Beginning of Jan.21, 2022

We can also rewrite the Gaussian familty has a two parameter exponential family *in canonical form*:

$$w_1(\theta) = \frac{\mu}{\sigma^2} \qquad \text{and} \qquad w_2(\theta) = -\frac{1}{2\sigma^2}$$

so we try to rewrite $a(w)$ in terms of w_1, w_2 by

$$\begin{aligned} a(w) = \frac{\mu^2}{2\sigma^2} + \log\sigma &= -\left(\frac{\mu}{\sigma^2}\right)^2 \cdot \left(-\frac{1}{2\sigma^2}\right)^{-1} - \frac{1}{2}\log\left(-2 \cdot \frac{-1}{2\sigma^2}\right) \\ &= -\frac{w_1^2}{4w_2} - \frac{\log(-2w_2)}{2}. \end{aligned}$$

Originally we had the restriction $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, so this is equivalent to the constraint $\{(w_1, w_2) \in \mathbb{R}^2 : w_2 < 0\}$.

Example: (3.4) Location Family. Let X be a random variable with continuous density $f : \mathbb{R} \to [0, \infty)$. Let $\mu \in \mathbb{R}$. Then the densities $\{f(x+\mu)\}_{\mu\in\mathbb{R}}$ is called the **location family** of X. This may *or may not* be an exponential family.

An example: Gaussian densities with a fixed variance — shifting the pdf simply results in a new Gaussian pdf with shifted mean and same variance.

A non-example: if X is uniform on $[0,1]$ then the location family $1_{[-\mu,1-\mu]}$ do not form an exponential family.

Example: (3.6) Scale Family. Let X be a random variable. The densities $\{\sigma^{-1}f(x/\sigma)\}_{\sigma>0}$ are called the **scale family** of X. (Divide by $1/\sigma$ because we need to ensure the integral is 1.) This family may *or may not* be an exponential family.

Example: (3.7) Location and Scale Family. Combining the two examples above, $\{\sigma^{-1}f((x+\mu)/\sigma)\}$ is caled the **location and scale family** of X. Again, this may *or may not* be an exponential family.

0.1 Differential Identities

Sometimes exponential families make certain computations easier. One obvious example is via differentiation.

Let X be a standard Gaussian. Then its moment generating function (MGF) is

$$\mathbb{E}e^{tX} = e^{t^2/2} \qquad \text{for al } t \in \mathbb{R}.$$

Using this we have

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\Big|_{t=0} \mathbb{E}e^{tX} = \mathbb{E}X^m,$$

so for example

$$\mathbb{E}X^2 = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big|_{t=0} e^{t^2/2} = 1.$$

We can do similar things for exponential families. If

$$a(w) = \log \int_{\mathbb{R}^n} h(x) \exp\Big(\sum_{i=1}^k w_i t_i(x)\Big)\, \mathrm{d}\mu(x),$$

and let W be the natural parameter space (i.e., where $a(w) < \infty$), then we claim that

Lemma: (3.8)

$a(w)$ is continuous and has continuous partial derivatives on the interior of W (i.e. where $a(\cdot)$ is finite). Moreover, the derivative can be obtained by differentiating under the integral sign.

Proof. We prove the existence of first order partial derivative with respect to w_1 and the rest follows by iteration. Let $e_1 := (1, 0, \ldots, 0) \in \mathbb{R}^k$. Exponential is analytic so it suffices to show that $\exp(a(w))$ has continuous partial derivative along e_1. The difference quotient is

$$\begin{aligned}\frac{\exp(a(w+\epsilon e_1)) - \exp(a(w))}{\epsilon} &= \frac{1}{\epsilon}\int_{\mathbb{R}^n} h(x)\left[\exp\left(\epsilon t_1(x) + \sum_{i=1}^k w_i t_i(x)\right) - \exp\left(\sum_{i=1}^k w_i t_i(x)\right)\right] \mathrm{d}\mu(x) \\ &= \int_{\mathbb{R}^n} h(x)\frac{\exp(\epsilon t_1(x)) - 1}{\epsilon} \exp\left(w_i t_i(x)\right) \mathrm{d}\mu(x).\end{aligned}$$

By the MVT, for any $\alpha \in (0,1)$ and for all $\beta \in \mathbb{R}$,

$$|e^{\alpha\beta - 1}| \leqslant |\alpha\beta| e^{|\beta|} \leqslant |\alpha| e^{2|\beta|} \leqslant |\alpha|(e^{2\beta} + e^{-2\beta}). \tag{*}$$

Beginning of Jan.24, 2022

Therefore, for $\delta > 0$, $\alpha := \epsilon/\delta$ and $\beta := \delta t_1(x)$,

$$\begin{aligned}\left| h(x)\frac{\exp(\epsilon t_1(x)) - 1}{\epsilon} \exp\left(\sum_{i=1}^k w_i t_i(x)\right)\right| &\leqslant h(x)\left|\frac{\exp(\epsilon t_1(x)) - 1}{\epsilon}\right| \exp\left(\sum_{i=1}^k w_i t_i(x)\right) &(1) \\ &\leqslant \frac{1}{\delta} h(x)[\exp(2\delta t_1(x) + \exp(-2\delta t_1(x))] \exp\left(\sum_{i=1}^k w_i t_i(x)\right). &(2)\end{aligned}$$

