on Θ through the property of linear posterior expectation of the mean parameter of $X : E\{E(X|\theta)|X = x\} = ax + b$. We also delineate some necessary conditions for the hyperparameters a and b, and find a necessary and sufficient condition that $0 < a < 1$. Besides the power series distribution with parameter space bounded above (for example, the negative binomial distribution and the logarithmic series distribution) and the Poisson distribution, we apply the result to the log-zeta distribution and all hyper-Poisson distributions.

Key words and phrases: Bounded analytic functions, characterization theorems, conjugate priors, discrete exponential families, posterior expectations.

1. Introduction

Let X come from an exponential family with natural parameter θ and density

$$dP_\theta(x) = e^{x\theta - M(\theta)} d\mu(x), \tag{1.1}$$

where μ is a discrete measure with support $\{x_0, x_1, \ldots, 0 \le x_0 < x_1 < x_2 < \cdots\}$. Let χ denote the interior of the convex hull of the support of μ, and let Θ be the natural parameter space, i.e., $\Theta = \{\theta, \int e^{\theta x} d\mu(x) < \infty\}$. It is obvious that $\chi = (x_0, x^0)$ where x^0 is the least upper bound of $\{x_0, x_1, \ldots\}$, and that Θ is convex and unbounded to the left since the $x_i, i = 1, 2, \ldots,$ are nonnegative. Throughout the paper we assume Θ is nonempty and open, and write $\Theta = (-\infty, \theta^0)$. By differentiation of $\int dP_\theta(x) = 1$, we find $E(X|\theta) = M'(\theta)$ and $\mathrm{Var}\,(X|\theta) = E(X - M'(\theta))^2 = M''(\theta)$, where $M''(\theta)$ is positive for any θ since χ is nonempty and open. For some α, β, let $\tilde{\pi}_{\alpha,\beta}$ be a measure on the Borel sets of Θ according to

$$d\tilde{\pi}_{\alpha,\beta}(\theta) = e^{\alpha\theta - \beta M(\theta)} d\theta, \tag{1.2}$$

where $d\theta$ denotes Lebesgue measure. If $\int d\tilde{\pi}_{\alpha,\beta} < \infty$, then $\tilde{\pi}_{\alpha,\beta}$ can be normalized to be a probability measure $\pi_{\alpha,\beta}$ with

$$d\pi_{\alpha,\beta}(\theta) = ke^{\alpha\theta - \beta M(\theta)} d\theta = f_{\alpha,\beta}(\theta) d\theta, \tag{1.3}$$

where $k^{-1} = \int_\Theta e^{\alpha\theta - \beta M(\theta)} d\theta$. For any support point x_i, $e^{M(\theta)} = \int e^{x\theta} d\mu(x) \geq e^{x_i\theta}\mu(x_i)$. Then for any prior measure τ on Θ with $\int_\Theta d\tau(\theta) < \infty$, $\int_\Theta e^{x_i\theta - M(\theta)} d\tau(\theta) < \infty$. Thus if a prior measure is proper, then all the posteriors are proper and can be normalized to be probability measures. We take prior measures to be proper here.

For the prior $\pi_{\alpha,\beta}$ at (1.3) and any given $x = x_i$, $i = 0, 1, \ldots$, the posterior has the form (1.3) with parameters α, β replaced by $\alpha + x_i$ and $\beta + 1$. Therefore the family $\{\pi_{\alpha,\beta}\}$ of (1.3) is closed under sampling, and is termed a distribution conjugate to the exponential family $\{P_\theta\}$ of (1.1) (Lindley (1972), pp.22-23, or Raiffa and Schlaifer (1961), pp.43-57). If θ has the conjugate prior distribution $\pi_{\alpha,\beta}$ of (1.3), then $E(M'(\theta)|X = x_i) = \frac{1}{\beta+1}x_i + \frac{\alpha}{\beta+1}$, $i = 0, 1, \ldots$, provided the posterior density $f_{\alpha+x_i,\beta+1}(\theta) \propto e^{(\alpha+x_i)\theta - (\beta+1)M(\theta)}$ approaches zero as θ approaches θ^0 (Chou (1988)).

