Notes on Information Geometry

Charles Zheng

August 24, 2015

1 Definitions

The α-connection

$$\Gamma^{(\alpha)}_{ij,k}(\theta) = \mathbf{E}_\theta[((\partial_i \partial_j \ell_\theta) - (\partial_i \ell_\theta)(\partial_j \ell_\theta))\partial_k \ell_\theta]$$

2 Exponential families

2.1 Definitions

$$p_\theta(x) = C(x)\exp[\Sigma_i \theta_i t_i(x) - \psi(\theta)]$$

$$\ell_\theta(x) = \log p_\theta(x) = \log C(x) + \sum_i \theta_i t_i(x) - \psi(\theta)$$

$$\psi(\theta) = \log \int_x C(x) e^{\theta^T t(x)} dx$$

$$\eta_i(\theta) = \mathbf{E}_\theta[t_i(X)]$$

$$g_{ij}(\theta) = \mathrm{Cov}_\theta[t_i(X), t_j(X)] = g_{ji}$$

2.2 Properties under natural parameterization

$$\partial_i \triangleq \frac{\partial}{\partial \theta_i}$$

$$\partial_i \psi(\theta) = \eta_i(\theta)$$

$$\partial_i \partial_j \psi(\theta) = g_{ij}(\theta)$$

$$\partial_i \ell(x) = t_i(x) - \eta_i$$

Note that $\partial_i \partial_j \ell(x)$ is constant as a function of x.

$$\partial_i \partial_j \ell(x) = -g_{ij}$$

$$\partial_i \partial_j \partial_k \ell(x) = -\partial_k g_{ij} = -\mathbb{E}(\partial_i \ell)(\partial_j \ell)(\partial_k \ell) \stackrel{D}{=} -T_{ijk}$$

$$T_{ijk} = \partial_k g_{ij} = T_{ikj} = \cdots = T_{kji}$$

2.3 Natural parameterization is e-affine

$$\Gamma^{(1)}_{ij,k}(\theta) = \mathbf{E}_\theta[(\partial_i \partial_j \ell)(\partial_k \ell)] = -g_{ij}\mathbf{E}[\partial_k \ell] = 0$$

2.4 Properties under mean (η) parameterization

$$\frac{\partial \eta_i}{\partial \theta_j} = \partial_j(\partial_i \psi) = g_{ij}$$

$$\tilde{\partial}_i \triangleq \frac{\partial}{\partial \eta_i} = \sum_j \frac{\partial \theta_j}{\partial \eta_i} \partial_j = \sum_j \frac{\partial_j}{g_{ij}}$$

2.4.1 Properties of T_{ijk}

We have

$$\frac{\partial^2 \theta_k}{\partial \eta_i \partial \eta_j} = \tilde{\partial}_j \frac{1}{g_{ik}} = \tilde{\partial}_i \frac{1}{g_{jk}}$$

therefore

$$\sum_m \frac{1}{g_{jm}} \frac{T_{ikm}}{g_{ik}^2} = \sum_m \frac{1}{g_{im}} \frac{T_{jkm}}{g_{jk}^2}$$

Hence, if we define

$$C_{ijk} \triangleq \frac{1}{g_{ik}^2} \sum_m \frac{T_{ikm}}{g_{jm}} = -\frac{\partial^2 \theta_k}{\partial \eta_i \partial \eta_j}$$

then we have

$$C_{ijk} = C_{jik}$$

by definition, and from symmetry of T_{ijk} we have

$$C_{ijk} = C_{kji}$$

hence $C_{ijk} = C_{kji} = C_{jki} = \cdots$, i.e. is symmetric with respect to indices.

2.4.2 Derivatives

$$\tilde{\partial}_i \ell = \sum_k \frac{\partial_k \ell}{g_{ik}}$$

$$\tilde{\partial}_i \tilde{\partial}_j \ell = -\left(\sum_k C_{ijk} \partial_k \ell\right) - \left(\sum_{k,m} \frac{g_{km}}{g_{ik} g_{jm}}\right)$$

2.4.3 Mean parameterization is m-affine

$$\begin{aligned}
\Gamma^{(-1)}_{ij,k}(\eta) &= \mathbf{E}_\eta[(\tilde{\partial}_i \tilde{\partial}_j \ell)(\tilde{\partial}_k \ell) + (\tilde{\partial}_i \ell)(\tilde{\partial}_j \ell)(\tilde{\partial}_k \ell)] \\
&= \mathbf{E}\left[\left(-\left(\sum_k C_{ijk} \partial_k \ell + \sum_{k,m} \frac{g_{km}}{g_{ik} g_{jm}}\right) + (\tilde{\partial}_i \ell)(\tilde{\partial}_j \ell)\right)(\tilde{\partial}_k \ell)\right]
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}[(\tilde{\partial}_i \ell)(\tilde{\partial}_j \ell)(\tilde{\partial}_k \ell)] &= \sum_{a,b,c} \frac{\mathbf{E}[(\partial_a \ell)(\partial_b \ell)(\partial_c \ell)]}{g_{ia} g_{jb} g_{kc}} \\
&= \sum_{a,b,c} \frac{T_{abc}}{g_{ia} g_{jb} g_{kc}}
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}_\eta[(\tilde{\partial}_i \tilde{\partial}_j \ell)(\tilde{\partial}_k \ell)] &= \mathbf{E}\left[\left(-\sum_k C_{ijk} \partial_k \ell - \sum_{k,m} \frac{g_{km}}{g_{ik} g_{jm}}\right)(\tilde{\partial}_k \ell)\right] \\
&= -\mathbf{E}\left[\left(\sum_a C_{ija} \partial_a \ell\right)\left(\sum_b \frac{\partial_b \ell}{g_{kb}}\right)\right] \\
&= -\sum_{a,b} \frac{C_{ija}}{g_{kb}} \mathbf{E}[(\partial_a \ell)(\partial_b \ell)] = -\sum_{a,b} \frac{C_{ija} g_{ab}}{g_{kb}}
\end{aligned}$$