same as for X. In particular, there is a measure $\nu_{\mathcal{T}}$ such that $P_{\theta,T} \ll \nu_{\mathcal{T}}$ for all θ and $dP_{\theta,T}/d\nu_{\mathcal{T}}(t) = c(\theta)\exp(\theta^\top t)$.

PROOF. Apply Lemma 2.24 with $m_2(T(x),\theta) = c(\theta)\exp(\sum_{i=1}^k \theta_i t_i(x))$ and $m_1(x) = h(x)$. □

Example 2.59 (Continuation of Example 2.56; see page 103). In the case of n conditionally IID $N(\mu,\sigma^2)$ random variables, the natural sufficient statistics are $T_1 = \sum_{i=1}^n X_i$ and $T_2 = \sum_{i=1}^n X_i^2$. It is well known that T_1 and $W = T_2 - T_1^2/n$ are independent, with T_1 having $N(n\mu, n\sigma^2)$ distribution and W having $\Gamma([n-1]/2, 1/[2\sigma^2])$ distribution. It follows that the joint density of (T_1,T_2) is

$$\left[\sqrt{\pi}\sigma^n\Gamma\left(\frac{n-1}{2}\right)2^n\right]^{-1}\left(t_2-\frac{t_1^2}{n}\right)^{\frac{n-1}{2}}\exp\left(-\frac{1}{2\sigma^2}\left[\frac{(t_1-n\mu)^2}{n}+t_2-\frac{t_1^2}{n}\right]\right).$$

This can be simplified to $c(\theta)h(t_1,t_2)\exp(\theta_1 t_1+\theta_2 t_2)$, with $c(\theta)$ the same as in Example 2.56 and $h(t_1,t_2)$ a constant times $(t_2 - t_1^2/n)^{[n-1]/2}$.

There are degenerate exponential families. That is, it is possible for some linear function of X to be constant (the same constant for all θ) with probability 1 given $\Theta = \theta$ for all θ. For example, let Y_1 be k_1-dimensional and have conditional density given $\Theta = \theta$ (with respect to a measure ν_1) $c(\theta)\exp(y_1^\top\theta)$. Define $\Gamma^\top = (\Theta^\top, \Psi^\top)$, where Ψ is k_2-dimensional. Let ν_2 be the measure that puts a mass of 1 on the point $r \in \mathbb{R}^{k_2}$ and puts 0 mass on the rest of $\mathbb{R}^{k_2}$. Define $\nu = \nu_1 \times \nu_2$, and let $Y^\top = (Y_1^\top, r^\top)$. Then Y has conditional density given $\Gamma = (\theta,\psi) = \gamma$ with respect to ν,

$$c^*(\gamma)\exp(y^\top\gamma) = c(\theta)\exp(y_1^\top\theta),$$

where $c^*(\gamma) = c(\theta)\exp(-r^\top\psi)$. The natural parameter space of Γ values is $\Omega \times \mathbb{R}^{k_2}$, where Ω is the natural parameter space of Θ values. For this reason, we introduce a definition.

Definition 2.60. An exponential family of distributions for X is *degenerate* if there exists at least one vector α and a scalar r such that $P_\theta(\alpha^\top X = r) = 1$ for all θ. If the exponential family is not degenerate, it is called *nondegenerate.*

Example 2.61. Let $X \sim Mult_k(n; p_1,\ldots,p_k)$ given $P = (p_1,\ldots,p_k)$. The natural parameter is $\Theta = (\log P_1,\ldots,\log P_k)$. We know that $P_\theta(\mathbf{1}^\top X = n) = 1$ for all θ, where $\mathbf{1}$ is a vector of k 1s.

When X has a degenerate exponential family distribution, then there exists a linear function $X' = AX$ of smaller dimension such that X' has nondegenerate exponential family distribution and X is still a function of X' with probability 1. The natural parameter space of a degenerate exponential family typically will contain distinct vectors that produce the same distribution for X. For instance, in Example 2.61, the natural parameter space is $\mathbb{R}^k$. But for every real a, $\theta = (a,\ldots,a)$ produces the same distribution for X, $Mult_k(n; 1/k,\ldots,1/k)$. For this reason, we shall assume that all exponential families are nondegerate in the remainder of this text.