In this paper we consider the converse result: for any X with distribution P_θ of (1.1), if

$$E(M'(\theta)|X = x_i) = ax_i + b, \quad i = 0, 1, \ldots, \tag{1.4}$$

then θ has a conjugate prior $\pi_{\alpha,\beta}$ as at (1.3) with $\alpha = a^{-1}b$, $\beta = a^{-1}(1 - a)$. We extend Diaconis and Ylvisaker's (1979) result for a parameter space that is bounded above, and Johnson's (1957, 1967) result for the Poisson distribution. We find necessary conditions on a and b in (1.4), $a > 0$ and $b > (1 - a)x_0$, and have $0 < a < 1$ if and only if $\lim_{\theta\to\theta^0} f_{\alpha,\beta}(\theta) = 0$, where $f_{\alpha,\beta}$ is the prior density. (An example with $a > 1$ and $\lim_{\theta\to\theta^0} f_{\alpha,\beta}(\theta) \neq 0$ is supplied in Section 2.1.) Besides the well-known distributions, the negative binomial and the Poisson, we introduce some interesting distributions, for example the zeta distribution, and apply our results to them.

In Section 2 we define important functions at (2.2) and (2.3), and we prove both of them are bounded and have infinitely many zeros. For any X with P_θ of (1.1), if the support points are $x_i = i$, $i = 0, 1, \ldots$, and $\Theta = (-\infty, \theta^0), \theta^0 < \infty$, Diaconis and Ylvisaker (1979) proved that the linearity at (1.4) characterizes conjugate priors. In Section 2.1 we extend their result to X with support containing x_i's where $x_i \neq 0$, $i = 1, 2, \ldots$, and $\sum_{i=1}^\infty \frac{1}{x_i} = \infty$. Therefore, besides characterizing conjugate priors for any power series distribution with parameter space bounded above, we can apply our result to non-power series distributions such as the log-zeta. In Section 2.2, we deal with general Θ. Subject to the condition that there exist a point θ^* in Θ and a positive constant c such that the mean parameter $E(X|\theta) > ce^\theta$ for all $\theta > \theta^*$, we prove that (1.4) implies Θ has a conjugate prior $\pi_{a^{-1}b,a^{-1}(1-a)}$ at (1.3), with $0 < a < 1$, $b > (1 - a)x_0$. Our condition is satisfied for the Poisson, so Johnson's result (1957, 1967) is a special case of our result. Moreover, for all hyper-Poisson distributions, we can characterize the conjugate prior information through the linearity of the posterior expectation of the mean parameter.

2. Main Result

For achieving Theorem 2.1 and Theorem 2.2, we need the following Lemma.

Lemma 2.1. *Let X have the distribution P_θ of* (1.1) *with $x_0 = 0$. Suppose that τ is a nondegenerate prior distribution on $\Theta = (-\infty, \theta^0)$ and that*

$$E(M'(\theta)|X = x) = ax + b \text{ for } x = x_i, \quad i = 0, 1, \ldots. \tag{2.1}$$

Then there exists $\tilde{\theta} \in \Theta$ such that $M'(\tilde{\theta}) = b > 0$. Moreover, if $a = 0$, *then*

$$f(z) = \int_\Theta e^{(\theta-\tilde{\theta})z}(M'(\theta) - b)e^{-M(\theta)}d\tau(\theta) \tag{2.2}$$

is a bounded function in the region $S = \{z,\ z = x + iy,\ x \in \chi,\ -\infty < y < \infty\}$, where χ is the interior of the convex hull of the support of X, and $f(x_i) = 0$ for all $i = 0, 1, \ldots$. If $a > 0$ and $\theta^0 < \infty$, then

$$\tilde{f}(z) = \int_\Theta e^{(\theta-\theta^0)z}\Big\{ae^{-M(\theta)}d\tau(\theta) - \Big[-\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(y)}d\tau(y)\Big]d\theta\Big\} \tag{2.3}$$

is a bounded function in the region $\tilde{S} = \{z, z = x + iy,\ x > x^*,\ -\infty < y < \infty\}$, where x^* is a fixed point with $0 < x^* < x_1$, and $\tilde{f}(x_i) = 0$ for all $i = 1, 2, \ldots$.

Proof. See the Appendix.

2.1. The natural parameter space is bounded above

For any random variable X having distribution P_θ of (1.1) with natural parameter space Θ bounded above, the following theorem generalizes and extends the result of Theorem 4 of Diaconis and Ylvisaker (1979). Note that if $\Theta = (-\infty, \theta^0)$, $\theta^0 < \infty$, then it is obvious that the interior χ is not bounded above, i.e., $\chi = (x_0, \infty)$.

