Chapter 3

Exponential Families

3.1 Regular Exponential Families

The theory of exponential families will be used in the following chapters to study some of the most important topics in statistical inference such as minimal and complete sufficient statistics, maximum likelihood estimators (MLEs), uniform minimum variance estimators (UMVUEs) and the Fréchet Cramér Rao lower bound (FCRLB), uniformly most powerful (UMP) tests and large sample theory.

Often a "brand name distribution" such as the normal distribution will have three useful parameterizations: the *usual parameterization* with parameter space Θ_U is simply the formula for the probability distribution or mass function (pdf or pmf, respectively) given when the distribution is first defined. The *k-parameter exponential family parameterization* with parameter space Θ, given in Definition 3.1 below, provides a simple way to determine if the distribution is an exponential family while the *natural parameterization* with parameter space Ω, given in Definition 3.2 below, is used for *theory* that requires a complete sufficient statistic.

Definition 3.1. A *family* of joint pdfs or joint pmfs $\{f(\boldsymbol{y}|\boldsymbol{\theta}) : \boldsymbol{\theta} = (\theta_1, ..., \theta_j) \in \Theta \}$ for a random vector $\boldsymbol{Y}$ is an **exponential family** if

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = h(\boldsymbol{y})c(\boldsymbol{\theta}) \exp\left[\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(\boldsymbol{y})\right] \tag{3.1}$$

for $\boldsymbol{y} \in \mathcal{Y}$ where $c(\boldsymbol{\theta}) \geq 0$ and $h(\boldsymbol{y}) \geq 0$. The functions c, h, t_i, and w_i are real valued functions. The parameter $\boldsymbol{\theta}$ can be a scalar and $\boldsymbol{y}$ can be a scalar.

It is crucial that $c, w_1, ..., w_k$ do not depend on $\boldsymbol{y}$ and that $h, t_1, ..., t_k$ do not depend on $\boldsymbol{\theta}$. The support of the distribution is $\mathcal{Y}$ and the parameter space is Θ. The family is a **k-parameter exponential family** if k is the smallest integer where (3.1) holds.

Notice that the distribution of Y is an exponential family if

$$f(y|\boldsymbol{\theta}) = h(y)c(\boldsymbol{\theta}) \exp\left[\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(y)\right] \tag{3.2}$$

and the distribution is a one parameter exponential family if

$$f(y|\theta) = h(y)c(\theta) \exp[w(\theta)t(y)]. \tag{3.3}$$

The parameterization is not unique since, for example, w_i could be multiplied by a nonzero constant a if t_i is divided by a. Many other parameterizations are possible. If $h(y) = g(y)I_{\mathcal{Y}}(y)$, then usually $c(\boldsymbol{\theta})$ and $g(y)$ are positive, so another parameterization is

$$f(y|\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(y) + d(\boldsymbol{\theta}) + S(y)\right] I_{\mathcal{Y}}(y) \tag{3.4}$$

where $S(y) = \log(g(y))$, $d(\boldsymbol{\theta}) = \log(c(\boldsymbol{\theta}))$, and $\mathcal{Y}$ does not depend on $\boldsymbol{\theta}$.

To demonstrate that $\{f(\boldsymbol{y}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ is an exponential family, find $h(y), c(\boldsymbol{\theta}), w_i(\boldsymbol{\theta})$ and $t_i(\boldsymbol{y})$ such that (3.1), (3.2), (3.3) or (3.4) holds.

Theorem 3.1. Suppose that $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ are iid random vectors from an exponential family. Then the joint distribution of $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ follows an exponential family.

Proof. Suppose that $f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i)$ has the form of (3.1). Then by independence,

$$f(\boldsymbol{y}_1, ..., \boldsymbol{y}_n) = \prod_{i=1}^{n} f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i) = \prod_{i=1}^{n} h(\boldsymbol{y}_i)c(\boldsymbol{\theta}) \exp\left[\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(\boldsymbol{y}_i)\right]$$

$$= [\prod_{i=1}^{n} h(\boldsymbol{y}_i)][c(\boldsymbol{\theta})]^n \prod_{i=1}^{n} \exp\left[\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(\boldsymbol{y}_i)\right]$$

$$= [\prod_{i=1}^n h(\boldsymbol{y}_i)][c(\boldsymbol{\theta})]^n \exp\left(\sum_{i=1}^n \left[\sum_{j=1}^k w_j(\boldsymbol{\theta}) t_j(\boldsymbol{y}_i)\right]\right)$$

$$= [\prod_{i=1}^n h(\boldsymbol{y}_i)][c(\boldsymbol{\theta})]^n \exp\left[\sum_{j=1}^k w_j(\boldsymbol{\theta}) \left(\sum_{i=1}^n t_j(\boldsymbol{y}_i)\right)\right].$$

To see that this has the form (3.1), take $h^*(\boldsymbol{y}_1, ..., \boldsymbol{y}_n) = \prod_{i=1}^n h(\boldsymbol{y}_i)$, $c^*(\boldsymbol{\theta}) = [c(\boldsymbol{\theta})]^n$, $w_j^*(\boldsymbol{\theta}) = w_j(\boldsymbol{\theta})$ and $t_j^*(\boldsymbol{y}_1, ..., \boldsymbol{y}_n) = \sum_{i=1}^n t_j(\boldsymbol{y}_i)$. QED

The parameterization that uses the **natural parameter $\boldsymbol{\eta}$** is especially useful for theory. See Definition 3.3 for the natural parameter space Ω.

Definition 3.2. Let Ω be the natural parameter space for $\boldsymbol{\eta}$. The **natural parameterization for an exponential family** is

$$f(\boldsymbol{y}|\boldsymbol{\eta}) = h(\boldsymbol{y}) b(\boldsymbol{\eta}) \exp\left[\sum_{i=1}^k \eta_i t_i(\boldsymbol{y})\right] \tag{3.5}$$

where $h(\boldsymbol{y})$ and $t_i(\boldsymbol{y})$ are the same as in Equation (3.1) and $\boldsymbol{\eta} \in \Omega$. The natural parameterization for a random variable Y is

$$f(y|\boldsymbol{\eta}) = h(y) b(\boldsymbol{\eta}) \exp\left[\sum_{i=1}^k \eta_i t_i(y)\right] \tag{3.6}$$

where $h(y)$ and $t_i(y)$ are the same as in Equation (3.2) and $\boldsymbol{\eta} \in \Omega$. Again, the parameterization is not unique. If $a \neq 0$, then $a\eta_i$ and $t_i(y)/a$ would also work.

