By MAKRAM TALIH

October 8, 2006

Likelihood for geodesics started at the identity matrix

Consider the geodesic curve $t \mapsto P(t)$ emanating from $P(0) = I \in \mathcal{P}$ with initial velocity vector $\dot{P}(0) = A \in \mathcal{T}_{\mathcal{P}}(I) \equiv \mathcal{S}$, and let $K_t = P(t) = e^{tA}$. Suppose one has n IID observations $X_1, \cdots, X_n$ from the multivariate Normal distribution $N_d(0, K_t^{-1})$, which has precision matrix K_t. Then, for the sample variance-covariance matrix $\mathbb{S} = n^{-1}\mathbb{X}\mathbb{X}^T$ where $\mathbb{X}$ is the $d \times n$ matrix whose column vectors are the X_k, we have that, up to an additive constant, twice the log-likelihood is given by

$$\begin{aligned} 2\ell(t, A) &= n \ln \det(e^{tA}) - n \operatorname{tr}(\mathbb{S}e^{tA}) \\ &= nt \operatorname{tr}(A) - n \operatorname{tr}(\mathbb{S}e^{tA}), \quad \text{since } \det(e^{tA}) = e^{t \operatorname{tr}(A)}. \end{aligned}$$

For values of t near 0, our goal is to examine how the likelihood varies as a function of the initial velocity vector $\dot{P}(0)$. The hope is to find "good" initial velocity vectors in that they are optimal in the vicinity of $t = 0\cdot$

As a function of A, we have

$$\begin{aligned} 2\frac{\partial \ell(t, A)}{\partial A} &= ntI - n \sum_{l=0}^{\infty} \frac{t^l}{l!} \frac{\partial}{\partial A} \operatorname{tr}(\mathbb{S}A^l) \\ &= ntI - n \sum_{l=0}^{\infty} \frac{t^l}{l!} \sum_{k=0}^{l-1} A^k \mathbb{S} A^{l-k-1} \\ &= ntI - n \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{t^{l+k+1}}{(l+k+1)!} A^k \mathbb{S} A^l\cdot \end{aligned}$$

In the above, we have used:

$$(\partial/\partial A) \operatorname{tr}(QA^l) = \sum_{k=0}^{l-1} A^k Q A^{l-k-1}\cdot$$

Since we hope to find an optimal direction in the vicinity of $t = 0$, let's look at derivatives of (twice) the score function in A with respect to the parameter t. The first derivative with respect to t is given by

$$2\frac{\partial^2 \ell(t,A)}{\partial t \partial A} = nI - n\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{t^{l+k}}{(l+k)!}A^k \mathbb{S} A^l \cdot$$

At $t = 0$, we get

$$2\frac{\partial^2 \ell(t,A)}{\partial t \partial A}\mid_{t=0} = n(I - \mathbb{S})\cdot$$

Next, notice that all derivatives of order $2m, m = 1, 2, \cdots$ in t yield 0 when evaluated at $t = 0$, while derivatives of order $2m+1, m = 1, 2, \cdots$ in t are given by

$$2\frac{\partial^{2m+2} \ell(t,A)}{\partial t^{2m+1} \partial A} = -n\sum_{k=m}\sum_{l=m}\frac{t^{l+k-2m}}{(l+k-2m)!}A^k \mathbb{S} A^l \cdot$$

Therefore, at $t = 0$, we get

$$2\frac{\partial^{2m+2} \ell(t,A)}{\partial t^{2m+1} \partial A}\mid_{t=0} = -nA^m \mathbb{S} A^m,$$

for $m = 1, 2, \cdots$.

The score function is well behaved in the Gaussian case, so we can certainly appeal to a Taylor series expansion around $t = 0$ to obtain

$$2\frac{\partial \ell(t,A)}{\partial A} - 2\frac{\partial \ell(0,A)}{\partial A} = nt(I - \mathbb{S}) - n\sum_{m=1}^{\infty}\frac{t^{2m+1}}{(2m+1)!}A^m \mathbb{S} A^m \cdot$$

Since $2\frac{\partial \ell(0,A)}{\partial A} = 0$, we obtain

$$2\frac{\partial \ell(t,A)}{\partial A} = ntI - n\sum_{m=0}^{\infty}\frac{t^{2m+1}}{(2m+1)!}A^m \mathbb{S} A^m \cdot$$

Now, setting $(\partial/\partial A)\ell(t,A) = 0$ for t near 0 gives

$$tI = \sum_{m=0}^{\infty}\frac{t^{2m+1}}{(2m+1)!}A^m \mathbb{S} A^m \cdot \tag{1}$$