Theorem 2.1. *Let X have the distribution P_θ of* (1.1) *with $0 \le x_0 < x_1 < \cdots$, $\sum_{i=1}^\infty \frac{1}{x_i} = \infty$ and $\Theta = (-\infty, \theta^0)$, $\theta^0 < \infty$. Suppose θ has a nondegenerate prior distribution τ. If*

$$E(E(X|\theta)|X = x) = ax + b \textit{ for } x = x_i\ , \quad i = 0, 1, \ldots, \tag{2.4}$$

then $a > 0$, $b > (1 - a)x_0$, τ is absolutely continuous with respect to Lebesgue measure, and $d\tau(\theta) = ce^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)}d\theta$. Moreover, $0 < a < 1$ if and only if the prior density approaches 0 as θ approaches θ^0.

Proof. See the Appendix.

Remark 2.1. In Theorem 2.1, $0 < a < 1$ if and only if $\lim_{\theta\to\theta^0} f(\theta) = 0$, $f(\theta)$ the prior density. The following is an example where $a \geq 1$ and the prior density does not approach zero as $\theta \to \theta^0$. Let X have a negative binomial density

$$P(x) = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)}\lambda^x(1-\lambda)^r \propto e^{\theta x - M(\theta)}, \tag{2.5}$$

where $x = 0, 1, \ldots$, $r > 0$, $\lambda \in (0,1)$, so $\theta = \log\lambda \in (-\infty, 1) = (-\infty, \theta^0)$ and $M(\theta) = -\log(1-\lambda)^r$. By applying Theorem 2.1, if λ has a nondegenerate prior distribution and $E(E(X|\lambda) = \frac{r\lambda}{1-\lambda}|X = x) = ax + b$ for any $x = 0, 1, \ldots$, then $a > 0$, $b > 0$ and θ has the conjugate prior density $f_{a^{-1}b, a^{-1}(1-a)}(\theta) \propto e^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)}$. Then λ has the conjugate Beta distribution $Be(a^{-1}b, a^{-1}(1-a)r+1)$, with density $b(\lambda) \propto \lambda^{a^{-1}b-1}(1-\lambda)^{(a^{-1}(1-a)r+1)-1}$. Therefore we have $0 < a < \frac{r}{r-1}$ if $r > 1$, or $0 < a < \infty$ if $r = 1$. However $\lim_{\theta\to\theta^0} f_{a^{-1}b, a^{-1}(1-a)}(\theta) \neq 0$ if $a \geq 1$.

Remark 2.2. Let $X_1, \ldots, X_n$ be a sample of size n from X satisfying the conditions of Theorem 2.1. Suppose τ is a nondegenerate prior distribution on Θ with

$$E(E(X|\theta)|X_1, \ldots, X_n) = a\overline{X} + b, \tag{2.6}$$

where $\overline{X}$ is the sample mean. Consider the complete and sufficient statistic $S = X_1 + \cdots + X_n$, which has the distribution from P_θ of (1.1) with x being replaced by s, $M(\theta)$ by $nM(\theta)$ and, from (2.6), $E(E(S|\theta)|S = s) = as + (nb)$ with $s = s_i$, $i = 0, 1, \ldots, 0 \leq s_0 = nx_0 < s_1 < \cdots$. Applying Theorem 2.1, we have $a > 0$, $b > (1-a)x_0$ and $d\tau(\theta) = ce^{na^{-1}b\theta - na^{-1}(1-a)M(\theta)}d\theta$.

Example 2.1 (*the power series distribution with parameter space bounded above*). Let X have the distribution P_θ of (1.1) with $x_i = i$, $i = m, m+1, m+2, \ldots, m$ a nonnegative integer and the natural parameter $\Theta = (-\infty, \theta^0)$, $\theta^0 < \infty$. Note that with $\lambda^0 = e^{\theta^0}$ and $\lambda = e^\theta \in (0, \lambda^0) = \mathbf{\Lambda}$. The density of the power series distribution X is usually in the form $f(x) = \frac{a_x\lambda^x}{\eta(\lambda)}$ where $a_x = d\mu(x)$, $\eta(\lambda) = e^{M(\ln\lambda)}$. If λ has a nondegenerate prior and if $E(E(X|\lambda)|X = x) = ax + b$ for $x = m, m+1, m+2, \ldots$, then $a > 0$ and λ has a conjugate prior with density $g(\lambda) \propto \lambda^{ab-1}\eta(\lambda)^{-a^{-1}(1-a)}$. Two well-known examples are the negative binomial distribution (see Remark 2.1), and the logarithmic series distribution with density $f(x) = (-\log(1-\lambda))^{-1}\frac{\lambda^x}{x}$, $x = 1, 2, \ldots$, and $\lambda \in (0,1) = \mathbf{\Lambda}$. For analytic properties of the logarithmic series distribution, see Johnson, Kotz and Kemp (1992).