Notice that the natural parameterization (3.6) has the same form as (3.2) with $\boldsymbol{\theta}^* = \boldsymbol{\eta}$, $c^*(\boldsymbol{\theta}^*) = b(\boldsymbol{\eta})$ and $w_i(\boldsymbol{\theta}^*) = w_i(\boldsymbol{\eta}) = \eta_i$. In applications often $\boldsymbol{\eta}$ and Ω are of interest while $b(\boldsymbol{\eta})$ is not computed.

The next important idea is that of a regular exponential family (and of a full exponential family). Let $d_i(x)$ denote $t_i(y)$, $w_i(\boldsymbol{\theta})$ or η_i. A *linearity constraint* is satisfied by $d_1(x), ..., d_k(x)$ if $\sum_{i=1}^k a_i d_i(x) = c$ for some constants a_i and c and for all x in the sample or parameter space where not all of the $a_i = 0$. If $\sum_{i=1}^k a_i d_i(x) = c$ for all x only if $a_1 = \cdots = a_k = 0$, then the $d_i(x)$ do not satisfy a linearity constraint. In linear algebra, we would say that the $d_i(x)$ are *linearly independent* if they do not satisfy a linearity constraint.

Let $\tilde{\Omega}$ be the set where the integral of the kernel function is finite:

$$\tilde{\Omega} = \{\boldsymbol{\eta} = (\eta_1, ..., \eta_k) : \frac{1}{b(\boldsymbol{\eta})} \equiv \int_{-\infty}^{\infty} h(y) \exp[\sum_{i=1}^{k} \eta_i t_i(y)] dy < \infty\}. \quad (3.7)$$

Replace the integral by a sum for a pmf. An interesting fact is that $\tilde{\Omega}$ is a convex set.

Definition 3.3. Condition E1: the natural parameter space $\Omega = \tilde{\Omega}$.
Condition E2: assume that in the natural parameterization, neither the η_i nor the t_i satisfy a linearity constraint.
Condition E3: Ω is a k-dimensional open set.
If conditions E1), E2) and E3) hold then the exponential family is a **regular exponential family** (REF).
If conditions E1) and E2) hold then the exponential family is a *full exponential family.*

Notation. A kP–REF is a k parameter regular exponential family. So a 1P–REF is a 1 parameter REF and a 2P–REF is a 2 parameter REF.

Notice that every REF is full. Any k–dimensional open set will contain a k–dimensional rectangle. A k–fold cross product of nonempty open intervals is a k–dimensional open set. For a one parameter exponential family, a one dimensional rectangle is just an interval, and the only type of function of one variable that satisfies a linearity constraint is a constant function. In the definition of an exponential family, $\boldsymbol{\theta}$ is a $j \times 1$ vector. Typically $j = k$ if the family is a kP–REF. If $j < k$ and k is as small as possible, the family will usually not be regular.

Some care has to be taken with the definitions of Θ and Ω since formulas (3.1) and (3.6) need to hold for every $\boldsymbol{\theta} \in \Theta$ and for every $\boldsymbol{\eta} \in \Omega$. For a continuous random variable or vector, the pdf needs to exist. Hence all degenerate distributions need to be deleted from Θ_U to form Θ and Ω. For continuous and discrete distributions, the natural parameter needs to exist (and often does not exist for discrete degenerate distributions). As a rule of thumb, remove values from Θ_U that cause the pmf to have the form 0^0. For example, for the binomial(k, ρ) distribution with k known, the natural parameter $\eta = \log(\rho/(1-\rho))$. Hence instead of using $\Theta_U = [0, 1]$, use $\rho \in \Theta = (0, 1)$, so that $\eta \in \Omega = (-\infty, \infty)$.

These conditions have some redundancy. If Ω contains a k-dimensional rectangle, no η_i is completely determined by the remaining $\eta_j's$. In particular, the η_i cannot satisfy a linearity constraint. If the η_i do satisfy a linearity constraint, then the η_i lie on a hyperplane of dimension at most k, and such a surface cannot contain a k-dimensional rectangle. For example, if $k = 2$, a line cannot contain an open box. If $k = 2$ and $\eta_2 = \eta_1^2$, then the parameter space does not contain a 2-dimensional rectangle, although η_1 and η_2 do not satisfy a linearity constraint.

The most important 1P–REFs are the binomial (k, ρ) distribution with k known, the exponential (λ) distribution, and the Poisson (θ) distribution.

Other 1P–REFs include the Burr (ϕ, λ) distribution with ϕ known, the double exponential (θ, λ) distribution with θ known, the two parameter exponential (θ, λ) distribution with θ known, the generalized negative binomial (μ, κ) distribution if κ is known, the geometric (ρ) distribution, the half normal (μ, σ^2) distribution with μ known, the largest extreme value (θ, σ) distribution if σ is known, the smallest extreme value (θ, σ) distribution if σ is known, the inverted gamma (ν, λ) distribution if ν is known, the logarithmic (θ) distribution, the Maxwell–Boltzmann (μ, σ) distribution if μ is known, the negative binomial (r, ρ) distribution if r is known, the one sided stable (σ) distribution, the Pareto (σ, λ) distribution if σ is known, the power (λ) distribution, the Rayleigh (μ, σ) distribution if μ is known, the Topp-Leone (ν) distribution, the truncated extreme value (λ) distribution, the Weibull (ϕ, λ) distribution if ϕ is known and the Zeta (ν) distribution. A one parameter exponential family can often be obtained from a k–parameter exponential family by holding $k - 1$ of the parameters fixed. Hence a normal (μ, σ^2) distribution is a 1P–REF if σ^2 is known. Usually assuming scale, location or shape parameters are known is a bad idea.

