Lab notebook: DDPM object tracking

Carl Smith

April 11, 2012 - May 29, 2012

Notes

Tuesday, May 29, 2012

A thought: you can construct transition kernels that encourage cyclic motion, like the following:

$$p(\theta_t|\theta_{t-1}) \propto \sum_i e^{\kappa \mu_i^T \mathbf{h}_{t-1}} e^{\kappa \mu_{i+1}^T \mathbf{h}_t}$$

where the $\{\mu_i\}$ are organized in, e.g., a circular path around the sphere.

Okay, so I think I have found a way to construct stationary distributions whose modes have different concentration parameters. The trick is that we also have to change the weights of the modes just so. Say that our stationary distribution is

$$\mathbb{G}_0(\theta) \propto \sum_i \alpha_i e^{\kappa_i \mu^T \mathbf{h}}$$

where to be concrete we constrain the $\{\alpha_i\}$ to sum to one. And suppose we have the same transition kernel from Equation (1). Then, following Equation (2), we require for stationarity that

$$\sum_j \alpha_j \frac{||\kappa_i \mu_i + \kappa_j \mu_j||}{\sinh(||\kappa_i \mu_i + \kappa_j \mu_j||)} = \alpha_i \quad \forall i$$

In fact, in this case we're free to choose the directions $\{\mu_i\}$ however we please, so long as we use the same $\{\mu_i\}$ for the transition kernel and so long as we vary the $\{\kappa_i\}$ and $\{\alpha_i\}$ appropriately. (Another question is, can we also allow for using different $\{\kappa_i\}$ for the stationary distribution and for the transition kernel?) As a concrete example, if we choose a bimodal distribution with $\mu_2 = -\mu_1$ and $\kappa_2 = 2\kappa_1$ so that one mode is opposite the other with twice the precision, then $\alpha_1 = 0.9203$ and $\alpha_2 = 0.3913$. That is, the more precise mode must be given less weight, in this case.

Wednesday, May 9, 2012

First off, Equation (2) may be correct after all, so long as the sum over j is the same for each i. Such is the case, as already mentioned, if the modes lie on the vertices of a Platonic solid inscribed in the sphere.

Met with Frank today. He wants to figure out a way to vary the κ's in the von Mises-von Mises transition kernel and/or stationary distribution. Also he'd like to see samples from the generative model. First I'll have to figure out how to sample from a multivariate von Mises distribution.

I've found `randvonMisesFisherm.m` by Sungkyu Jung, which samples from the von Mises-Fisher distribution of arbitrary dimension.

For our stationary distribution, we're constrained to keep the modes on the sphere symmetric. That is, rotating the sphere in any way that leaves the positions of the (indistinguishable) modes the same must be an identity element. For instance, we can inscribe any Platonic solid in the sphere and put modes at its vertices. Then we get a distribution that looks like Figure 1 for 10,000 samples with $\kappa = 1$, and like Figure 2 for 10,000 samples with $\kappa = 10$, where in the latter the segregation into modes is obvious, but the former distribution looks pretty uniform to me. These are for the modes being placed on the vertices of an inscribed octahedron, i.e. simply on the axes. Figures 3, 4, and 5 show random walks via the octahedral transition kernel with varying κ and 10 time steps. As expected, for higher κ the walker is effectively trapped in one mode. These were all hand picked so that most of the density was on the side of the sphere facing the viewer.

In other news, it is fine for the stationary distribution to have a different κ from that of the transition kernel, i.e.

$$\mathbb{G}_0(\theta) \propto \sum_i e^{\kappa_s \mu_i^T \mathbf{h}}$$

where as usual $\mathbf{h}$ is the unit vector in 3-space corresponding to the angle(s) θ, while

$$P(\theta_{t+1}|\theta_t) = \sum_{z_t}^R p(z_t|\theta_t)p(\theta_{t+1}|z_t)$$

$$\propto \sum_{z_t}^R e^{\kappa_t \mu_{z_t}^T \mathbf{h_{t+1}}} e^{\kappa_t \mu_{z_t}^T \mathbf{h_t}}. \tag{1}$$

But, different κ's for the different modes? I don't think there's a stationary distribution in that case.
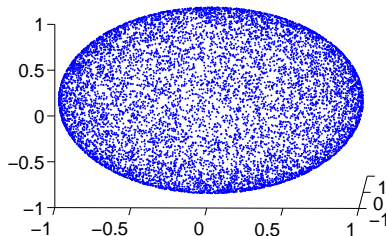

Figure 1: 10,000 draws on the octahedral von Mises sphere with $\kappa = 1$.

