

Hierarchical Bayesian Small Area Estimation for Diurnal Data with Application to Recreational Fisheries

Hernandez-Stumpfhauser, Daniel
Colorado State University, Department of Statistics
 102 Statistics
 Fort Collins, CO 80523-1877 USA
 E-mail: hernande@stat.colostate.edu

Opsomer, Jean D.
Colorado State University, Department of Statistics
 102 Statistics
 Fort Collins, CO 80523-1877 USA
 E-mail: jopsomer@stat.colostate.edu

Breidt, F. Jay
Colorado State University, Department of Statistics
 201 Statistics
 Fort Collins, CO 80523-1877 USA
 E-mail: jbreidt@stat.colostate.edu

Background

In this application, we are interested in obtaining predictions of the daily distributions of the departures of recreational anglers along the coasts of the United States, as a function of the type of fishing trip, its location and time of year. In order to reflect the circular nature of the departure times, we model them as projected bivariate normal random variables. We propose a new latent hierarchical Bayesian regression model that makes it possible to incorporate covariates and allows for spatial prediction and inference. We investigate a number of issues related to model specification, model selection and computational efficiency. The approach is applied to a large dataset collected by the US National Oceanic and Atmospheric Administration (NOAA).

The data of interest are collected through the Marine Recreational Fisheries Statistics Survey (MRFSS), which is a national survey of recreational fishing activities in saltwater. Such activities constitute a multi-billion dollar industry in the United States. The survey is conducted by an agency within NOAA. Its major goal is to estimate recreational fish catch by species and size class. These estimates are also broken down by region, by mode of fishing (shore, charter boat, private boat, party boat) and by season:

$$(17 \text{ coastal states}) \times (4 \text{ modes}) \times (6 \text{ two-month waves}) = 408 \text{ domains}$$

(NOAA oversees all US coastal fishing, but only 17 Atlantic coast and Gulf coast states are included because the Pacific coast states and Texas have their own surveys.) The estimates are used in fisheries stock assessments and in fisheries regulations, such as setting quotas on species, start and end dates for the fishing season, etc.

MRFSS actually consists of two separate and complementary surveys. The Access Point Angler Intercept Survey (APAIS) collects data on catch at the fishing site. An on-site interviewer “intercepts” anglers as they leave the site. Data from this survey are used to estimate average catch per angler trip. Data are also used to make coverage adjustments to the Coastal Household Telephone Survey (CHTS), which collects fishing activity data from a telephone survey of coastal households. The CHTS uses random digit dialing (RDD) to find saltwater recreational anglers, which is expensive and

time-consuming. It focuses on coastal households to increase the “hit rate” of anglers, then makes adjustments for the obvious undercoverage problem by using the APAIS data on intercepted anglers from non-coastal households. CHTS data are then used to estimate total angler trips. Finally, the estimated total catch is obtained as the product:

$$(\text{estimated average catch per trip}) \times (\text{estimated total trips}).$$

APAIS consists of two or more stages of sampling. The first stage consists of stratified unequal-probability selection of site-days in each fishing mode, with probabilities proportional to a known index of expected fishing pressure. Subsequent stages depend on the mode. For boat modes, intermediate stages consist of selection of fishing boats and groups of anglers within boats, with each stage approximated by simple random sampling without replacement. The final stage of sampling in any mode consists of equal-probability selection of anglers observed leaving the site during the interviewer’s on-site assignment within the selected day.

A serious problem with estimation in the original MRFSS was that it ignored essentially all aspects of this design in the estimation, as noted by a US National Academy of Sciences panel [8]. Valid design-based estimation requires sampling weights accounting for the survey design. Consider two stages of selection. Let U_I denote the set of all site-days in a domain of interest, and U_d denote the set of all anglers departing the site during that day. A sample $s_I \subset U_I$ of site-days is selected via a probability sampling design, with first-order inclusion probabilities $\pi_{Id} > 0$ for $d \in U_I$. Within each selected site-day, a sample $s_d \subset U_d$ of anglers is intercepted by the interviewer as they leave the site. Let $\pi_{a|d} > 0$ denote the probability that angler $a \in U_d$ is intercepted by the interviewer. Then an unbiased estimator of the total $t_y = \sum_{d \in U_I} \sum_{a \in U_d} y_{da}$ for $y_{da} =$ catch characteristic of angler a on site-day d is given by

$$\hat{t}_y = \sum_{d \in s_I} \sum_{a \in s_d} \frac{y_{da}}{\pi_{Id} \pi_{a|d}}.$$