The most important 2P–REFs are the beta (δ, ν) distribution, the gamma (ν, λ) distribution and the normal (μ, σ^2) distribution. The chi (p, σ) distribution and the lognormal (μ, σ^2) distribution are also 2–parameter exponential families. Example 3.9 will show that the inverse Gaussian distribution is full but not regular. The two parameter Cauchy distribution is not an exponential family because its pdf cannot be put into the form of Equation (3.1).

The natural parameterization can result in a family that is much larger than the family defined by the usual parameterization. See the definition of

$\Omega = \tilde{\Omega}$ given by Equation (3.7). Casella and Berger (2002, p. 114) remarks that

$$\{\boldsymbol{\eta} : \boldsymbol{\eta} = (w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta}))|\boldsymbol{\theta} \in \Theta\} \subseteq \Omega, \tag{3.8}$$

but often Ω is a strictly larger set.

Remark 3.1. For the families in Chapter 10 other than the χ^2_p and inverse Gaussian distributions, make the following assumptions. Assume that $\eta_i = w_i(\boldsymbol{\theta})$ and that $\dim(\Theta) = k = \dim(\Omega)$. Assume the usual parameter space Θ_U is as big as possible (replace the integral by a sum for a pmf):

$$\Theta_U = \{\boldsymbol{\theta} \in \Re^k : \int f(y|\boldsymbol{\theta})dy = 1\},$$

and let

$$\Theta = \{\boldsymbol{\theta} \in \Theta_U : w_1(\boldsymbol{\theta}), ..., w_k(\boldsymbol{\theta}) \text{ are defined } \}.$$

Then assume that the natural parameter space satisfies condition E1) with

$$\Omega = \{(\eta_1, ..., \eta_k) : \eta_i = w_i(\boldsymbol{\theta}) \text{ for } \boldsymbol{\theta} \in \Theta\}.$$

In other words, simply define $\eta_i = w_i(\boldsymbol{\theta})$. For many common distributions, $\boldsymbol{\eta}$ is a one to one function of $\boldsymbol{\theta}$, and the above map is correct, especially if Θ_U is an open interval or cross product of open intervals.

Example 3.1. Let $f(x|\mu, \sigma)$ be the $N(\mu, \sigma^2)$ family of pdfs. Then $\boldsymbol{\theta} = (\mu, \sigma)$ where $-\infty < \mu < \infty$ and $\sigma > 0$. Recall that μ is the mean and σ is the standard deviation (SD) of the distribution. The usual parameterization is

$$f(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(x-\mu)^2}{2\sigma^2}) I_{\Re}(x)$$

where $\Re = (-\infty, \infty)$ and the indicator $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. Notice that $I_{\Re}(x) = 1 \;\; \forall x$. Since

$$f(x|\mu, \sigma) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-\mu}{2\sigma^2})}_{c(\mu,\sigma)\geq 0} \exp(\underbrace{\frac{-1}{2\sigma^2}}_{w_1(\boldsymbol{\theta})} \underbrace{x^2}_{t_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\boldsymbol{\theta})} \underbrace{x}_{t_2(x)}) \underbrace{I_{\Re}(x)}_{h(x)\geq 0},$$

this family is a 2-parameter exponential family. Hence $\eta_1 = -0.5/\sigma^2$ and $\eta_2 = \mu/\sigma^2$ if $\sigma > 0$, and $\Omega = (-\infty, 0) \times (-\infty, \infty)$. Plotting η_1 on the horizontal axis and η_2 on the vertical axis yields the left half plane which certainly contains a 2-dimensional rectangle. Since t_1 and t_2 lie on a quadratic rather than a line, the family is a 2P–REF. Notice that if $X_1, ..., X_n$ are iid $N(\mu, \sigma^2)$ random variables, then the joint pdf $f(\boldsymbol{x}|\boldsymbol{\theta}) = f(x_1, ..., x_n|\mu, \sigma) =$

$$\underbrace{[\frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-\mu}{2\sigma^2})]^n}_{C(\mu,\sigma)\geq 0} \exp(\underbrace{\frac{-1}{2\sigma^2}}_{w_1(\boldsymbol{\theta})} \underbrace{\sum_{i=1}^n x_i^2}_{T_1(\boldsymbol{x})} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\boldsymbol{\theta})} \underbrace{\sum_{i=1}^n x_i}_{T_2(\boldsymbol{x})}) \underbrace{1}_{h(\boldsymbol{x})\geq 0},$$

and is thus a 2P–REF.

Example 3.2. The χ^2_p distribution is not a REF since the usual parameter space Θ_U for the χ^2_p distribution is the set of integers, which is neither an open set nor a convex set. Nevertheless, the natural parameterization is the gamma($\nu, \lambda = 2$) family which is a REF. Note that this family has uncountably many members while the χ^2_p family does not.

Example 3.3. The binomial(k, ρ) pmf is

$$f(x|\rho) = \binom{k}{x} \rho^x (1-\rho)^{k-x} \; I_{\{0,...,k\}}(x)$$

$$= \underbrace{\binom{k}{x} I_{\{0,...,k\}}(x)}_{h(x)\geq 0} \; \underbrace{(1-\rho)^k}_{c(\rho)\geq 0} \; \exp[\underbrace{\log(\frac{\rho}{1-\rho})}_{w(\rho)} \underbrace{x}_{t(x)}]$$

where $\Theta_U = [0, 1]$. Since the pmf and $\eta = \log(\rho/(1-\rho))$ is undefined for $\rho = 0$ and $\rho = 1$, we have $\Theta = (0, 1)$. Notice that $\Omega = (-\infty, \infty)$.

Example 3.4. The uniform(0,θ) family is not an exponential family since the support $\mathcal{Y}_\theta = (0, \theta)$ depends on the unknown parameter θ.

Example 3.5. If Y has a half normal distribution, $Y \sim \text{HN}(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{2}{\sqrt{2\pi}\;\sigma} \exp{(\frac{-(y-\mu)^2}{2\sigma^2})}$$

where $\sigma > 0$ and $y \geq \mu$ and μ is real. Notice that

$$f(y) = \frac{2}{\sqrt{2\pi}\;\sigma} I(y \geq \mu) \exp\left[(\frac{-1}{2\sigma^2})(y-\mu)^2 \right]$$

is a 1P–REF if μ is known. Hence $\Theta = (0, \infty)$, $\eta = -1/(2\sigma^2)$ and $\Omega = (-\infty, 0)$. Notice that a different 1P–REF is obtained for each value of μ when μ is known with support $\mathcal{Y}_\mu = [\mu, \infty)$. If μ is not known, then this family is not an exponential family since the support depends on μ.