Example 2.2. (the log-zeta distribution) Let Y be a random variable with density $f(y) = cy^{-\rho}$, $y = 1, 2, \ldots, \rho > 1$ and $c^{-1} = \sum_{y=1}^{\infty} y^{-\rho} = \xi(\rho)$, where

ξ denotes the Riemann zeta function. The distribution of Y has been used in linguistic studies by Estoup (1916) and Zipf (1949), and some interesting properties and comments can be found in Kendall (1961). Let $X = \log Y$. Note that with $\theta = -\rho$, the natural parameter space of X is $\Theta = (-\infty, -1)$, and the support $x_i = \log i$, $i = 2, 3, \ldots$, satisfies $\sum_{i=2}^{\infty} \frac{1}{x_i} = \infty$. Then if ρ has a nondegenerate prior distribution and $E(E(X|\rho)|X = x) = ax + b$ for $x = \log i$, $i = 1, 2, \ldots$, we have $a > 0$, $b > 0$ and the density of the prior distribution on ρ is $g(\rho) \propto e^{a^{-1}b\rho}\xi(\rho)^{-a^{-1}(1-a)}$.

2.2. The natural parameter space is not bounded above

If X is Poisson variable, Johnson (1957, 1967) proved that linearity at (1.4) implies the prior has a conjugate distribution. The following theorem generalizes his result.

Theorem 2.2. *Let X have the distribution P_θ of* (1.1) *with the natural parameter space $\Theta = (-\infty, \infty)$ and support points $x_i = li + k$, $i = 0, 1, \ldots$, $l > 0$, $k \geq 0$. Suppose there exists some point in Θ, say θ^*, and some positive constant c such that the mean parameter $M'(\theta) \geq ce^\theta$ for all $\theta \geq \theta^*$. Assume that θ has a nondegenerate prior distribution τ. If*

$$E(E(X|\theta)|X = x) = ax + b,\ x = x_i,\ i = 0, 1, \ldots, \tag{2.7}$$

then $0 < a < 1$, $b > (1-a)x_0$, τ is absolutely continuous with respect to Lebesgue measure and $d\tau(\theta) \propto e^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)}d\theta$.

Proof. See the Appendix.

Remarks 2.3. Let $X_1, \ldots, X_n$ be a sample of size n from X with distribution P_θ of (1.1). If the conditions of Theorem 2.2. are satisfied and $E(E(X|\theta)|x_1, \ldots, x_n) = a\overline{x} + b$, $\overline{x}$ the sample mean, then $0 < a < 1$, $b > (1-a)x_0$, the prior distribution τ is absolutely continuous with respect to Lebesgue measure, and $f(\theta) \propto e^{na^{-1}b\theta - na^{-1}(1-a)M(\theta)}$.

Example 2.3. (Johnson 1957, 1967) Let X be the Poisson variable with density $f(x) = e^{-\lambda}\frac{x^\lambda}{x!}$, $x = 0, 1, \ldots$, $\lambda > 0$. The natural parameter is $\theta = \ln \lambda$ and the mean parameter $\lambda = e^\theta$ satisfies the condition in Theorem 2.2. Thus if $E(E(X|\lambda)|X = x) = ax + b$ for $x = 0, 1, \ldots$, then $0 < a < 1$, $b > 0$ and τ has the Gamma distribution on λ with density

$$g(\lambda) = \frac{1}{\Gamma(a^{-1}b)[a(1-a)^{-1}]^{a^{-1}b}}\lambda^{a^{-1}b-1}e^{-a^{-1}(1-a)\lambda}.$$

Example 2.4. (the Hyper-Poisson distribution) Let X be a random variable with density $f_\lambda(x) = c_{\lambda,k}^{-1}\frac{\lambda^x}{k(k+1)\cdots(k+x-1)}$, where $x = 0, 1, \ldots$, $\lambda > 0$, $k > 0$, and $c_{\lambda,k} = \sum_{x=1}^{\infty} \frac{\lambda^x}{k(k+1)\cdots(k+x-1)}$. Bandwell and Crow (1964) term this family of distributions Hyper-Poisson. They classified them as sub-Poisson ($0 < k < 1$), Poisson ($k = 1$) and super-Poisson ($k > 1$), and discuss some properties of the distributions. Note that the natural parameter space $\Theta = \{\theta, \theta = \log\lambda, \lambda > 0\}$ is not bounded above and, from the recurrence relation $((k-1)+(x+1))f_\lambda(x+1) = \lambda f_\lambda(x)$, $x = 0, 1, \ldots$, the mean parameter $E(X|\lambda) = \lambda + (1-k)(1-f_\lambda(0)) \geq e^\theta - |1-k|$. Applying Theorem 2.2, if $E(E(X|\lambda)|X = x) = ax + b$ for $x = 0, 1, \ldots$, then $0 < a < 1$, $b > 0$ and the prior density of $\lambda \in (0, \infty)$ is $g(\lambda) \propto \lambda^{a^{-1}b-1} c_{\lambda,k}^{-a^{-1}(1-a)}$.