Monday, April 30, 2012

It seems to be important, in coming up with a stationary distribution $\mathbb{G}_0$ to have a uniform grid of points on the (hyper-)sphere. One thing I've come across in the 3D case is inscribing an icosahedron and taking the vertices to be the grid points. Unfortunately this rather fixes the resolution of the smoothing, so that's not the most desirable. It would be preferable to somehow write down a stationary distribution for the normal latitude-longitude grid.
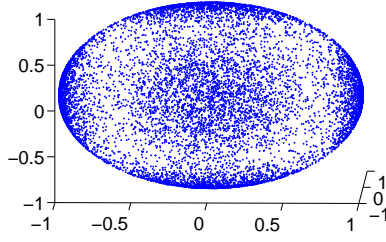
Figure 2: 10,000 draws on the octahedral von Mises sphere with $\kappa = 10$.
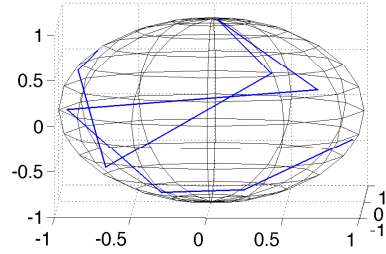

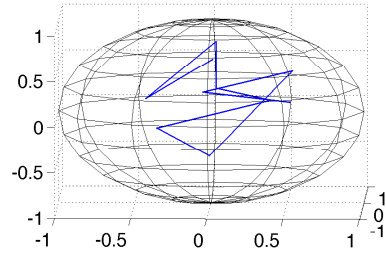Figure 3: 10 steps of random walk on octahedral von Mises sphere with $\kappa = 1$.

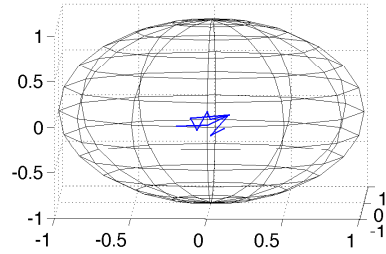Figure 4: 10 steps of random walk on octahedral von Mises sphere with $\kappa = 10$.

Figure 5: 10 steps of random walk on octahedral von Mises sphere with $\kappa = 100$.

Maybe, and I'm not sure why I ruled this out before, the stationary density is simply uniform on the sphere:

$$\int d\theta_{t-1} \mathbb{G}_0(\theta_{t-1}) \sum_i p(\mu_i|\theta_{t-1})p(\theta_t|\mu_i)$$

$$\propto \int d\theta_{t-1} \sum_i e^{\kappa \mu_i^T \mathbf{h}_t} e^{\kappa \mu_i^T \mathbf{h}_{t-1}}$$

Nope.

Thursday, April 26, 2012

On the unit circle, the von Mises distribution can be parametrized by the "mean" angle ϕ where the mode of the density resides:

$$f(x|\phi,\kappa) = \frac{e^{\kappa \cos(x-\phi)}}{2\pi I_0(\kappa)}$$

where $I_0(x)$ is the order 0 modified Bessel function. (Assume throughout that κ is fixed.) Instead, however, consider the unit vector $\mathbf{x}$ from the circle's center to the point on the perimeter an angle x away from the x-axis. Similarly define μ to be the vector with tail at the circle's center and point on the perimeter making the angle ϕ with the x-axis. This we may write

$$f(\mathbf{x}|\mu,\kappa) = \frac{e^{\kappa \mu^T \mathbf{x}}}{2\pi I_0(\kappa)}.$$

In the same vein, for higher dimensional $\mathbf{x}$ we have the von Mises-Fisher distribution:

$$f_p(\mathbf{x};\mu,\kappa) = C_p(\kappa)e^{\kappa \mu^T \mathbf{x}}$$

where $C_p(\kappa)$ is a normalization constant that involves Bessel functions of κ. ($p = 2$ for the simple von Mises distribution on the circle.)

In (Neiswanger and Wood, 2012) we have that we need to choose $\mathbb{G}_0(\theta)$, $p(\theta|z)$, and $p(z|\theta)$ such that

$$\int \mathbb{G}_0(\theta_{t-1})P(\theta_t|\theta_{t-1})d\theta_{t-1}$$

$$= \int \mathbb{G}_0(\theta_{t-1}) \sum_z P(\theta_t|z_t)P(z_t|\theta_{t-1})d\theta_{t-1}$$

$$= \mathbb{G}_0(\theta_t)$$

To this end, choose the following transition kernel: (Note that θ denotes some cluster parameter(s) – the *angle* in the distribution. Denote by $\mathbf{h}$ the corresponding unit vector in the p-dimensional space.)

$$p(\theta_t|\theta_{t-1}) \propto \sum_i e^{\kappa \mu_i^T \mathbf{h_t}} e^{\kappa \mu_i^T \mathbf{h_{t-1}}}$$

4

where the $\{\mu_i\}$ form a grid in angle-space. (Imagine a point on a globe wherever a latitude and longitude line intersect. This obviously results in anisotropic coverage of the sphere. We'll return to this point below.)