The site-day inclusion probability, π_{Id} , is based on known fishing pressure and can be readily calculated. The angler inclusion probability $\pi_{a|d}$, on the other hand, is unknown. Let Δ represent the “slice” of time during which the interviewer visited the site on a given day. Let $F_{\Delta d}$ denote the fraction of anglers who departed the site during that time slice. The inverse of the angler inclusion probability is interpreted as a weight: the n_d anglers intercepted while departing the site during a Δ_d represent all N_d anglers departing the site during Δ_d , and those N_d anglers in turn represent $(1/F_{\Delta d})$ anglers departing the site during the entire day. The weight is then

$$\pi_{a|d}^{-1} = (N_{\Delta d}/n_{\Delta d})(1/F_{\Delta d}).$$

A problem with this weight is that it is unknown because $F_{\Delta d}$ is unknown.

Small area estimation of time slice distributions

In our work we have created estimated sampling weights for the APAIS by estimating the “time slice distribution,” $F_{\Delta d}$. Our approach is to use CHTS to create estimated within-site-day weights by estimating the fraction of total daily departures $F_{\Delta d}$ from site-day d for a given time interval Δ . CHTS data are available for 980,000 trips by 215,000 households between 1990 and 2008. Each trip record available for this project contains:

- departure time
- fishing mode (shore, charter boat, private boat, party boat)
- fishing date, coarsened to: wave (2-month interval)

- fishing site, coarsened to: state (17 states).

To estimate the time slice distribution, we consider as our target of estimation the finite population quantity

$$F_{N,ijk}(t_0, \Delta) = \frac{\sum_{t \in U_{ijk}} I_{[t_0, t_0 + \Delta]}(T_{ijkt})}{N_{ijk}}$$

with

$$\begin{aligned} U_{ijk} &= \text{population of fishing trips in state } i, \text{ wave } j, \text{ mode } k \\ N_{ijk} &= \text{total number of trips in state } i, \text{ wave } j, \text{ mode } k \\ T_{ijkt} &= \text{departure time for trip } t \\ I_{[t_1, t_2]}(T) &= \begin{cases} 1 & \text{if } T \text{ is in the interval } [t_1, t_2) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

This quantity $F_{N,ijk}(t_0, \Delta)$ can be estimated directly from the CHTS data:

$$\hat{F}_{ijk}(t_0, \Delta) = \frac{\sum_{s_{ijk}} I_{[t_0, t_0 + \Delta]}(T_{ijkt})}{n_{ijk}}.$$

This direct estimator is unbiased but will be highly variable in many (i, j, k) domains with low sample size. We thus consider the use of small area estimation methods to borrow strength across domains. The small area estimation method should allow for:

- circular nature of time-of-day data (24-hour clock)
- spatial and temporal distribution of fishing behavior
- correlated observations (multiple trips per respondent)
- “design consistency”: when the sample size is large within a cell, the estimator should be close to $\hat{F}_{ijk}(t_0, \Delta)$.

In what follows, we begin by describing the Bayesian hierarchical model we have fitted to the CHTS data. We then describe two different approaches to create small area estimates. The first is based on composite estimation, and is the one currently implemented by NOAA. The second “internalizes” the small area predictions into the hierarchical model itself, using a Dirichlet process specification.

Hierarchical modeling with projected normal distributions

Our approach to this problem is to build a hierarchical regression model for the departure times using a *spherically projected multivariate linear* model (see [6]). This model is an immediate generalization of the *projected normal* distribution. We conduct Bayesian inference for the model using Markov chain Monte Carlo (MCMC). We have used an “optimal” *composite estimation* (e.g. [4]) to combine observed and model-predicted distributions, and in ongoing work, are building the compositing into the hierarchical model structure.