Acknowledgement

This paper was done during my visit to the Department of Mathematics of UCLA in 1998. I would like to thank my host, and former Ph.D. advisor, Donald Ylvisaker.

Appendix

Proof of Lemma 2.1. From (2.1) with $x = x_0 = 0$, $E(M'(\theta)|X = 0) = b$. Since $E(X|\theta) = M'(\theta)$ is positive, strictly increasing and continuous on Θ, $b \in M'(\Theta)$ and there exists $\tilde{\theta} \in \Theta$ such that $b = M'(\tilde{\theta}) > 0$.

If $a = 0$, we have $e^{\tilde{\theta}x_i} f(x_i) = 0$ from (2.1), and then $f(x_i) = 0$ for all x_i, $i = 0, 1, \ldots$. For any $z = x + iy \in S$,

$$|f(z)| \leq \Big|\int_{-\infty}^{\tilde{\theta}} e^{(\theta-\tilde{\theta})z}(M'(\theta) - b)e^{-M(\theta)} d\tau(\theta)\Big| + \Big|\int_{\tilde{\theta}}^{\theta^0} e^{(\theta-\tilde{\theta})z}(M'(\theta) - b)e^{-M(\theta)} d\tau(\theta)\Big|$$
$$\leq \int_{-\infty}^{\tilde{\theta}} e^{(\theta-\tilde{\theta})z}(b - M'(\theta))e^{-M(\theta)} d\tau(\theta) + \int_{\tilde{\theta}}^{\theta^0} e^{(\theta-\tilde{\theta})z}(M'(\theta) - b)e^{-M(\theta)} d\tau(\theta) \leq 2c \tag{A.1}$$

where $c = bk_0^{-1} + \int_\Theta M'(\theta)e^{-M(\theta)} d\tau(\theta)$, $k_0^{-1} = \int_\Theta e^{-M(\theta)} d\tau(\theta)$. The second inequality in (A.1) uses $|t^z| = t^x$ for any $z = x + iy$ and $t > 0$. For the last inequality, since $x \in \chi$, there exists a support point $x_i \geq x$ for which

$$\int_{\tilde{\theta}}^{\theta^0} e^{(\theta-\tilde{\theta})x}(M'(\theta) - b)e^{-M(\theta)} d\tau(\theta) \leq \int_{\tilde{\theta}}^{\theta^0} e^{(\theta-\tilde{\theta})x_i}(M'(\theta) - b)e^{-M(\theta)} d\tau(\theta)$$
$$= \int_{-\infty}^{\tilde{\theta}} e^{(\theta-\tilde{\theta})x_i}(b - M'(\theta))e^{-M(\theta)} d\tau(\theta) \leq c.$$

The last equality in the above follows from $f(x_i) = 0$.

Now assume that $a > 0$ and $\theta^0 < \infty$. From (2.1),

$$\int_{-\infty}^{\theta^0} e^{\theta x}(M'(\theta) - b)e^{-M(\theta)} d\tau(\theta) = ax \int_{-\infty}^{\theta^0} e^{x\theta} e^{-M(\theta)} d\tau(\theta), \; x = x_i, \; i = 0, 1, \ldots. \tag{A.2}$$

The argument from (A.3) to (A.4) proceeds along the lines (4.5)-(4.6) of Theorem 4 of Diaconis and Ylvisaker (1979). Transform the left side of (A.2) as follows:

$$\begin{aligned}
&\int_{-\infty}^{\theta^0}\Big[\int_{-\infty}^{\theta} xe^{xy}dy\Big](M'(\theta)-b)e^{-M(\theta)}d\tau(\theta)\\
&=\int_{-\infty}^{\theta^0} xe^{xy}\Big[\int_{y}^{\theta^0}(M'(\theta)-b)e^{-M(\theta)}d\tau(\theta)\Big]dy\\
&=-\int_{-\infty}^{\theta^0} xe^{x\theta}\Big[\int_{-\infty}^{\theta}(M'(y)-b)e^{-M(y)}d\tau(y)\Big]d\theta. \qquad \text{(A.3)}
\end{aligned}$$

