Detailed Local KL Geometry

The local KL geometry of statistical manifolds deviates from euclidean geometry in two ways: its "metric" has *superquadratic* terms, most notably cubic (asymmetry) and quartic (barrier); and its quadratic component induces a *curved* connection. More precisely, we have a divergence $D : \mathcal{M} \times \mathcal{M} \to [0, \infty]$ that vanishes on and only on the diagonal; D is smooth on a neighborhood of this diagonal; on this diagonal, the hessian is invariant with respect to interchange of the two $\mathcal{M}$ factors and otherwise generic. With $v = (a, b)$, we thus write:

$$D(p + a : p + b) = (1/2)H_p(v, v) + (1/6)J_p(v, v, v) + (1/24)Q_p(v, v, v, v) + o(v^{\otimes 4})$$

Here, H is positive definite and H, J, Q — as well as the little o constant — vary smoothly with p. The situation has a hidden symmetry under interchange of v with $\bar{v} = (b, a)$. For example, $H(v, v) = H(\bar{v}, \bar{v})$. To see how this arises, let us focus on a submanifold $\mathcal{M} \subseteq \Delta^o(X)$ of the open simplex on a finite set X. Since $\log(1 + z) = z - z^2/2 + z^3/3 - z^4/4 + \cdots$ and $1/(p + b) = (1/p)(1 - (b/p) + (b/p)^2 - (b/p)^3 + \cdots)$, we have

$$\begin{aligned}
&D(p + a, p + b) \\
&= \sum (p + a) \log\left(1 + \frac{a - b}{p + b}\right) \\
&= \sum (p + a) \left(\left(\frac{a - b}{p + b}\right) - \frac{1}{2}\left(\frac{a - b}{p + b}\right)^2 + \frac{1}{3}\left(\frac{a - b}{p + b}\right)^3 - \frac{1}{4}\left(\frac{a - b}{p + b}\right)^4\right) \\
&= \sum p(1 + \alpha) \left(\delta(1 - \beta + \beta^2 - \beta^3) - \frac{1}{2}\delta^2(1 - 2\beta + 3\beta^2) + \frac{1}{3}\delta^3(1 - 3\beta) - \frac{1}{4}\delta^4\right)
\end{aligned}$$

Here, $\alpha = a/p$, $\beta = b/p$, $\gamma = \alpha + \beta$, $\delta = \alpha - \beta$. We expand to:

$$\begin{aligned}
= \sum p\delta(1 + \alpha)((1) + (-\beta - \delta/2) \\
+ \left(\beta^2 + 2\delta\beta/2 + \delta^2/3\right) \\
+ \left(-\beta^3 - 3\delta\beta^2/2 - 3\delta^2\beta/3 - \delta^3/4\right))
\end{aligned}$$

The degree one sum vanishes by normalization, as expected. The degree two term has the familiar χ^2 form:

$$p\delta(-\beta - \delta/2 + \alpha) = p\delta^2/2$$

The degree three term witnesses asymmetry:

$$p\delta(\beta^2 + 2\delta\beta/2 + \delta^2/3 - \alpha\beta - \alpha\delta/2) = p(-(5/12)\delta^3 + \gamma\delta^2/4)$$

The degree four term is:

$$p\delta(-\beta^3 - 3\delta\beta^2/2 - 3\delta^2\beta/3 - \delta^3/4 + \alpha\beta^2 + 2\alpha\delta\beta/2 + \alpha\delta^2/3) = p(\delta^4/4 - (5/6)\alpha^2\delta^2 - 2\alpha\beta^2\delta)$$

We now explore the basics of this geometry. We especially examine how the new geometry distorts our euclidean picture of a large-N sample as a tight Gaussian on an inner product space.