Let \mathbf{I}_2 denote the 2×2 identity matrix. A random angle θ has a *projected normal distribution* $\mathcal{PN}(\boldsymbol{\mu}, \mathbf{I}_2)$ if $\mathbf{X} = (X_1, X_2)^T \sim \mathcal{N}_2(\boldsymbol{\mu}, \mathbf{I}_2)$, with

$$X_1 = \|\mathbf{X}\| \cos \theta, \quad X_2 = \|\mathbf{X}\| \sin \theta.$$

In other words, the random angle θ can be generated by drawing $(X_1, X_2)^T \sim \mathcal{N}_2(\boldsymbol{\mu}, \mathbf{I}_2)$, converting to polar coordinates, and discarding the length $\|\mathbf{X}\|$. Explicit expressions for the probability density function and the moments of θ are available. In this distribution, the parameter $\boldsymbol{\mu}$ plays the role of both

the “mean” and the “variance” (in the somewhat loose sense of measures of center and spread). To see this, consider that as $\boldsymbol{\mu} \rightarrow \mathbf{0}$, the distribution of θ becomes uniform on the unit circle. Conversely, letting $\boldsymbol{\mu} = (\mu_1, \mu_2) = (\|\boldsymbol{\mu}\| \cos \psi, \|\boldsymbol{\mu}\| \sin \psi)$, if $\|\boldsymbol{\mu}\| \rightarrow \infty$ with ψ held fixed, then the distribution of θ becomes a point mass at ψ .

The key advantage of the projected normal distribution vs. other circular distributions (such as von Mises) is normality. This structure makes it conceptually easy to construct regression models. For hierarchical model building and Bayesian inference, conjugate priors are available and are easy to sample. The main disadvantage is the latent structure: each random angle has a corresponding latent random variable corresponding to the “discarded length” of the bivariate normal random vector. Estimation methods therefore need to account for a number of latent variables equal to the number of observations.

We use Bayesian inferential techniques built upon [5], which discusses Bayesian inference for $\mathcal{PN}(\boldsymbol{\mu}, \mathbf{I}_2)$. To handle the latent variables, they augment the observed angle data, denoted T_{ijkt} , with the latent lengths, denoted R_{ijkt} :

$$\mathbf{X}_{ijkt} = R_{ijkt} \begin{bmatrix} \cos T_{ijkt} \\ \sin T_{ijkt} \end{bmatrix}.$$

In this formulation, the joint distribution of (R_{ijkt}, T_{ijkt}) and the conditional distributions are easily obtained. Given the latent lengths, the problem becomes inference for a hierarchical model for normal random variables. Additional steps are then needed to account for the distribution of R_{ijkt} .

We extended this formulation to allow for a regression model in the mean structure, starting with conjugate normal and inverse gamma as prior distributions. We then applied a *sweeping reparameterization* [9] to deal with the issue of correlated random effects (e.g., [3]), so that we end up with the following hierarchical model specification:

$$\begin{aligned} \boldsymbol{\mu}' &\sim \mathcal{N}_2(\boldsymbol{\mu}_0, \sigma_0'^2) \\ \mathbf{m}'_{-K,c} &\sim \mathcal{N}_{K-1}\left(\mathbf{0}, \sigma_m^2 \left(\mathbf{I}_{K-1} - \frac{1}{K} \mathbf{J}_{K-1}\right)\right) \quad c = 1, 2 \\ \mathbf{s}'_{-I,c} \mid \sigma_s^2 &\sim \mathcal{N}_{I-1}\left(\mathbf{0}, \sigma_s^2 \left(\mathbf{I}_{I-1} - \frac{1}{I} \mathbf{J}_{I-1}\right)\right) \quad c = 1, 2 \\ \mathbf{w}'_{-J,c} \mid \sigma_w^2 &\sim \mathcal{N}_{J-1}\left(\mathbf{0}, \sigma_w^2 \left(\mathbf{I}_{J-1} - \frac{1}{J} \mathbf{J}_{J-1}\right)\right) \quad c = 1, 2 \\ \sigma_s^2 &\sim \text{Inverse Gamma}(\alpha_s, \beta_s) \\ \sigma_w^2 &\sim \text{Inverse Gamma}(\alpha_w, \beta_w), \end{aligned}$$

where the $-K, -I, -J$ indicate vectors with the last element deleted. The last element \mathbf{m}'_K was obtained as $\mathbf{m}'_K = -\sum_{k=1}^{K-1} \mathbf{m}'_k$, with $\mathbf{s}'_I, \mathbf{w}'_J$ defined similarly. After this reparameterization, we obtained the full conditional distributions and used the Gibbs sampler.