In (A.3) the interchange of integration can be easily checked and (A.2) has been invoked with $x = 0$ to produce the final equality. From (A.2) and (A.3) we have

$$\begin{aligned}
&\int_{-\infty}^{\theta^0} e^{x\theta}\Big[-\int_{-\infty}^{\theta}(M'(y)-b)e^{-M(y)}d\tau(y)\Big]d\theta\\
&=a\int_{-\infty}^{\theta^0} e^{x\theta}e^{-M(\theta)}d\tau(\theta) \text{ for } x=x_i,\ i=1,2,\ldots. \qquad \text{(A.4)}
\end{aligned}$$

Multiply both sides of (A.4) by $e^{-\theta^0 x_i}$ to get $\tilde{f}(x_i) = 0$ for $x = x_i$, $i = 1, 2, \ldots$. Let $h(\theta) = -\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(y)}d\tau(y)$. Since $(M'(y) - b) \le 0$ if and only if $y \le \tilde{\theta}$, $h(\theta)$ is nondecreasing for $\theta \le \tilde{\theta}$ and nonincreasing for $\theta \ge \tilde{\theta}$. From (2.1) with $x = x_0 = 0$, $\lim_{\theta\to\theta^0} h(\theta) = 0$. So $h(\theta)$ is always nonnegative and achieves the maximum value at $\theta = \tilde{\theta}$. For any $z = x + iy \in \tilde{S} = \{z,\ z = x + iy,\ x > x^*,\ -\infty < y < \infty\}$,

$$\begin{aligned}
|\tilde{f}(z)| &\le \Big|\int_{\Theta} e^{(\theta-\theta^0)z}ae^{-M(\theta)}d\tau(\theta)\Big| + \Big|\int_{\Theta} e^{(\theta-\theta^0)z}h(\theta)d\theta\Big|\\
&\le \int_{-\infty}^{\theta^0} e^{(\theta-\theta^0)x}ae^{-M(\theta)}d\tau(\theta) + \int_{-\infty}^{\theta^0} e^{(\theta-\theta^0)x}h(\theta)d\theta\\
&\le a\int_{-\infty}^{\theta^0} e^{-M(\theta)}d\tau(\theta) + \int_{-\infty}^{\theta^0} e^{(\theta-\theta^0)x^*}h(\tilde{\theta})d\theta\\
&\le a\int_{-\infty}^{\theta^0} e^{-M(\theta)}d\tau(\theta) + h(\tilde{\theta})\frac{1}{x^*} \le \tilde{c}.
\end{aligned}$$

Therefore $\tilde{f}(z)$ is bounded for all $z \in \tilde{S}$.

Proof of Theorem 2.1. First we show $a > 0$ if $x_0 = 0$. If a is zero, define $f(z)$ as at (2.2) and Lemma 2.1 has $f(z)$ bounded in the region $S = \{z,\ z = x + iy,\ x \in \chi = (0, \infty), -\infty < y < \infty\}$. Moreover f is continuous and then analytic by Moresa's theorem (Rudin (1987), pp.208). For $z \in \boldsymbol{u} = \{z, |z| < 1\}$, define $g(z) = f(\frac{1+z}{1-z})$. Then g is bounded, analytic and $g(\alpha_i) = 0$, $\alpha_i = \frac{x_i-1}{x_i+1}$, $i = 1, 2, \ldots$. Since

$\sum_{i=1}^{\infty} \frac{1}{x_i} = \infty$ implies $\sum_{i=1}^{\infty}(1-|\alpha_i|) = \infty$, $g \equiv 0$ on $\boldsymbol{u}$ (Rudin (1987), pp.312), and then $f \equiv 0$ on S. Therefore $\int_\Theta (M'(\theta)-b)e^{x\theta - M(\theta)} d\tau(\theta) = 0$ for all $x \in \chi$, an open interval in R^1. It follows that $M'(\theta) - b$ vanishes on the support of τ, violating the strictly increasing property of $M'(\theta)$. So $a \neq 0$. If $a < 0$, choose $x_j \in \chi = (0,\infty)$ with $ax_j + b < 0$. Then $E(M'(\theta)|X = x_j) = ax_j + b < 0$, which violates the fact that $M'(\theta) = E(X|\theta) > 0$ for any θ. Therefore $a > 0$.

