Research statement

My research as a PhD student has focused on methodology for multivariate data, probability distributions on manifolds, Bayesian computation, and informative prior specification. I am also interested in inference with intractable likelihoods, robustness to model misspecification, and applications in the environmental, health, and social sciences. Below I describe completed and ongoing research projects and lay out possible future directions.

Bayesian analysis with orthogonal matrix parameters

Statistical models for multivariate data are often naturally parametrized by the set of orthogonal matrices $\mathcal{V}(k,p) = \{\boldsymbol{Q} \in \mathbb{R}^{p\times k} \mid \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}\}$, referred to as the Stiefel manifold. For example, we might assume a covariance matrix $\boldsymbol{\Sigma}$ has the structure $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T + \sigma^2\boldsymbol{I}$, i.e. that it is the sum of a low-rank component, represented by its eigendecomposition, and a scaled identity matrix. Parametrization in terms of orthogonal matrices is most common in models based on matrix or tensor decompositions and other models which assume low dimensional structure underlying high dimensional data. Bayesian analyses of these models can be prohibitively difficult due to two major obstacles: prior specification and posterior simulation. In each case, the challenge relates to constrained nature of the parameter space. The following three projects with my advisers David Dunson and Peter Hoff were motivated by these obstacles and contribute methodology to overcome them. Their practical impact is to allow for flexible prior specification and routine posterior simulation for models with orthogonal matrix parameters, even in statistical software such as Stan [3].

Random orthogonal matrices and the Cayley transform The Cayley transform is a function which parametrizes the Stiefel manifold in terms of Euclidean elements [13]. In Jauch et al. [6], we draw on tools from geometric measure theory to describe how to map a probability distribution defined on the Stiefel manifold to Euclidean space using the inverse Cayley transform. From a practical point of view, this gives us a three-step scheme for simulating from a distribution on the Stiefel manifold: transform the distribution to Euclidean space, use familiar Markov chain Monte Carlo (MCMC) techniques such Hamiltonian Monte Carlo (HMC) [12] as implemented in Stan, then map the resulting Markov chain back to the Stiefel manifold with the Cayley transform. However, in my view, the most interesting parts of this project are the more theoretical contributions. Building upon recent results in the probability literature [8], we prove a theorem which states (roughly) that the pushforward measure of a tall and skinny, uniformly-distributed orthogonal matrix by the inverse Cayley transform can be approximated by independent normals. We also establish that the Grassmann manifold – the set of k-dimensional subspaces of $\mathbb{R}^p$ – can be represented by a subset of orthogonal matrices and that this subset can also be parametrized using the Cayley transform. We then give a change of variables result which allows for simulation from distributions on the Grassmann manifold.

An auxiliary variable approach to simulation This project develops an MCMC approach to simulating from probability distributions on the Stiefel manifold which can be routinely implemented in software such as Stan, scales to problems of realistic size, and has strong theoretical support. Suppose we want to simulate from the distribution of the random $p \times k$ orthogonal matrix $\boldsymbol{Q}$. We imagine introducing a $k \times k$ symmetric positive definite matrix $\boldsymbol{S}$

so that the joint distribution of $\boldsymbol{Q}$ and $\boldsymbol{S}$ admits a density. Using the Jacobian associated with the polar decomposition, we can then write down the density of the matrix $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{S}^{1/2}$ which is supported on $\mathbb{R}^{p\times k}$. We simulate from the distribution of $\boldsymbol{X}$ using familiar MCMC techniques and then transform the realized Markov chain back to the Stiefel manifold via the polar decomposition. This procedure yields a sample from the target distribution for $\boldsymbol{Q}$. But how do we choose the conditional distribution of $\boldsymbol{S}$ given $\boldsymbol{Q}$? We propose one choice which leads to simple $\boldsymbol{X}$ distributions for familiar $\boldsymbol{Q}$ distributions. Drawing on results of Durmus et al. [4], we show that HMC chains which target the resulting distribution for $\boldsymbol{X}$ are geometrically ergodic for a wide class of $\boldsymbol{Q}$ distributions. We illustrate the practical performance of our approach by fitting Bayesian models for a protein interaction network and gene expression data.

Marginal matching priors Defining probability distributions on the Stiefel manifold which reflect our prior information is not always straightforward. In models for multivariate data, for example, one commonly expects that an orthogonal matrix parameter is nearly sparse, i.e. its entries are mostly near zero with a small fraction of relatively large values. Yet incorporating this prior information into our analyses presents technical challenges. Motivated by this example, we describe an approach to constructing prior distributions for orthogonal matrices having prescribed element-wise marginals. If we have prior information that our orthogonal matrix is nearly sparse, we choose these marginals to match conventional sparsity inducing priors – up to rescaling. Our construction allows for posterior simulation via Stan.