Included among these conditionals are the latent bivariate normal lengths. The conditional distribution of R_{ijkt} given the observed data and parameters is

$$p_R(r \mid \cdot) \propto r \exp \left[-\frac{1}{2} r^2 + b_{ijkt} r \right]$$

with $b_{ijkt} = (\cos T_{ijkt}, \sin T_{ijkt}) \boldsymbol{\mu}_{ijk}$. Further, expressions for the cumulative distribution function (cdf) $F_R(r \mid \cdot)$ and the mode r_m are readily available. The density p_R is log-concave, so it would be possible to use an adaptive rejection scheme to select R , but for computational efficiency we developed a sampling method using the inverse cdf technique: if $u \sim U(0, 1)$, $r = F_R^{-1}(u \mid \cdot)$ is a random draw from $p_R(r \mid \cdot)$. We numerically solved for the inverse as follows:

- define $g(r, u) = F_R(r) - u$
- for given u , find r that solves $g(r, u) = 0$ by Newton-Raphson:
 - initialize by setting $r_0 = r_m$, the known mode
 - do

$$r_{s+1} = r_s - \frac{g(r_s, u)}{p_R(r_s | \cdot)}$$

for a small, fixed number of iterations.

This inverse cdf technique allowed for rapid implementation in R.

Using this Bayesian methodology, we explored a number of model specifications that included state, wave, and mode effects along with various interactions. We considered both fixed and random specifications for the various effects. We also allowed for correlated observations within households by introducing a factor \mathbf{h}_l for respondent household l (unlike the other factors, the household effect did not require centering because of the very large number of levels). Finally, we conducted model selection by comparing models using the Deviance Information Criterion (DIC). The final model we selected is as follows:

$$\boldsymbol{\mu}_{ijklt} = \boldsymbol{\mu} + \mathbf{m}_k + \mathbf{sw}_{ij},$$

where mode \mathbf{m}_k is fixed and centered and the state-wave interaction \mathbf{sw}_{ij} is random and centered.

We return now to estimation of the time-slice departure distribution. Given the posterior distribution for $\boldsymbol{\mu}_{ijk}$ estimated as described above, the posterior distribution of

$$F_{ijk}(t_0, \Delta) = \int_{t_0}^{t_0+\Delta} f_T(t | \boldsymbol{\mu}_{ijk}) dt$$

is in principle obtainable directly from the Gibbs sampler, since it is a function of only $\boldsymbol{\mu}_{ijk}$. But computing this directly would require integration at each iteration of the sampler, which would be extremely time-consuming. Instead, we streamlined the calculations by first restricting attention to $\Delta = 1$ hour and estimating 24 time slices (setting $F_{ijk}(t_0) \equiv F_{ijk}(t_0, 1)$ from now on). Next, we used trigonometric identities and analytic calculations to take advantage of the circular nature of model. In particular, it can be shown that

$$\int_{\theta}^{\theta+\pi/2} f_T(\theta | \boldsymbol{\mu}) d\theta = \Phi(-\rho \sin(\theta - \omega)) \Phi(\rho \cos(\theta - \omega)),$$

where $\Phi(\cdot)$ is the standard normal cdf and $\boldsymbol{\mu}^T = \rho(\cos \omega, \sin \omega)$, so that 24 integrals are replaced by 5 integrals plus successive differencing. Posterior distributions for the fractions of daily departures by hour are shown in Figure 1 by wave and mode, for one state. These distributions are depicted as boxplots within each hour-long increment. The boxplots are superimposed on the original departure distribution data from the CHTS, shown as histograms.

Small area estimation via estimated optimal compositing

The estimates obtained from the Bayesian hierarchical model just described are fully model-based. If we consider the main target of estimation to be the finite population quantity

$$F_{N,ijk}(t_0) = \frac{\sum_{U_{ijk}} I_{[t_0, t_0+1)}(T_{ijk t})}{N_{ijk}}$$

for $t_0 = 1, \dots, 24$, we do not necessarily want to fully trust the model to be a completely correct representation of the population. Figure 1 shows that we have two possible estimators: the direct