The following two examples are important examples of REFs where $\dim(\Theta) > \dim(\Omega)$.

Example 3.6. If the t_i or η_i satisfy a linearity constraint, then the number of terms in the exponent of Equation (3.1) can be reduced. Suppose that $Y_1, ..., Y_n$ follow the multinomial $M_n(m, \rho_1, ..., \rho_n)$ distribution which has $\dim(\Theta) = n$ if m is known. Then $\sum_{i=1}^n Y_i = m$, $\sum_{i=1}^n \rho_i = 1$ and the joint pmf of $\boldsymbol{Y}$ is

$$f(\boldsymbol{y}) = m! \prod_{i=1}^{n} \frac{\rho_i^{y_i}}{y_i!}.$$

The support of $\boldsymbol{Y}$ is $\mathcal{Y} = \{\boldsymbol{y} : \sum_{i=1}^n y_i = m \text{ and } 0 \le y_i \le m \text{ for i} = 1, ..., \text{n}\}$.

Since Y_n and ρ_n are known if $Y_1, ..., Y_{n-1}$ and $\rho_1, ..., \rho_{n-1}$ are known, we can use an equivalent joint pmf f_{EF} in terms of $Y_1, ..., Y_{n-1}$. Let

$$h(y_1, ..., y_{n-1}) = \left[\frac{m!}{\prod_{i=1}^n y_i!}\right] \ I[(y_1, ..., y_{n-1}, y_n) \in \mathcal{Y}].$$

(This is a function of $y_1, ..., y_{n-1}$ since $y_n = m - \sum_{i=1}^{n-1} y_i$.) Then $Y_1, ..., Y_{n-1}$ have a $M_n(m, \rho_1, ..., \rho_n)$ distribution if the joint pmf of $Y_1, ..., Y_{n-1}$ is

$$f_{EF}(y_1, ..., y_{n-1}) = \exp[\sum_{i=1}^{n-1} y_i \log(\rho_i) + (m - \sum_{i=1}^{n-1} y_i) \log(\rho_n)] \ h(y_1, ..., y_{n-1})$$

$$= \exp[m \log(\rho_n)] \ \exp[\sum_{i=1}^{n-1} y_i \log(\rho_i/\rho_n)] \ h(y_1, ..., y_{n-1}). \qquad (3.9)$$

Since $\rho_n = 1 - \sum_{j=1}^{n-1} \rho_j$, this is an $n-1$ dimensional REF with

$$\eta_i = \log(\rho_i/\rho_n) = \log\left(\frac{\rho_i}{1 - \sum_{j=1}^{n-1} \rho_j}\right)$$

and $\Omega = \Re^{n-1}$.

Example 3.7. Similarly, let $\boldsymbol{\mu}$ be a $1 \times j$ row vector and let $\boldsymbol{\Sigma}$ be a $j \times j$ positive definite matrix. Then the usual parameterization of the multivariate normal $\text{MVN}_j(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution has $\dim(\Theta) = j + j^2$ but is a $j + j(j+1)/2$ parameter REF.

A **curved exponential family** is a k-parameter exponential family where the elements of $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ are completely determined by $d < k$ of the elements. For example if $\boldsymbol{\theta} = (\theta, \theta^2)$ then the elements of $\boldsymbol{\theta}$ are completely determined by $\theta_1 = \theta$. A curved exponential family is neither full nor regular since it places a restriction on the parameter space Ω resulting in a new parameter space Ω_C where Ω_C does not contain a k-dimensional rectangle.

Example 3.8. The $N(\theta, \theta^2)$ distribution is a 2-parameter exponential family with $\eta_1 = -1/(2\theta^2)$ and $\eta_2 = 1/\theta$. Thus

$$\Omega_C = \{(\eta_1, \eta_2) | \eta_1 = -0.5\eta_2^2, -\infty < \eta_1 < 0, -\infty < \eta_2 < \infty, \eta_2 \neq 0\}.$$

The graph of this parameter space is a quadratic and cannot contain a 2-dimensional rectangle.

3.2 Properties of $(t_1(Y), ..., t_k(Y))$

This section follows Lehmann (1983, p. 29-35) closely. Write the *natural parameterization for the exponential family* as

$$f(y|\boldsymbol{\eta}) = h(y) b(\boldsymbol{\eta}) \exp\left[\sum_{i=1}^k \eta_i t_i(y)\right]$$

$$= h(y) \exp\left[\sum_{i=1}^k \eta_i t_i(y) \quad - \quad a(\boldsymbol{\eta})\right] \tag{3.10}$$

where $a(\boldsymbol{\eta}) = -\log(b(\boldsymbol{\eta}))$. The kernel function of this pdf or pmf is

$$h(y) \exp\left[\sum_{i=1}^k \eta_i t_i(y)\right].$$

Lemma 3.2. Suppose that Y comes from an exponential family (3.10) and that $g(y)$ is any function with $E_{\boldsymbol{\eta}}[|g(Y)|] < \infty$. Then for any $\boldsymbol{\eta}$ in the interior of Ω, the integral $\int g(y)f(y|\theta)dy$ is continuous and has derivatives of all orders. These derivatives can be obtained by interchanging the derivative and integral operators. If f is a pmf, replace the integral by a sum.

Proof. See Lehmann (1986, p. 59).

Hence
$$\frac{\partial}{\partial \eta_i} \int g(y) f(y|\boldsymbol{\eta}) dy = \int g(y) \frac{\partial}{\partial \eta_i} f(y|\boldsymbol{\eta}) dy \tag{3.11}$$
if f is a pdf and
$$\frac{\partial}{\partial \eta_i} \sum g(y) f(y|\boldsymbol{\eta}) = \sum g(y) \frac{\partial}{\partial \eta_i} f(y|\boldsymbol{\eta}) \tag{3.12}$$
if f is a pmf.