Suppose $0 < x^* < x_1$ and let $\tilde{f}(z)$ and $\tilde{S}$ be as given in (2.3). Define $\tilde{g}(z) = \tilde{f}(\frac{1+z}{1-z} + x^*)$. As before we can prove $\tilde{g}(z) \equiv 0$ on $\boldsymbol{u} = \{z, |z| < 1\}$, and then $\tilde{f}(z) \equiv 0$ on $\tilde{S} = \{z, z = x + iy, x \in (x^*,\infty), -\infty < y < \infty\}$. In particular $\tilde{f}(x) = 0$ for all $x \in (x^*,\infty)$, an open interval of R, so $\{ae^{-M(\theta)}d\tau(\theta) - [-\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(\theta)}d\tau(y)]d\theta\}$ is a zero measure on Θ. Therefore $ae^{-M(\theta)}d\tau(\theta) = [-\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(\theta)}d\tau(y)]d\theta$ and τ is absolutely continuous $(d\theta)$ with density f which satisfies the differential equation $af'(\theta) - aM'(\theta)f(\theta) = -(M'(\theta) - b)f(\theta)$ for any $\theta \in \Theta$. Solving the equation, we have $d\tau(\theta) = f(\theta)d\theta = ce^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)}d\theta$.

For the general case $x_0 \geq 0$, let $\tilde{X} = X - x_0$, $\tilde{x}_i = x_i - x_0$. Since $\tilde{X}$ is also from P_θ of (1.1) with x being replaced by $\tilde{x}$, $M(\theta)$ by $\tilde{M}(\theta) = M(\theta) - x_0\theta$, (2.4) implies $E[E(\tilde{X}|\theta)|\tilde{X} = \tilde{x}] = a\tilde{x} + b'$, where $\tilde{x} = \tilde{x}_i$, $i = 0,1,\ldots$, $b' = b + (a-1)x_0$. Applying the previous result for the first support point $\tilde{x}_0$ being zero, we have $a > 0$ and τ is absolutely continuous with $d\tau(\theta) = f(\theta)d\theta = ce^{a^{-1}(b+(a-1)x_0)\theta - a^{-1}(1-a)(M(\theta)-x_0\theta)}d\theta = ce^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)}d\theta$. For any θ, $E(X|\theta) \in \chi = (x_0,\infty)$, $E(E(X|\theta)|X = x) = ax + b \in \chi$ for any $x = x_0, x_1, \ldots$. Therefore $ax_0 + b \in \chi$, i.e., $b > (1-a)x_0$. Moreover, since $f(\theta) \propto e^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)}$, $a > 0$, $b > (1-a)x_0$, $x_0 \geq 0$ and $e^{M(\theta)} = \int e^{x\theta}d\mu(x)$ (which approaches infinite as θ approaches $\theta^0 < \infty$), it is obvious that $0 < a < 1$ if and only if $\lim_{\theta\to\theta^0} f(\theta) = 0$.

Proof of Theorem 2.2. For $l = 1$, $k = 0$, i.e., $x_i = i$, $i = 0,1,\ldots$, we first prove $a > 0$ in (2.7). If $a = 0$, define $f(z)$ as at (2.2) and follow the same argument as that in the proof of Theorem 2.1 to get $a > 0$. From (2.7), by the argument for (A.3) and (A.4), we have

$$\int_{-\infty}^{\theta^0} e^{x\theta}\Big[-\int_{-\infty}^{\theta}(M'(y) - b)e^{-M(y)}d\tau(y)\Big]d\theta = a\int_{-\infty}^{\theta^0} e^{x\theta}e^{-M(\theta)}d\tau(\theta), \quad \text{for } x = 1,2,\ldots. \tag{A.5}$$

Let $t = e^\theta$ and define a probability distribution m on $(0, e^{\theta^0})$ ($e^{\theta^0} \equiv \infty$ if $\theta^0 = \infty$) according to $\int_E dm(t) = \int_{h^{-1}(E)} ke^{\theta - M(\theta)}d\tau(\theta)$, where $k^{-1} = \int_\theta e^{\theta - M(\theta)}d\tau(\theta)$, E is any Borel set of $(0, e^{\theta^0})$ and $h(\theta) = e^\theta = t$. Let μ_n denote the nth moment of

m. For $n = 0, 1, \ldots,$

$$\frac{\mu_{n+1}}{\mu_n} = \frac{\displaystyle\int t^{n+1} dm\,(t)}{\displaystyle\int t^n dm\,(t)} = \frac{\displaystyle\int_{-\infty}^{\theta^0} e^{(n+1)\theta} e^{\theta - M(\theta)} d\tau\,(\theta)}{\displaystyle\int_{-\infty}^{\theta^0} e^{n\theta} e^{\theta - M(\theta)} d\tau\,(\theta)}$$

