THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

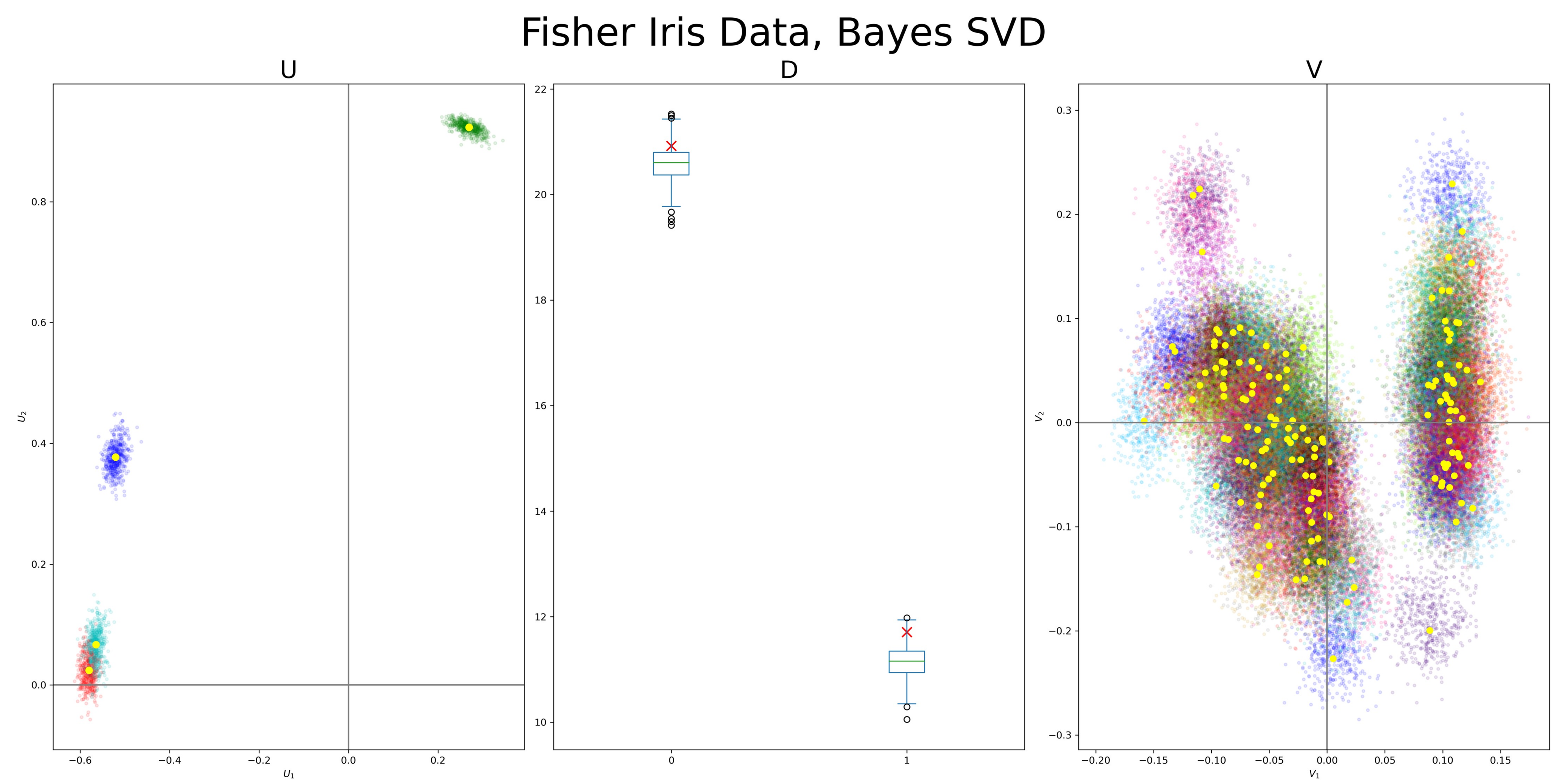
FULL BAYES SVD IN STAN

Thomas H. Keefe, Jan Hannig, J.S. Marron

Department of Statistics & O.R., UNC-Chapel Hill

ABSTRACT

We provide the first Stan implementation of a full-Bayes SVD, $Y = UDV^T + \epsilon$. Given observed matrix $Y \in \mathbb{R}^{D\times N}$, we provide posterior distributions for U, D, V, and var(ϵ), where $U \in \mathbb{R}^{D\times K}$ and $V \in \mathbb{R}^{N\times K}$ are the left and right singular vectors of Y, and $D \in \mathbb{R}^K$ is the vector of singular values. The key technical challenge of sampling orthonormal matrix parameters in Stan is achieved by the scheme of Nirwan & Bertschinger (2019), whereby the constrained space of orthonormal matrices is mapped to an unconstrained representation using Householder transformations. We apply this procedure to the Bayes SVD described by Hoff (2007).