Remark 3.2. If $\boldsymbol{Y}$ comes from an exponential family (3.1), then the derivative and integral (or sum) operators can be interchanged. Hence
$$\frac{\partial}{\partial \theta_i} \int \cdots \int g(\boldsymbol{y}) f(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{y} = \int \cdots \int g(\boldsymbol{y}) \frac{\partial}{\partial \theta_i} f(\boldsymbol{y}|\boldsymbol{\theta}) d\boldsymbol{x}$$
for any function $g(\boldsymbol{y})$ with $E_\theta |g(\boldsymbol{Y})| < \infty$.

The behavior of $(t_1(Y), ..., t_k(Y))$ will be of considerable interest in later chapters. The following result is in Lehmann (1983, p. 29-30). Also see Johnson, Ladella, and Liu (1979).

Theorem 3.3. Suppose that Y comes from an exponential family (3.10). Then a)
$$E(t_i(Y)) = \frac{\partial}{\partial \eta_i} a(\boldsymbol{\eta}) = -\frac{\partial}{\partial \eta_i} \log(b(\boldsymbol{\eta})) \tag{3.13}$$
and b)
$$\mathrm{Cov}(t_i(Y), t_j(Y)) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} a(\boldsymbol{\eta}) = -\frac{\partial^2}{\partial \eta_i \partial \eta_j} \log(b(\boldsymbol{\eta})). \tag{3.14}$$
Notice that $i = j$ gives the formula for $\mathrm{VAR}(t_i(Y))$.

Proof. The proof will be for pdfs. For pmfs replace the integrals by sums. Use Lemma 3.2 with $g(y) = 1 \; \forall y$. a) Since $1 = \int f(y|\boldsymbol{\eta})dy$,
$$0 = \frac{\partial}{\partial \eta_i} 1 = \frac{\partial}{\partial \eta_i} \int h(y) \exp\left[\sum_{m=1}^k \eta_m t_m(y) \quad - \quad a(\boldsymbol{\eta})\right] dy$$

$$= \int h(y) \frac{\partial}{\partial \eta_i} \exp\left[\sum_{m=1}^k \eta_m t_m(y) \quad - \quad a(\boldsymbol{\eta})\right] dy$$

$$= \int h(y) \exp\left[\sum_{m=1}^k \eta_m t_m(y) \quad - \quad a(\boldsymbol{\eta})\right] (t_i(y) - \frac{\partial}{\partial \eta_i} a(\boldsymbol{\eta})) dy$$

$$= \int (t_i(y) - \frac{\partial}{\partial \eta_i} a(\boldsymbol{\eta})) f(y|\boldsymbol{\eta}) dy$$

$$= E(t_i(Y)) - \frac{\partial}{\partial \eta_i} a(\boldsymbol{\eta}).$$

b) Similarly,

$$0 = \int h(y) \frac{\partial^2}{\partial \eta_i \partial \eta_j} \exp\left[\sum_{m=1}^k \eta_m t_m(y) \quad - \quad a(\boldsymbol{\eta})\right] dy.$$

From the proof of a),

$$0 = \int h(y) \frac{\partial}{\partial \eta_j} \left[\exp\left[\sum_{m=1}^k \eta_m t_m(y) \quad - \quad a(\boldsymbol{\eta})\right] (t_i(y) - \frac{\partial}{\partial \eta_i} a(\boldsymbol{\eta}))\right] dy$$

$$= \int h(y) \exp\left[\sum_{m=1}^k \eta_m t_m(y) \quad - \quad a(\boldsymbol{\eta})\right] (t_i(y) - \frac{\partial}{\partial \eta_i} a(\boldsymbol{\eta}))(t_j(y) - \frac{\partial}{\partial \eta_j} a(\boldsymbol{\eta})) dy$$

$$- \int h(y) \exp\left[\sum_{m=1}^k \eta_m t_m(y) \quad - \quad a(\boldsymbol{\eta})\right] (\frac{\partial^2}{\partial \eta_i \partial \eta_j} a(\boldsymbol{\eta})) dy$$

$$= \mathrm{Cov}(t_i(Y), t_j(Y)) - \frac{\partial^2}{\partial \eta_i \partial \eta_j} a(\boldsymbol{\eta})$$

since $\frac{\partial}{\partial \eta_j} a(\boldsymbol{\eta}) = E(t_j(Y))$ by a). QED

Theorem 3.4. Suppose that Y comes from an exponential family (3.10), and let $\boldsymbol{T} = (t_1(Y), ..., t_k(Y))$. Then for any $\boldsymbol{\eta}$ in the interior of Ω, the moment generating function of $\boldsymbol{T}$ is

$$m_{\boldsymbol{T}}(\boldsymbol{s}) = \exp[a(\boldsymbol{\eta} + \boldsymbol{s}) - a(\boldsymbol{\eta})] = \exp[a(\boldsymbol{\eta} + \boldsymbol{s})] / \exp[a(\boldsymbol{\eta})].$$

Proof. The proof will be for pdfs. For pmfs replace the integrals by sums. Since $\boldsymbol{\eta}$ is in the interior of Ω there is a neighborhood of $\boldsymbol{\eta}$ such that if $\boldsymbol{s}$ is in that neighborhood, then $\boldsymbol{\eta} + \boldsymbol{s} \in \Omega$. (Hence there exists a $\delta > 0$ such that if $\|\boldsymbol{s}\| < \delta$, then $\boldsymbol{\eta} + \boldsymbol{s} \in \Omega$.) For such $\boldsymbol{s}$ (see Definition 2.25),

$$m_{\boldsymbol{T}}(\boldsymbol{s}) = E[\exp(\sum_{i=1}^k s_i t_i(Y))] \equiv E(g(Y)).$$

