# waspr: an R package for computing Wasserstein barycenters of subset posteriors

Jolien Cremers

August 19th, 2020

# Parallel computation for MCMC

- Averaging
- Consensus Monte Carlo (Scott et al., 2016)
- Semiparametric density product estimators (Neiswanger et al., 2013)
- Wasserstein barycenters for subset posteriors (Srivastava et al., 2015, 2018)

# Wasserstein distance

A geometric distance between probability measures used in transportation theory.

E.g.: $X \sim L$, $Y \sim Q$ with densities $l$ and $q$ with finite $p$th moments and $X, Y \in \mathbb{R}^d$. Let $\mathcal{J}(L, Q)$ denote all joint distributions $J$ for $(X, Y)$ that have marginals $L$ and $Q$, then the Wasserstein distance is:

$$W_p(L, Q) = \left( \inf_{J \in \mathcal{J}(L,Q)} \int ||x - y||^p dJ(x, y \right)^{1/p},$$

where $p \geq 1$. The minimizer $J^*$ is called the ‘optimal transport plan’ or ‘optimal coupling’

# Wasserstein barycenters

Suppose we have a set of distributions $Q_1, \ldots, Q_N$, the Wasserstein barycenter is then defined as the distribution $Q$ that minimizes:

$$\sum_{j=1}^{N} W(Q, Q_j)$$

# Computation

When $d = 1$ computation of $W(L, Q)$ is easy.

If $L$ and $Q$ are the empirical distributions of two datasets $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ the distance takes the following function of the order statistics:

$$W_p(L, Q) = \left( \sum_{i=1}^{n} ||X_{(i)} - Y_{(i)}||^p \right)^{1/p}$$

Otherwise we need linear programming or other algorithms to compute the Wasserstein distance and barycenter.

# Iterative swapping algorithm

Barycenter $\mu$ of $n$ probability measures $\mu_1, \ldots, \mu_n \in P_2(\mathbb{R}^d)$.

**Puccetti et al. (2020):**

- $X_i$ is an optimal n-coupling if and only if the distribution $\bar{\mu}_n$ of $\bar{S}_n := \sum_{i=1}^{n} x_i/n$ is a barycenter of $\mu_i$
- Optimal n-couplings $X_i = (X_1, \ldots, X_n)$ are defined as solutions of
$$sup \{\mathbb{E}[f(X_1, \ldots, X_n)]; X_i \sim \mu_i, 1 \leq i < n\}$$
where $f(X_1, \ldots, X_n) = \sum_{i=1}^{n} \langle x_i, \sum_{j \neq i} x_j \rangle$.
- Use an iterative version of the swapping algorithm from Puccetti (2017).

# Iterative swapping algorithm

$n =$ number of pre-assigned probability measures (subposteriors)
$k =$ number of atoms of empirical measures (MCMC samples)
$d =$ dimensionality of the space $\mathbb{R}^d$ (parameters)

1. Start with an assignment $x^1_{\sigma_1}, \ldots, x^n_{\sigma_n}$ where $\sigma_1, \ldots, \sigma_n \in \Sigma_k$ and $\Sigma_k$ denotes the set of all permutations of $\{1, \ldots, k\}$.
2. Swapping condition for a fixed $i \in \{1, \ldots, n\}$, and indices $1 \leq k_1 < k_2 \leq k$:

$$\langle x^i_{\sigma_i(k_1)}, \sum_{j \neq 1} x^j_{\sigma_j(k_1)} \rangle + \langle x^i_{\sigma_i(k_2)}, \sum_{j \neq 1} x^j_{\sigma_j(k_2)} \rangle < \langle x^i_{\sigma_i(k_2)}, \sum_{j \neq 1} x^j_{\sigma_j(k_1)} \rangle + \langle x^i_{\sigma_i(k_1}$$

3. If condition holds, swap $\sigma_i(k_1)$ and $\sigma_i(k_2)$. A new assignment $\{\sigma'_1, \ldots, \sigma'_n\}$ is found.
4. Repeat 3 with $\sigma = \sigma'$ until no further swaps are possible and output the final assignment $\{\hat{\sigma}_1, \ldots, \hat{\sigma}_n\}$

# Assumptions WASP

The subset posterior distributions should provide a noisy approximation fo the full data posterior. Problematic in rare event data.