Then the inner products to be computed are of the form

$$\int d\theta \; e^{\kappa(\mu_i+\mu_j)^T \mathbf{h}} = C_p\left(\kappa \cdot ||\mu_i + \mu_j||\right)$$

The stationary distribution $\mathbb{G}_0$ puts equal density on each bump:

$$\mathbb{G}_0(\theta_t) = \sum_i e^{\kappa \mu_i^T \mathbf{h}_t}$$

$$\int d\theta_{t-1} \mathbb{G}_0(\theta_{t-1}) \sum_i p(\mu_i|\theta_{t-1})p(\theta_t|\mu_i)$$

$$\propto \int d\theta_{t-1} \sum_{ij} e^{\kappa \mu_i^T \mathbf{h}_t} e^{\kappa \mu_j^T \mathbf{h}_{t-1}} e^{\kappa \mu_i^T \mathbf{h}_{t-1}}$$

$$\propto \sum_{ij} C_p\left(\kappa \cdot ||\mu_i + \mu_j||\right) e^{\kappa \mu_i^T \mathbf{h}_t}$$

$$\propto \sum_i e^{\kappa \mu_i^T \mathbf{h}_t} = \mathbb{G}_0(\theta_t) \tag{2}$$

That last step follows because for each i, the range of values taken by $||\mu_i + \mu_j||$ is the same, and so the entire $\sum_j C_p$ coefficient is a common factor.

Wednesday, April 11, 2012

Frank thinks the low-rank models ideas can straightforwardly be applied to his work with an undergraduate described in (Neiswanger and Wood, 2012). I met with him March 8 and here are some notes from that meeting and an email I sent him:

These cluster parameters $\theta_{k,t}$ change over time. And the cluster parameters change smoothly; adjacent parameters are coupled using auxiliary variables $z_{k,t,m}$. So we have in mind this graphical model like in Figure 6. Actually these particular distributions aren't important. What is important is that $\theta_{k,t} \approx \theta_{k,t-1}$ $\forall k$. But there are a couple GPUDDP caveats. We need that $\theta_{k,t}$ can be analytically integrated out given observations: we need to be able to compute $\int d\theta_{k,t} p(x|\theta_{k,t})G_0(\theta_{k,t}|\theta)$, where $G_0(\theta)$ is the stationary distribution of $p(\theta_{t+1}|\theta_t)$. The question is, can you meet this requirement for arbitrary $F(\theta_{k,t}) = p(x_{k,t+1}|\theta_{k,t})$? for F from an exponential family? for F normal?

What are these parameters exactly? Is $\theta_{k,t}$ the mean and variance of the Gaussian or do we assume unit variance or something? And what are the parameters of the normal inverse Wishart? The normal-inverse-Wishart is parametrized in terms of hyperparamters $(\mu_0, \Lambda_0/\kappa_0; \nu_0, \Lambda)$:

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1})$$
$$\mu|\Sigma \sim \mathcal{N}(\mu_0, \Sigma/\kappa_0)$$

So how do the z's and θ's relate to these? *Update: see (Neiswanger and Wood, 2012) for all these details.*
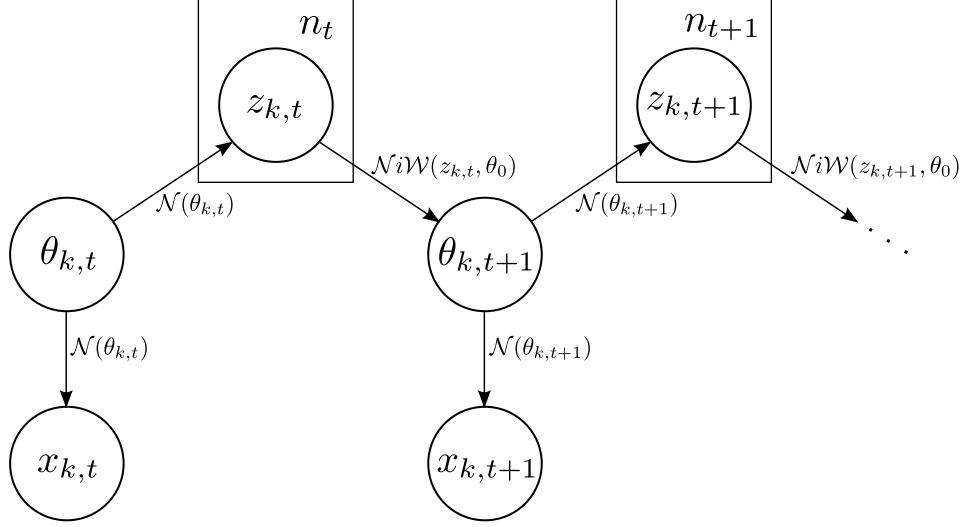
Figure 6: Cluster parameters $\theta_{k,t}$ vary smoothly over time, enforced by the $z_{k,t,m}$'s.

References

Neiswanger, W. and Wood, F. (2012). Unsupervised detection and tracking of arbitrary objects with dependent Dirichlet process mixtures. Forthcoming.