While I don't see how to solve this matrix equation in A directly, I can work on characterizing the solution. For instance, I can multiply by A on both sides of (1) and take the trace:

$$\begin{aligned} t\,\mathrm{tr}(A) &= \sum_{m=0}^{\infty} \frac{t^{2m+1}}{(2m+1)!}\,\mathrm{tr}(A^m \mathbb{S} A^{m+1}) \\ &= \sum_{m=0}^{\infty} \frac{t^{2m+1}}{(2m+1)!}\,\mathrm{tr}(A^{2m+1}\mathbb{S}) \\ &= \mathrm{tr}\Big(\sum_{m=0}^{\infty} \frac{1}{(2m+1)!}(tA)^{2m+1}\mathbb{S}\Big) \\ &= \mathrm{tr}(\sinh(tA)\mathbb{S}), \end{aligned}$$

where the hyperbolic matrix sine $\sinh(tA)$ is formally defined for matrices via the power series expansion in a manner similar to the definition of the matrix exponential. Dividing by t and letting $t \to 0$ yields

$$\mathrm{tr}(A) = \mathrm{tr}(\cosh(tA)\mid_{t=0} \mathbb{S}) = \mathrm{tr}(I\mathbb{S}) = \mathrm{tr}(\mathbb{S})\cdot$$

Well, my guess is that for values of t near zero, the likelihood maximizing direction is simply $\hat{A} = \mathbb{S} = n^{-1}\mathbb{X}\mathbb{X}^T\cdot$ Let's verify equation (1) with this choice of A for values of t near zero:

$$t^{-1}\sum_{m=0}^{\infty} \frac{t^{2m+1}}{(2m+1)!}\mathbb{S}^{2m+1} = t^{-1}\sinh(t\mathbb{S}) \to \cosh(t\mathbb{S})\mid_{t=0} = I$$

as $t \to 0$, as required.

Thus, the resulting geodesic curve $t \mapsto P(t)$ with initial position $P(0) = I$ and velocity vector $\dot{P}(0) = \hat{A}$ has velocity vector at zero given by $\dot{P}(0) = \mathbb{S}$.

General case

In general, we would start from $P(0) = K \in \mathcal{P}$, and ride along the geodesic curve $t \mapsto P(t)$ emanating from K with initial velocity vector $\dot{P}(0) = K^{1/2}AK^{1/2} \in \mathcal{T}_{\mathcal{P}}(K) \equiv \mathcal{S}$, and let

$$K_t = P(t) = K^{1/2}e^{tA}K^{1/2}\cdot$$

Suppose one has n IID observations $X_1, \cdots, X_n$ from the multivariate Normal distribution $N_d(0, K_t^{-1})$, which has precision matrix K_t. Then one makes the change of variables $\mathbb{Y} = K^{1/2}\mathbb{X}$, thus working directly with the n IID observations $Y_1, \cdots, Y_n$ from the multivariate Normal distribution $N_d(0, K^{1/2}K_t^{-1}K^{1/2})$, which has precision matrix $K^{-1/2}K_tK^{-1/2} = e^{tA}$. Then, the sample variance-covariance matrix is given by $\mathbb{S}' = n^{-1}\mathbb{Y}\mathbb{Y}^T$, and we proceed as in the previous section whence, up to an additive constant, twice the log-likelihood given K is

$$2\ell(t, A) = n \ln \det(e^{tA}) - n \operatorname{tr}(\mathbb{S}'e^{tA}) = nt \operatorname{tr}(A) - n \operatorname{tr}(\mathbb{S}'e^{tA})\cdot$$

Just as before, we can proceed setting $(\partial/\partial A)\ell(t, A) = 0$ for t near 0 to obtain

$$tI = \sum_{m=0}^{\infty} \frac{t^{2m+1}}{(2m+1)!} A^m \mathbb{S}' A^m, \tag{2}$$

Which, for t near zero, has its solution

$$\hat{A} = \mathbb{S}' \stackrel{\text{here}}{=} K^{1/2}\mathbb{S}K^{1/2}$$

for $\mathbb{S} = n^{-1}\mathbb{X}\mathbb{X}^T\cdot$

Note that the resulting geodesic curve $t \mapsto P(t)$ with $P(0) = K$ and $\dot{P}(0) = K^{1/2}\hat{A}K^{1/2}$ now has velocity vector at zero given by $\dot{P}(0) = K\mathbb{S}K$, which accounts explicitely for the fact that the starting point was an arbitrary $K \in \mathcal{P}$ and not I in general.