It is important to notice that we are finding the mgf of $\boldsymbol{T}$, not the mgf of Y. Hence we can use the kernel method of Section 1.5 to find $E(g(Y)) = \int g(y)f(y)dy$ without finding the joint distribution of $\boldsymbol{T}$. So

$$m_{\boldsymbol{T}}(\boldsymbol{s}) = \int \exp(\sum_{i=1}^k s_i t_i(y)) h(y) \exp\left[\sum_{i=1}^k \eta_i t_i(y) \;-\; a(\boldsymbol{\eta})\right] dy$$

$$= \int h(y) \exp\left[\sum_{i=1}^k (\eta_i + s_i) t_i(y) \;-\; a(\boldsymbol{\eta}+\boldsymbol{s}) \;+\; a(\boldsymbol{\eta}+\boldsymbol{s}) \;-\; a(\boldsymbol{\eta})\right] dy$$

$$= \exp[a(\boldsymbol{\eta}+\boldsymbol{s}) \;-\; a(\boldsymbol{\eta})] \int h(y) \exp\left[\sum_{i=1}^k (\eta_i + s_i) t_i(y) \;-\; a(\boldsymbol{\eta}+\boldsymbol{s})\right] dy$$

$$= \exp[a(\boldsymbol{\eta}+\boldsymbol{s}) \;-\; a(\boldsymbol{\eta})] \int f(y|[\boldsymbol{\eta}+\boldsymbol{s}])dy = \exp[a(\boldsymbol{\eta}+\boldsymbol{s}) \;-\; a(\boldsymbol{\eta})]$$

since the pdf $f(y|[\boldsymbol{\eta}+\boldsymbol{s}])$ integrates to one. QED

Theorem 3.5. Suppose that Y comes from an exponential family (3.10), and let $\boldsymbol{T} = (t_1(Y), ..., t_k(Y)) = (T_1, ..., T_k)$. Then the distribution of $\boldsymbol{T}$ is an exponential family with

$$f(\boldsymbol{t}|\boldsymbol{\eta}) = h^*(\boldsymbol{t}) \exp\left[\sum_{i=1}^k \eta_i t_i \;-\; a(\boldsymbol{\eta})\right].$$

Proof. See Lehmann (1986, p. 58).

The main point of this section is that $\boldsymbol{T}$ is well behaved even if Y is not. For example, if Y follows a one sided stable distribution, then Y is from an exponential family, but $E(Y)$ does not exist. However the mgf of T exists, so all moments of T exist. If $Y_1, ..., Y_n$ are iid from a one parameter exponential family, then $T \equiv T_n = \sum_{i=1}^n t(Y_i)$ is from a one parameter exponential family. One way to find the distribution function of T is to find the distribution of $t(Y)$ using the transformation method, then find the distribution of $\sum_{i=1}^n t(Y_i)$ using moment generating functions or Theorems 2.17 and 2.18. This technique results in the following two theorems. Notice that T often has a gamma distribution.

Theorem 3.6. Let $Y_1, ..., Y_n$ be iid from the given one parameter exponential family and let $T \equiv T_n = \sum_{i=1}^n t(Y_i)$.

a) If Y_i is from a binomial (k, ρ) distribution, then $t(Y) = Y \sim \text{BIN}(\text{k}, \rho)$ and $T_n = \sum_{i=1}^n Y_i \sim \text{BIN}(\text{nk}, \rho)$.

b) If Y is from an exponential (λ) distribution then, $t(Y) = Y \sim \text{EXP}(\lambda)$ and $T_n = \sum_{i=1}^n Y_i \sim G(n, \lambda)$.

c) If Y is from a gamma (ν, λ) distribution with ν known, then $t(Y) = Y \sim G(\nu, \lambda)$ and $T_n = \sum_{i=1}^n Y_i \sim G(n\nu, \lambda)$.

d) If Y is from a geometric (ρ) distribution, then $t(Y) = Y \sim \text{geom}(\rho)$ and $T_n = \sum_{i=1}^n Y_i \sim \text{NB}(\text{n}, \rho)$ where NB stands for negative binomial.

e) If Y is from a negative binomial (r, ρ) distribution with r known, then $t(Y) = Y \sim \text{NB}(\text{r}, \rho)$ and $T_n = \sum_{i=1}^n Y_i \sim \text{NB}(\text{nr}, \rho)$.

f) If Y is from a normal (μ, σ^2) distribution with σ^2 known, then $t(Y) = Y \sim N(\mu, \sigma^2)$ and $T_n = \sum_{i=1}^n Y_i \sim N(n\mu, n\sigma^2)$.

g) If Y is from a normal (μ, σ^2) distribution with μ known, then $t(Y) = (Y - \mu)^2 \sim G(1/2, 2\sigma^2)$ and $T_n = \sum_{i=1}^n (Y_i - \mu)^2 \sim G(n/2, 2\sigma^2)$.

h) If Y is from a Poisson (θ) distribution, then $t(Y) = Y \sim \text{POIS}(\theta)$ and $T_n = \sum_{i=1}^n Y_i \sim \text{POIS}(\text{n}\theta)$.

Theorem 3.7. Let $Y_1, ..., Y_n$ be iid from the given one parameter exponential family and let $T \equiv T_n = \sum_{i=1}^n t(Y_i)$.

a) If Y_i is from a Burr (ϕ, λ) distribution with ϕ known, then $t(Y) = \log(1 + Y^\phi) \sim \text{EXP}(\lambda)$ and $T_n = \sum \log(1 + Y_i^\phi) \sim G(n, \lambda)$.

b) If Y is from a chi(p, σ) distribution with p known, then $t(Y) = Y^2 \sim G(p/2, 2\sigma^2)$ and $T_n = \sum Y_i^2 \sim G(np/2, 2\sigma^2)$.

c) If Y is from a double exponential (θ, λ) distribution with θ known, then $t(Y) = |Y - \theta| \sim \text{EXP}(\lambda)$ and $T_n = \sum_{i=1}^n |Y_i - \theta| \sim G(n, \lambda)$.

d) If Y is from a two parameter exponential (θ, λ) distribution with θ known, then $t(Y) = Y_i - \theta \sim \text{EXP}(\lambda)$ and $T_n = \sum_{i=1}^n (Y_i - \theta) \sim G(n, \lambda)$.

e) If Y is from a generalized negative binomial GNB(μ, κ) distribution with κ known, then $T_n = \sum_{i=1}^n Y_i \sim GNB(n\mu, n\kappa)$

f) If Y is from a half normal (μ, σ^2) distribution with μ known, then $t(Y) = (Y - \mu)^2 \sim G(1/2, 2\sigma^2)$ and $T_n = \sum_{i=1}^n (Y_i - \mu)^2 \sim G(n/2, 2\sigma^2)$.