$$\leq \frac{\displaystyle\int_{-\infty}^{\theta^*} e^{(n+1)\theta} e^{\theta - M(\theta)} d\tau\,(\theta)}{\displaystyle\int_{-\infty}^{\theta^0} e^{n\theta} e^{\theta - M(\theta)} d\tau\,(\theta)} + \frac{\frac{1}{c}\displaystyle\int_{\theta^*}^{\theta^0} e^{(n+1)\theta} M'\,(\theta)\, e^{-M(\theta)} d\tau\,(\theta)}{\displaystyle\int_{-\infty}^{\theta^0} e^{n\theta} e^{\theta - M(\theta)} d\tau\,(\theta)}$$

$$\leq e^{\theta^*} + \frac{1}{c}\left(a\,(n+1) + b\right) = \alpha n + \beta, \tag{A.6}$$

where $\alpha = \frac{a}{c}$, $\beta = e^{\theta^*} + \frac{a}{c} + b$. The inequalities of (A.6) follow from the fact that $M'(\theta) > ce^{\theta}$ for all $\theta > \theta^*$, and (2.7). Since $\mu_0 = 1$, it follows that $\mu_1 \leq \beta$, $\mu_2 \leq \beta(\beta+\alpha), \ldots, \mu_n \leq \beta(\beta+\alpha)\cdots(\beta+(n-1)\alpha)$. Hence $\mu_n \leq (\beta+(n-1)\alpha)^n$, and then $\sum_{n=0}^{\infty} \mu_{2n}^{-\frac{1}{2n}} = \infty$. From Carleman's Uniqueness Theorem (Akhiezer (1965, pp.85-86)), the distribution m is uniquely determined by its moments μ_n, $n = 0, 1, \ldots,$ and then (A.5) implies $ae^{\theta - M(\theta)} d\tau(\theta) = -e^{\theta}[\int_{-\infty}^{\theta} (M'(y) - b) e^{-M(y)} d\tau(y)] d\theta$. By the same argument as in Theorem 2.1, we have τ is absolutely continuous with density $f(\theta) \propto e^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)}$.

For the general case that $l > 0$, $k \geq 0$, we consider $\tilde{X} = l^{-1}(X - k)$. Since $\tilde{X}$ is also from P_θ of (1.1) with X being replaced by $\tilde{X}$, the natural parameter θ by $l\theta$, and $M(\theta)$ by $M(\theta) - k\theta$, (2.7) implies $E(E(\tilde{X}|\theta)|\tilde{X} = \tilde{x}) = a\tilde{x} + b'$, where $\tilde{x} = 0, 1, \ldots,$ $b' = l^{-1}[b + (a-1)k]$. Applying the previous result, we have $a > 0$ and $d\tau(\theta) \propto e^{a^{-1}b'(l\theta) - a^{-1}(1-a)(M(\theta) - k\theta)} d\theta \propto e^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)} d\theta$.

Since $M'\,(\theta) > ce^{\theta}$ for all $\theta > \theta^*$, from the Mean Value Theorem, $M\,(\theta)$ approaches infinity as $\theta \to \theta^0 = \infty$. Note that the prior density $f\,(\theta)$ approaches zero as $\theta \to \infty$. Therefore by the same argument as in the proof of Theorem 2.1, we find $0 < a < 1$, and $b > (1-a)\,x_0$.

References

Akhiezer, N. I. (1969). *The Classical Moment Problem.* Hafner, New York.

Bardwell, G. E. and Crow, E. L. (1964). A two-parameter family of hyper-Poisson distributions. *J. Amer. Statist. Assoc.* **59**, 133-141.

Chou, J. P. (1988). An identity for multidimensional continuous exponential families and its applications. *J. Multivariate Anal.* **19**, 314-328.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **6**, 269-281.

Estoup, J. B. (1916). *Les Gammes Sténographiques.* Institut Sténographique, Paris.

Johnson, W. L. (1957). Uniqueness of a result in the theory of accident proneness. *Biometrika* **44**, 430-431.

Johnson W. L. (1967). Note on a uniqueness relation in certain accident proneness models. *J. Amer. Statist. Assoc.* **62**, 288-289.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions.* Wiley, New York.

Kendall, M. G. (1961). Natural law in the social sciences. *J. Roy. Statist. Soc. Ser. A* **124**, 1-18.

Lindley, D. V. (1972). *Bayesian Statistics, A Review.* SIAM, Philadelphia.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Harvard University, Boston.

Rudin, W. (1987). *Real and Complex Analysis.* McGraw-Hill, New York.

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort.* Addison-Wesley, Reading, Mass.

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

E-mail: jpchou@stat.sinica.edu.tw

(Received October 1998; accepted September 2000)