Note that we have gotten rid of the dependence of ϵ.
If we define $X_\epsilon :=$ the LHS of (1) and $Y :=$ (2), then $|X_\epsilon| \leqslant Y$ for $0 < \epsilon < \delta < 1$.Letting $\epsilon \to 0$ and using DCT,

$$\begin{aligned}\frac{\partial}{\partial w_1}\exp(a(w)) &= \lim_{\epsilon \to 0}\int_{\mathbb{R}^n}\left| h(x)\frac{\exp(\epsilon t_1(x)) - 1}{\epsilon}\exp\left(\sum_{i=1}^k w_i t_i(x)\right)\right| \mathrm{d}\mu(x) \\ &= \int_{\mathbb{R}^n}\lim_{\epsilon\to 0} h(x)\left|\frac{\exp(\epsilon t_1(x)) - 1}{\epsilon}\exp\left(\sum_{i=1}^k w_i t_i(x)\right)\right| \mathrm{d}\mu(x) \\ &= \int_{\mathbb{R}^n} h(x) t_1(x)\exp\left(\sum_{i=1}^k w_i t_i(x)\right) \mathrm{d}\mu(x),\end{aligned}$$

where the dominance of an integrable function is given by the fact that w is in the interior of W, so there exists $\delta > 0$ such that

$$a(w + 2\delta e_1) < \infty \qquad \text{and} \qquad a(w - 2\delta e_1) < \infty.$$ □

Remark. We can rewrite the above formula, using definition of $e^{-a(w)}$, as

$$\exp(-a(w))\frac{\partial}{\partial w_1}\exp(a(w)) = \int_{\mathbb{R}^n} t_1(x) h(x)\exp\left(\sum_{i=1}^k w_i t_i(x) - a(w)\right) \mathrm{d}\mu(x) = \int_{\mathbb{R}^n} t_1(x) f_w(x)\, \mathrm{d}\mu(x).$$

That is, differentiating $a(w)$ gives moment information for the exponential family $\{f_w(x)\}$.
Since $f_w(x)$ can be thought of as a PDF with respect to the measure μ, i.e. $\int_{\mathbb{R}^n} t_i f_w(x)\, \mathrm{d}\mu(x) = 1$, for

convenience we define

$$\mathbb{E}_\theta t_i := \int_{\mathbb{R}^n} t_i f_w(x)\,\mathrm{d}\mu(x).$$

Remark. We proved the lemma for canonical exponential families. For non-canonical exponential families, a similar argument holds:

$$e^{-a(w(\theta))}\frac{\partial}{\partial\theta_1}e^{a(w(\theta))} = e^{-a(w(\theta))}\sum_{i=1}^{k}\frac{\partial e^{a(w)}}{\partial w_i}\frac{\partial w_i}{\partial\theta_1} = \sum_{i=1}^{k}\frac{\partial w_i}{\partial\theta_1}\mathbb{E}_\theta t_i = \mathbb{E}_\theta\Big(\sum_{i=1}^{k}\frac{\partial w_i}{\partial\theta_1}t_i\Big).$$

Example: (3.13) Gaussian revisited. Recall that, for Gaussians with $\mu\in\mathbb{R}$ and $\sigma^2>0$, we have $k=2, n=1$, and we defined $\theta=(\theta_1,\theta_2):=(\mu,\sigma^2)\in\mathbb{R}^2$, $t_1(x):=x$, $t_2(x):=x^2$,

$$w_1(\theta):=\frac{\theta_1}{\theta_2}=\frac{\mu}{\sigma^2},\qquad w_2(\theta):=-\frac{1}{2\theta_2}=-\frac{1}{2\sigma^2},$$

and finally

$$a(w(\theta)):=\frac{\theta_1^2}{2\theta_2}+\frac{\log\theta_2}{2}=\frac{\mu^2}{2\sigma^2}+\log\sigma.$$

Then,

$$\begin{aligned}e^{-a(w(\theta))}\frac{\partial}{\partial\theta_1}e^{a(w(\theta))} &= e^{-a(w(\theta))}\frac{\mathrm{d}}{\mathrm{d}\theta_1}\exp\left[\frac{\theta_1^2}{2\theta_2}+\frac{\log\theta_2}{2}\right]\\ &= (2\theta_1)/(2\theta_2)=\mu/\sigma^2,\end{aligned}$$

whereas the previous remark gives

$$\mathbb{E}_\theta\Big(\sum_{i=1}^{2}\frac{\partial w_i}{\partial\theta_1}t_i\Big)=\mathbb{E}_\theta\Big(\frac{\partial w_1}{\partial\theta_1}t_1+0\Big)\mathbb{E}_\theta(x/\theta_2)=\mathbb{E}_\theta(x)/\sigma^2.$$

That is,

$$\mathbb{E}_\theta(x)/\sigma^2=\mu/\sigma^2 \implies \mathbb{E}_\theta(x)=\mu.$$

In totality, we've shown that ***expected value of a Gaussian with mean μ is indeed μ!***

Remark. We can take *more* derivatives of $a(w(\theta))$ and obtain more moment information.