Optimal shrinkage estimation with nuisance coefficients

In a linear regression setting, substantive questions often pertain to a subset of regression coefficients, while the complementary subset might be thought of as nuisance parameters. For example, when analyzing a laboratory experiment designed to assess the relationship between particular covariates and a response, we may also have information indicating which observations were part of the same experimental batch. The regression coefficients associated with the batches are not likely to be of scientific interest, but the batch information must be taken into account when estimating the target coefficients. How, if at all, should these nuisance parameters enter into our loss function for estimating the coefficients of interest? In ongoing work with Peter Hoff, we propose a loss function, compare it to alternatives, and seek a minimax estimator with respect to this loss by minimizing Stein's unbiased risk estimate within a family of Bayes estimators.

Bayesian optimization with shape constraints

Bayesian optimization is a framework for optimization of black box functions which are costly to evaluate. In typical applications of Bayesian optimization, very little is assumed about the function being optimized. Instead, it is assigned a generic Gaussian process (GP) prior. In Jauch and Peña [7], we make the case that prior shape constraints are often appropriate in two important application areas of Bayesian optimization: hyperparameter tuning of machine learning algorithms and decision analysis with utility functions. We describe methodology for imposing (approximate) shape constraints through a GP prior and present results from simple applications which indicate the benefits of this approach.

Future directions

Inference with an intractable likelihoods Modern statistical applications often give rise to intractable likelihood functions or posterior distributions. For example, intractable likelihoods present problems in models for directional data, exponential random graph models for networks, and non-Gaussian Markov random field models in spatial statistics. As our appetite for complex statistical analyses grows, intractable likelihoods loom large. A number of creative ideas have appeared in response, e.g. MCMC approaches [10], inference based on homogeneous scoring rules [5], and minimum Wasserstein distance estimation [1]. I would like to further explore these ideas and their applications.

Robustness to model misspecification Likelihood-based inference requires one to specify a probability distribution for the data, if not a joint distribution for the data and parameters, and inferences are often sensitive to even minor misspecification. As we take on increasingly complex analyses, modeling the true data generating process becomes increasingly error-prone and cumbersome. Furthermore, a great deal of energy is spent modeling aspects of the data generating process which are incidental to the question of interest. I am interested in developing methodology which addresses these issues in important applied settings. Several approaches seem promising. The first, which has a long history, strategically discards some information in the data for the sake of simplicity and robustness. The second achieves a degree of robustness by conditioning on an alternative event [9]. The third proposes to connect parameters to observations through a loss function rather than a likelihood and develops a coherent framework for updating prior beliefs in this setting [2].

Informative prior specification A much touted advantage of the Bayesian paradigm is that it provides a coherent framework for incorporating prior information into our analyses. However, even when our prior information is easy to formulate, integrating that information into our inferences can present technical challenges. Identifying settings where prior information is available but ignored and addressing the related technical challenges is a common theme of my projects on Bayesian optimization and priors for orthogonal matrices. In my view, more work is needed in this direction. For example, I would like to apply the prior specification strategy of Neal [11] to the regression setting in which one has many observations on a small number of covariates and a small number of observations on a large set of covariates. In cases where imputation is impractical, we can specify an informative prior for the larger regression based on the posterior predictive distribution of the smaller regression.

References

[1] Bernton, E., P. E. Jacob, M. Gerber, and C. P. Robert (2017). Inference in generative models using the Wasserstein distance. *arXiv:1701.05146*.

[2] Bissiri, P. G., C. C. Holmes, and S. G. Walker (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 78*(5), 1103–1130.

[3] Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software 76*(1), 1–32.

[4] Durmus, A., E. Moulines, and E. Saksman (2017). On the convergence of Hamiltonian Monte Carlo. *arxiv:1705.00166*.

[5] Hyvarinen, A. (2005). Estimation of Non-Normalized Statistical Models by Score Matching. *Journal of Machine Learning Research 6*, 695–708.

[6] Jauch, M., P. D. Hoff, and D. B. Dunson (2018). Random orthogonal matrices and the Cayley transform. *arXiv:1810.02881*.

[7] Jauch, M. and V. Peña (2016). Bayesian Optimization with Shape Constraints. *NIPS Workshop on Bayesian Optimization*.

[8] Jiang, T. (2006). How many entries of a typical orthogonal matrix can be approximated by independent normals? *Annals of Probability 34*(4), 1497–1529.

[9] Miller, J. W. and D. B. Dunson (2018). Robust Bayesian Inference via Coarsening. *Journal of the American Statistical Association*.

[10] Murray, I., Z. Ghahramani, and D. J. C. MacKay (2006). MCMC for doubly-intractable distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pp. 359–366.

[11] Neal, R. M. (2001). Transferring Prior Information Between Models Using Imaginary Data. Technical report, Department of Statistics, University of Toronto.

[12] Neal, R. M. (2010). MCMC Using Hamiltonian Dynamics. *Handbook of Markov Chain Monte Carlo 54*, 113–162.

[13] Shepard, R., S. R. Brozell, and G. Gidofalvi (2015). The Representation and Parametrization of Orthogonal Matrices. *The Journal of Physical Chemistry A 119*(28), 7924–7939.