Exact assumptions and proofs for when WASP converges to the full data posterior given in Srivastava et al. (2018).

# waspr

```r
library(waspr)
```

```r
out <- wasp(pois_logistic,
            iter = 10,
            acc = 0.001,
            par.names = c("beta_s", "alpha_l", "beta_l",
                         "baseline_sigma", "baseline_mu",
                         "correlation", "sigma_s", "sigma_l"))
```

# waspr

```r
summary(out)
```

```
##                      mean       mode         sd      LB HPD     UB HPD
## beta_s          0.5527601  0.5518034 0.10988949  0.36598187  0.7896041
## alpha_l         2.6811079  2.6959176 0.19199304  2.30380675  3.0295802
## beta_l          0.7508520  0.7339988 0.21631011  0.37281283  1.1740767
## baseline_sigma  0.3563222  0.3811609 0.06859910  0.21910807  0.4870079
## baseline_mu    -0.8008872 -0.7516167 0.10867533 -1.01168299 -0.5944583
## correlation     0.1732170  0.1392670 0.07437737  0.02824474  0.3059979
## sigma_s         1.7225455  1.7535499 0.17920847  1.40126462  2.0610585
## sigma_l         1.2190297  1.2612822 0.07558163  1.06768047  1.3569757
```

# Poisson-logistic joint model

Hierarchical logistic:

$$\pi_{it} = P(Y_{it} = 1 \mid \boldsymbol{x}_{it}, b_i) = \frac{1}{1 + \exp(-\boldsymbol{x}_{it}\boldsymbol{\beta} + b_i)}$$

$b_i$ = random intercept

Hierarchical Poisson log-linear model (proportional hazards model with piece-wise constant baseline hazard):

$$\log \mu_{it} = \log t_{it} + \boldsymbol{x}_{it}\boldsymbol{\eta} + \alpha_t + u_i,$$

$\mu_{it}$ = hazard, $\alpha_t = \log \lambda_t = \mu_\lambda + N(0, \sigma_\lambda)$, $u_i$ = frailty, $t_{it}$ = offset

# Poisson-logistic joint model

Random intercept $b_i$ and frailty $u_i$ assumed to follow a multivariate normal distribution with the following variance-covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_b^2 & \sigma_{bu} \\ \sigma_{bu} & \sigma_u^2 \end{bmatrix}$$

# Simulation results

$N = 6000$, $n_i = 30$, 8 subsets

```
##                 real estimate  bias   sd coverage
## beta_s          0.50     0.50  0.00 0.04     0.98
## alpha_l         1.00     1.00  0.00 0.04     1.00
## beta_l          0.50     0.50  0.00 0.06     0.98
## baseline_sigma  0.30     0.31  0.01 0.08     0.72
## baseline_mu    -1.00    -0.96  0.04 0.12     0.73
## correlation     0.35     0.35 -0.01 0.03     0.99
## sigma_s         1.41     1.43  0.01 0.04     1.00
## sigma_l         1.00     1.01  0.01 0.06     0.86
```

```
## [1] 309.084
```

# References I

Neiswanger, W., Wang, C., and Xing, E. (2013). Asymptotically exact, embarrassingly parallel MCMC.

Puccetti, G. (2017). An algorithm to approximate the optimal expected inner product of two vectors with given marginals. *Journal of mathematical analysis and applications*, 451(1):132–145.

Puccetti, G., Rüschendorf, L., and Vanduffel, S. (2020). On the computation of Wasserstein barycenters. *Journal of multivariate analysis*, 176:104581.

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., and McCulloch, R. E. (2016). Bayes and big data: the consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, 11(2):78–88.

# References II

Srivastava, S., Cevher, V., Dinh, Q., and Dunson, D. (2015). WASP: Scalable Bayes via barycenters of subset posteriors. In *Artificial Intelligence and Statistics*, pages 912–920.

Srivastava, S., Li, C., and Dunson, D. B. (2018). Scalable bayes via barycenter in Wasserstein space. *Journal of machine learning research: JMLR*, 19(8):1–35.