DATA MODEL

$$Y = UDV^T + \epsilon$$
$$U \sim \text{Uniform Stiefel}(K, D)$$
$$V \sim \text{Uniform Stiefel}(K, N)$$
$$d_k \sim \text{Half Cauchy}(0,1)$$
$$\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$
$$p(\sigma) \propto 1/\sigma$$

SAMPLING ORTHOGONAL MATRIX PARAMETERS

Stan's Hamiltonian Monte Carlo sampler cannot directly sample constrained parameters; they must instead be sampled in an unconstrained space and then transformed, with a Jacobian adjustment to the joint probability density. For example, to sample a nonnegative parameter y, Stan samples $\log y \in \mathbb{R}$, and increments the log probability density by $\log \frac{d \log y}{dy} = -y$.

Continued ☞

Orthogonal and orthonormal matrices form a constrained subset of $\mathbb{R}^{D\times K}$ called the Stiefel manifold, $\{Q \in \mathbb{R}^{D\times K}: Q^T Q = I\}$, which has $DK - \frac{1}{2}K(K-1)$ degrees of freedom. Following Niwran & Bertschinger (2019), we rely on an unconstrained parameterization of the Stiefel manifold based on Householder transformations. Let vectors $v_D, v_{D-1}, \dots, v_{D-K+1}$ be uniformly distributed on the unit spheres in decreasing dimensions $S^{D-1}, S^{D-2}, \dots, S^{D-K+2}$, and let $\mathbf{H}_d(\boldsymbol{v}_d)$ be the Householder transformation in $\mathbb{R}^{D\times D}$ defined by each $\boldsymbol{v}_d$: that is, with

$$\boldsymbol{u}_d = \boldsymbol{v}_d + \operatorname{sgn}(\boldsymbol{v}_{d1})\boldsymbol{e}_1$$

we have

$$\mathbf{H}_d(\boldsymbol{v}_d) = \begin{pmatrix} \mathbf{I}_{D-d} & \mathbf{0} \\ \mathbf{0} & -\operatorname{sgn}(v_{d1})\left(\mathbf{I}_d - \frac{2\boldsymbol{u}_d\boldsymbol{u}_d^t}{\boldsymbol{u}_d^t\boldsymbol{u}_d}\right) \end{pmatrix} \in \mathbb{R}^{D\times D}$$

Then the product $\mathbf{H}_D(\boldsymbol{v}_D)\mathbf{H}_{D-1}(\boldsymbol{v}_{D-1}) \dots \mathbf{H}_{D-K+1}(\boldsymbol{v}_{D-K+1})$, after dropping the last K columns, are a uniform draw from Stiefel(K,D).

STAN MODEL

```
data{
  int<lower=1> N;
  int<lower=1> D;
  int<lower=1> K;
  matrix[D, N] Y;
}
transformed data{
  vector[D*N] Y_v = to_vector(Y);
}
parameters{
  vector[stiefel_dim(K,D)] u;
  positive_ordered[K]      d;
  vector[stiefel_dim(K,N)] v;
  real<lower=0>        sigma;
}
transformed parameters{
  matrix[D, K] U;
  matrix[N, K] V;
  U = ortho_matrix_lp(D, K, u);
  V = ortho_matrix_lp(N, K, v);
}
model{
  sigma ~ cauchy(0, 1);
  target += -log(sigma);
  matrix[D, N] MuMatrix = U*d*V';
  Y_v ~ normal(to_vector(MuMatrix),
sigma);
}
```

STAN FUNCTIONS

```
matrix Householder (int k, int D, vector v_col){
// Householder trans. normal to the vector v_col
  vector[D-k+1] v = v_col;
  real sgn = sign(v[1]);
  v[1] += sgn;
  real new_vtv = 2*sgn*v[1];
  matrix[D-k+1, D-k+1] uuT = (v * v')/(v'*v)
  matrix[D-k+1, D-k+1] Htilde;
  Htilde = -sgn * add_diag(-2 * uuT, 1.0);
  matrix[D,D] H = identity_matrix(D);
  H[k:D, k:D] = Htilde;
  return H;
}

matrix ortho_matrix_lp (int D, int K, vector v) {
// Return the ortho matrix corresponding to
// the unconstrained vector v
  // 1) Create V from v; take care of jacobians
  matrix[D, K] V = to_lower_triangular(D, K, v);
  vector[K] logjacobians;
  for (k in 1:K){
    vector[D] Vk = V[,k];
    real r = norm2(Vk);
    r ~ gamma(100, 100);
    logjacobians[k] = -log(r)*(D-k);
    V[,k] = Vk/r;
  }
  target += sum(logjacobians);

  // (2) Build the orthogonal matrix
  matrix[D, D] H = identity_matrix(D);
  for (j in 1:K){
    // iterate backwards over cols of V
    int k = K - j + 1;
    vector[D-k+1] v_col = V[k:D, k];
    H *= Householder(k, D, v_col);
  }
  return H[, 1:K];
}
```

REFERENCES

Hoff, P. D. (2007). Model Averaging and Dimension Selection for the Singular Value Decomposition. *Journal of the American Statistical Association* 102 674–85.

R. Nirwan and N. Bertschinger (2019). Rotation invariant householder parameterization for Bayesian PCA. *Proceedings of the 36th International Conference on Machine Learning*, volume 97, 4820–4828.

CONTACT

THK: tkeefe@live.unc.edu
JH: jan.hannig@live.unc.edu
JSM: marron@unc.edu