g) If Y is from an inverse Gaussian IG(θ, λ) distribution with λ known, then $T_n = \sum_{i=1}^n Y_i \sim IG(n\theta, n^2\lambda)$.

h) If Y is from an inverted gamma (ν, λ) distribution with ν known, then $t(Y) = 1/Y \sim G(\nu, \lambda)$ and $T_n = \sum_{i=1}^n 1/Y_i \sim G(n\nu, \lambda)$.

i) If Y is from a lognormal (μ, σ^2) distribution with μ known, then $t(Y) = (\log(Y) - \mu)^2 \sim G(1/2, 2\sigma^2)$ and $T_n = \sum_{i=1}^n (\log(Y_i) - \mu)^2 \sim G(n/2, 2\sigma^2)$.

j) If Y is from a lognormal (μ, σ^2) distribution with σ^2 known, then $t(Y) = \log(Y) \sim N(\mu, \sigma^2)$ and $T_n = \sum_{i=1}^n \log(Y_i) \sim N(n\mu, n\sigma^2)$.

k) If Y is from a Maxwell-Boltzmann (μ, σ) distribution with μ known, then $t(Y) = (Y - \mu)^2 \sim G(3/2, 2\sigma^2)$ and $T_n = \sum_{i=1}^n (Y_i - \mu)^2 \sim G(3n/2, 2\sigma^2)$.

l) If Y is from a one sided stable (σ) distribution, then $t(Y) = 1/Y \sim G(1/2, 2/\sigma)$ and $T_n = \sum_{i=1}^n 1/Y_i \sim G(n/2, 2/\sigma)$.

m) If Y is from a Pareto (σ, λ) distribution with σ known, then $t(Y) = \log(Y/\sigma) \sim \text{EXP}(\lambda)$ and $T_n = \sum_{i=1}^n \log(Y_i/\sigma) \sim G(n, \lambda)$.

n) If Y is from a power (λ) distribution, then $t(Y) = -\log(Y) \sim \text{EXP}(\lambda)$ and $T_n = \sum_{i=1}^n [-\log(Y_i)] \sim G(n, \lambda)$.

o) If Y is from a Rayleigh (μ, σ) distribution with μ known, then $t(Y) = (Y - \mu)^2 \sim \text{EXP}(2\sigma^2)$ and $T_n = \sum_{i=1}^n (Y_i - \mu)^2 \sim G(n, 2\sigma^2)$.

p) If Y is from a Topp-Leone (ν) distribution, then $t(Y) = -\log(2Y - Y^2) \sim \text{EXP}(1/\nu)$ and $T_n = \sum_{i=1}^n [-\ \log(2Y_i - Y_i^2)] \sim G(n, 1/\nu)$.

q) If Y is from a truncated extreme value (λ) distribution, then $t(Y) = e^Y - 1 \sim \text{EXP}(\lambda)$ and $T_n = \sum_{i=1}^n (e^{Y_i} - 1) \sim G(n, \lambda)$.

r) If Y is from a Weibull (ϕ, λ) distribution with ϕ known, then $t(Y) = Y^\phi \sim \text{EXP}(\lambda)$ and $T_n = \sum_{i=1}^n Y_i^\phi \sim G(n, \lambda)$.

3.3 Complements

Example 3.9. Following Barndorff–Nielsen (1978, p. 117), if Y has an inverse Gaussian distribution, $Y \sim \text{IG}(\theta, \lambda)$, then the pdf of Y is

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left[\frac{-\lambda(y-\theta)^2}{2\theta^2 y}\right]$$

where $y, \theta, \lambda > 0$.

Notice that

$$f(y) = \sqrt{\frac{\lambda}{2\pi}} e^{\lambda/\theta} \sqrt{\frac{1}{y^3}} I(y > 0) \exp\left[\frac{-\lambda}{2\theta^2} y - \frac{\lambda}{2} \frac{1}{y}\right]$$

is a two parameter exponential family.

Another parameterization of the inverse Gaussian distribution takes $\theta = \sqrt{\lambda/\psi}$ so that

$$f(y) = \sqrt{\frac{\lambda}{2\pi}} e^{\sqrt{\lambda\psi}} \sqrt{\frac{1}{y^3}} I[y > 0] \exp\left[\frac{-\psi}{2} y - \frac{\lambda}{2}\frac{1}{y}\right],$$

where $\lambda > 0$ and $\psi \geq 0$. Here $\Theta = (0, \infty) \times [0, \infty)$, $\eta_1 = -\psi/2$, $\eta_2 = -\lambda/2$ and $\Omega = (-\infty, 0] \times (-\infty, 0)$. Since Ω is not an open set, this is a **2 parameter full exponential family that is not regular**. If ψ is known then Y is a 1P–REF, but if λ is known then Y is a one parameter full exponential family. When $\psi = 0$, Y has a one sided stable distribution.

The following chapters show that exponential families can be used to simplify the theory of sufficiency, MLEs, UMVUEs, UMP tests and large sample theory. Barndorff-Nielsen (1982) and Olive (2005) are useful introductions to exponential families. Also see Bühler and Sehr (1987). Interesting subclasses of exponential families are given by Rahman and Gupta (1993) and Sankaran and Gupta (2005). Most statistical inference texts at the same level as this text also cover exponential families. History and references for additional topics (such as finding conjugate priors in Bayesian statistics) can be found in Lehmann (1983, p. 70), Brown (1986) and Barndorff-Nielsen (1978, 1982).

Barndorff-Nielsen (1982), Brown (1986) and Johanson (1979) are post–PhD treatments and hence very difficult. Mukhopadhyay (2000) and Brown (1986) place restrictions on the exponential families that make their theory less useful. For example, Brown (1986) covers linear exponential distributions. See Johnson and Kotz (1972).

3.4 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

Refer to Chapter 10 for the pdf or pmf of the distributions in the problems below.

3.1*. Show that each of the following families is a 1P–REF by writing the pdf or pmf as a one parameter exponential family, finding $\eta = w(\theta)$ and by showing that the natural parameter space Ω is an open interval.

a) The binomial (k, ρ) distribution with k known and $\rho \in \Theta = (0, 1)$.
b) The exponential (λ) distribution with $\lambda \in \Theta = (0, \infty)$.
c) The Poisson (θ) distribution with $\theta \in \Theta = (0, \infty)$.
d) The half normal (μ, σ^2) distribution with μ known and $\sigma^2 \in \Theta = (0, \infty)$.

