Product Space Decompositions for Continuous Representations of Brain Connectivity: Supplemental Notes

***

***

1 Table of notation

Var.	Type	Description
N	Integer	Number of subjects. Fixed.
n	Index	Subject index.
C_n	Integer	Number of clusters for subject n. Not fixed.
c	Index	Cluster index.
K	Integer	Number of prototypes. Not fixed.
k	Index	prototypes index.
i	Index	Streamline index. Fixed per subject.
t_i^n	Index	Cluster assignment indicator for streamline i and subject n, taking values in $\{1,\ldots,C\}$
g_c^n	Index	Prototype assignment indicator for cluster c and subject n, taking values in $\{1,\ldots,K\}$
η_c^n	Integer	Number of streamlines in subject n assigned cluster c.
η_k	Integer	Number of clusters associated with prototype k
C_*	Integer	Total number of clusters
μ	Pair of vectors	Mean parameter for tangent Gaussians. Actually a pair of means, one for each tangent Gaussian.
Σ	Pair of 2×2 matrices	Variance parameter for tangent Gaussians. Actually a pair of matrices, one for each tangent Gaussian.
Δ	2×2 matrix	Prior parameter on Σ
ν	Integer	Prior parameter on Σ (degrees of freedom)
α	Scalar	Concentration parameter for clusters.
γ	Scalar	Concentration parameter for prototypes.

Table 1. Notation.

2 Marginal Likelihood of a new cluster

We would like to calculate the probability $p(x|\mathit{\Delta},\nu)$ for only one data-point x_i, marginalizing out $\mathit{\Sigma}$ and μ. That is,

$$\begin{aligned}
p(x|\mathit{\Delta},\nu) &= \int_S \int_{M^{2\times 2}} p(x|\mathit{\Sigma},\mu)p(\mathit{\Sigma}|\mathit{\Delta})p(\mu)d\mathit{\Sigma}d\mu \\
&= \int_S \int_{M^{2\times 2}} \mathcal{N}(\mathrm{Log}(x;\mu)|0,\mathit{\Sigma})\mathrm{IW}(\mathit{\Delta}+A,\nu+1)d\mathit{\Sigma}d\mu \\
&= \underbrace{Const}_{\text{Conj. Prior}} \; |Sphere|^{-1} \int_S |\mathit{\Delta}+A|^{-\frac{\nu+1}{2}} d\mu
\end{aligned}$$

Let $\mathit{\Delta}$ be I_d and ν be 3. If A is the scatter matrix dd^T (with $d = x_{tan} - \mu$):

$$\begin{aligned}
|\mathit{\Delta}+A| &= \underbrace{(x_1x_1+1)}_{a}\underbrace{(x_2x_2+1)}_{d} - \underbrace{(x_1x_2)}_{b}\underbrace{(x_2x_1)}_{c} \\
&= (x_1^2+x_2^2+1) + (x_1^2x_2^2 - x_1^2x_2^2) \\
&= 1 + d^T d
\end{aligned}$$

Further consider any rotation of x_{tan}, Rx_{tan}. Then:

$$|\mathit{\Delta}+A'| = 1 + x^T R^T R x = 1 + x^T x$$

Thus, $|\mathit{\Delta}+A'|$ is rotationally invariant to x about μ. Rewriting our objective using spherical coordinates, we have:

$$\begin{aligned}
p(x|\mathit{\Delta},\nu) &= \underbrace{Const}_{\text{Conj. Prior}} \; |Sphere|^{-1} \int_S |\mathit{\Delta}+A|^{-\frac{\nu+1}{2}} d\mu \\
&= C \int_0^{\pi} 2\pi |I_d + \mathrm{Log}(x(\theta);\mu)\,\mathrm{Log}(x(\theta);\mu)^T|^{-2} \sin\theta d\theta \\
\mathrm{Log}(x(\theta);\mu) &= (x - \mu\cos\theta)\frac{\theta}{\sin\theta}
\end{aligned}$$

$$
\begin{aligned}
p(x|\Delta,\nu) &= 2\pi C \int_0^\pi \left(1 + \frac{\theta^2}{\sin^2\theta}(x - \mu\cos\theta)\cdot(x-\mu\cos\theta)\right)^{-2} \sin\theta d\theta \\
&= 2\pi C \int_0^\pi \left(1 + \frac{\theta^2}{\sin^2\theta}(x\cdot x - 2x\cdot\mu\cos\theta + \mu\cdot\mu)\right)^{-2} \sin\theta d\theta \\
&= 2\pi C \int_0^\pi \left(1 + \frac{\theta^2}{\sin^2\theta}(2 - 2\cos^2\theta)\right)^{-2} \sin\theta d\theta \\
&= 2\pi C \int_0^\pi \left(1 + \frac{2\theta^2}{\sin^2\theta}(\sin^2\theta)\right)^{-2} \sin\theta d\theta \\
&= 2\pi C \int_0^\pi \left(1 + 2\theta^2\right)^{-2} \sin\theta d\theta \\
Const &= \frac{|\Delta|^{\frac{\nu}{2}}\Gamma_p(\frac{\nu+n}{2})}{\pi^{\frac{np}{2}}\Gamma_p(\frac{\nu}{2})} = \frac{\Gamma_p(2)}{\pi\Gamma_p(\frac{3}{2})} = \frac{\Gamma(2)\Gamma(\frac{3}{2})}{\pi\Gamma(\frac{3}{2})\Gamma(1)} = \frac{2}{\pi}
\end{aligned}
$$

The resulting integral is easily numerically integrated, opposed to the initial integral. This is **not** a conjugate prior, and, to our knowledge, there isn't such a prior for tangent Gaussians. This same process can, in fact, be done for other choices of ν, though the numerical stability may vary. Other choices of Δ are also possible, though they must be diagonal. Note that both x and μ are constrained to be on the surface of the sphere. We should also note that the reason this isn't simply 1 is because we also have neglected to integrate over x. Finally, note that this is symmetric in μ and x.

3 Implementation Notes

By far the largest amount of computational time is spent updating the t_i^n. This is partially because of our data; there are a large number of tracts per subject, leading to a large number of evaluations of the tangent Gaussian densities. However, except for the actual assignment step, the evaluation of the densities can be done fully in parallel, even across subjects, distributed across different processors. In the event that a new cluster is drawn when assigning clusters (which must be done serially), we can propagate this change and evaluate each point for the new density. This approach is, in our experience, much faster, though it requires $O(C_*NS)$ memory, where S is the number of streamlines.