3.2*. Show that each of the following families is a 2P–REF by writing the pdf or pmf as a two parameter exponential family, finding $\eta_i = w_i(\theta)$ for $i = 1, 2$ and by showing that the natural parameter space Ω is a cross product of two open intervals.
a) The beta (δ, ν) distribution with $\Theta = (0, \infty) \times (0, \infty)$.
b) The chi (p, σ) distribution with $\Theta = (0, \infty) \times (0, \infty)$.
c) The gamma (ν, λ) distribution with $\Theta = (0, \infty) \times (0, \infty)$.
d) The lognormal (μ, σ^2) distribution with $\Theta = (-\infty, \infty) \times (0, \infty)$.
e) The normal (μ, σ^2) distribution with $\Theta = (-\infty, \infty) \times (0, \infty)$.

3.3. Show that each of the following families is a 1P–REF by writing the pdf or pmf as a one parameter exponential family, finding $\eta = w(\theta)$ and by showing that the natural parameter space Ω is an open interval.
a) The generalized negative binomial (μ, κ) distribution if κ is known.
b) The geometric (ρ) distribution.
c) The logarithmic (θ) distribution.
d) The negative binomial (r, ρ) distribution if r is known.
e) The one sided stable (σ) distribution.
f) The power (λ) distribution.
g) The truncated extreme value (λ) distribution.
h) The Zeta (ν) distribution.

3.4. Show that each of the following families is a 1P–REF by writing the pdf or pmf as a one parameter exponential family, finding $\eta = w(\theta)$ and by showing that the natural parameter space Ω is an open interval.
a) The $N(\mu, \sigma^2)$ family with $\sigma > 0$ known.
b) The $N(\mu, \sigma^2)$ family with μ known and $\sigma > 0$.
c) The gamma (ν, λ) family with ν known.
d) The gamma (ν, λ) family with λ known.
e) The beta (δ, ν) distribution with δ known.
f) The beta (δ, ν) distribution with ν known.

3.5. Show that each of the following families is a 1P–REF by writing the pdf or pmf as a one parameter exponential family, finding $\eta = w(\theta)$ and by showing that the natural parameter space Ω is an open interval.

a) The Burr (ϕ, λ) distribution with ϕ known.
b) The double exponential (θ, λ) distribution with θ known.
c) The two parameter exponential (θ, λ) distribution with θ known.
d) The largest extreme value (θ, σ) distribution if σ is known.
e) The smallest extreme value (θ, σ) distribution if σ is known.
f) The inverted gamma (ν, λ) distribution if ν is known.
g) The Maxwell–Boltzmann (μ, σ) distribution if μ is known.
h) The Pareto (σ, λ) distribution if σ is known.
i) The Rayleigh (μ, σ) distribution if μ is known.
j) The Weibull (ϕ, λ) distribution if ϕ is known.

3.6*. Determine whether the Pareto (σ, λ) distribution is an exponential family or not.

3.7. Following Kotz and van Dorp (2004, p. 35-36), if Y has a Topp–Leone distribution, $Y \sim TL(\nu)$, then the cdf of Y is $F(y) = (2y - y^2)^\nu$ for $\nu > 0$ and $0 < y < 1$. The pdf of Y is

$$f(y) = \nu(2 - 2y)(2y - y^2)^{\nu - 1}$$

for $0 < y < 1$. Determine whether this distribution is an exponential family or not.

3.8. In Spiegel (1975, p. 210), Y has pdf

$$f_Y(y) = \frac{2\gamma^{3/2}}{\sqrt{\pi}}\ y^2\ \exp(-\gamma\ y^2)$$

where $\gamma > 0$ and y is real. Is Y a 1P-REF?

3.9. Let Y be a (one sided) truncated exponential $TEXP(\lambda, b)$ random variable. Then the pdf of Y is

$$f_Y(y|\lambda, b) = \frac{\frac{1}{\lambda} e^{-y/\lambda}}{1 - \exp(-\ \frac{b}{\lambda})}$$

for $0 < y \le b$ where $\lambda > 0$. If b is known, is Y a 1P-REF? (Also see O'Reilly and Rueda (2007).)

Problems from old quizzes and exams.

3.10*. Suppose that X has a $\text{N}(\mu, \sigma^2)$ distribution where $\sigma > 0$ and μ **is known**. Then

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\mu^2/(2\sigma^2)} \exp[-\frac{1}{2\sigma^2}x^2 + \frac{1}{\sigma^2}\mu x].$$

Let $\eta_1 = -1/(2\sigma^2)$ and $\eta_2 = 1/\sigma^2$. Why is this parameterization not the regular exponential family parameterization? (Hint: show that η_1 and η_2 satisfy a linearity constraint.)

3.11. Let $X_1, ..., X_n$ be iid $N(\mu, \gamma_o^2\mu^2)$ random variables where $\gamma_o^2 > 0$ is **known** and $\mu > 0$.

a) Find the distribution of $\sum_{i=1}^n X_i$.

b) Find $E[(\sum_{i=1}^n X_i)^2]$.

c) The pdf of X is

$$f_X(x|\mu) = \frac{1}{\gamma_o \mu \sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\gamma_o^2\mu^2}\right].$$

Show that the family $\{f(x|\mu) : \mu > 0\}$ is a two parameter exponential family.

d) Show that the natural parameter space is a parabola. You may assume that $\eta_i = w_i(\mu)$. Is this family a regular exponential family?

3.12. Let $X_1, ..., X_n$ be iid $N(\alpha\sigma, \sigma^2)$ random variables where α is a **known** real number and $\sigma > 0$.

a) Find $E[\sum_{i=1}^n X_i^2]$.

b) Find $E[(\sum_{i=1}^n X_i)^2]$.

c) Show that the family $\{f(x|\sigma) : \sigma > 0\}$ is a two parameter exponential family.

d) Show that the natural parameter space Ω is a parabola. You may assume that $\eta_i = w_i(\sigma)$. Is this family a regular exponential family?