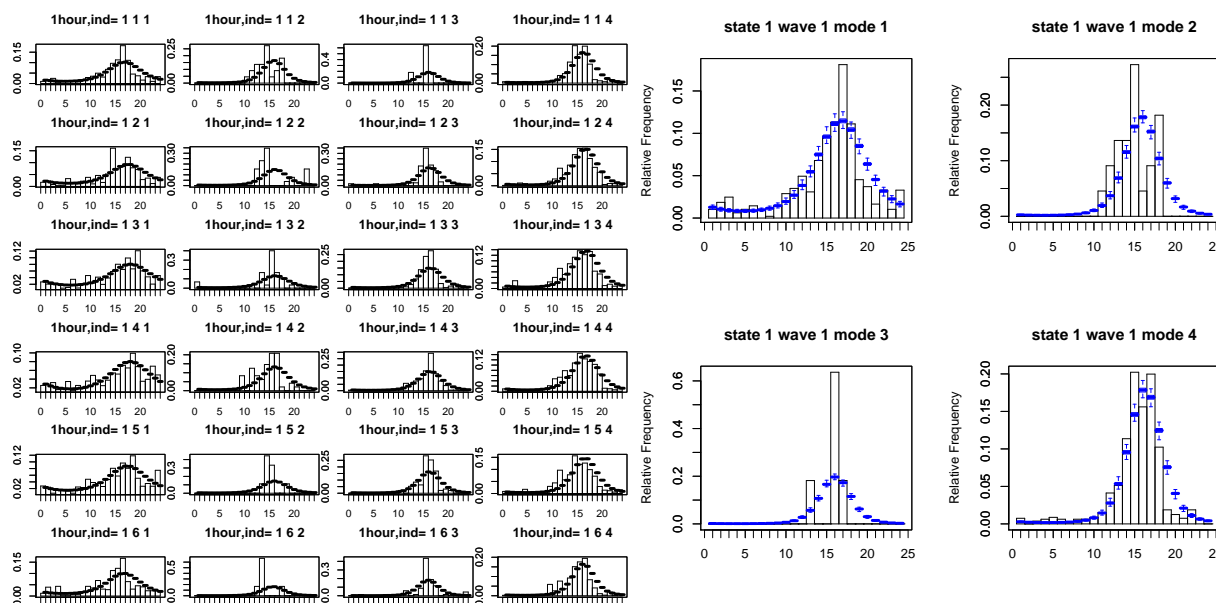


Figure 1, left: departure distribution data (histograms) by wave and mode for the state of Alabama; right: posterior distributions (boxplots) for proportion of daily departures for wave 1 in Alabama.

estimator from CHTS corresponding to the left plot,

$$\hat{F}_{ijk}^D(t_0) = \frac{\sum_{s_{ijk}} I_{[t_0, t_0+1)}(T_{ijk t})}{n_{ijk}},$$

and the model-based estimator (posterior mean from the hierarchical model) corresponding to the right plot,

$$\hat{F}_{ijk}^M(t_0) = E \left[\Pr(t_0 \leq T < t_0 + 1 \mid \boldsymbol{\mu}_{ijk}) \mid \text{data} \right].$$

The former is unbiased but highly variable when n_{ijk} is small, while the latter is very precise but subject to model misspecification bias. We therefore consider combining the strengths of the two estimators using composite estimation (e.g., [4]), which is of the form

$$\hat{F}_{ijk}^C(t_0) = w_{ijk} \hat{F}_{ijk}^D(t_0) + (1 - w_{ijk}) \hat{F}_{ijk}^M(t_0).$$

Finding a good choice of w_{ijk} is critical in composite estimation. We make a number of simplifying assumptions in order to select w_{ijk} for the departure distribution data. First, assume that $\text{Cov}(\hat{F}_{ijk}^D(t_0), \hat{F}_{ijk}^M(t_0)) = 0$ (or small enough to be ignorable), which is reasonable since $\text{Var}(\hat{F}_{ijk}^M(t_0))$ is vanishingly small compared to $\text{Var}(\hat{F}_{ijk}^D(t_0))$. Second, we assume that the bias of $\hat{F}_{ijk}^M(t_0)$ does not depend on (i, j, k) . Under those assumptions, we have

$$E \left(\hat{F}_{ijk}^M(t_0) - \hat{F}_{ijk}^D(t_0) \right)^2 = \text{MSE}(\hat{F}_{ijk}^M(t_0)) + \text{Var}(\hat{F}_{ijk}^D(t_0))$$

and the optimal value for w_{ijk} is given by

$$w_{ijk}^{\text{opt}}(t_0) = \frac{\text{MSE}(\hat{F}_{ijk}^M(t_0))}{\text{MSE}(\hat{F}_{ijk}^M(t_0)) + \text{Var}(\hat{F}_{ijk}^D(t_0))}.$$

We are interested in estimating $w_{ijk}^{\text{opt}}(t_0)$ by replacing the unknown quantities on the right-hand side by sample quantities. However, in order to maintain the fact that we require $\sum_{t=1}^{24} \hat{F}_{ijk}^C(t) = 1$, we will keep w_{ijk} fixed with respect to t_0 .

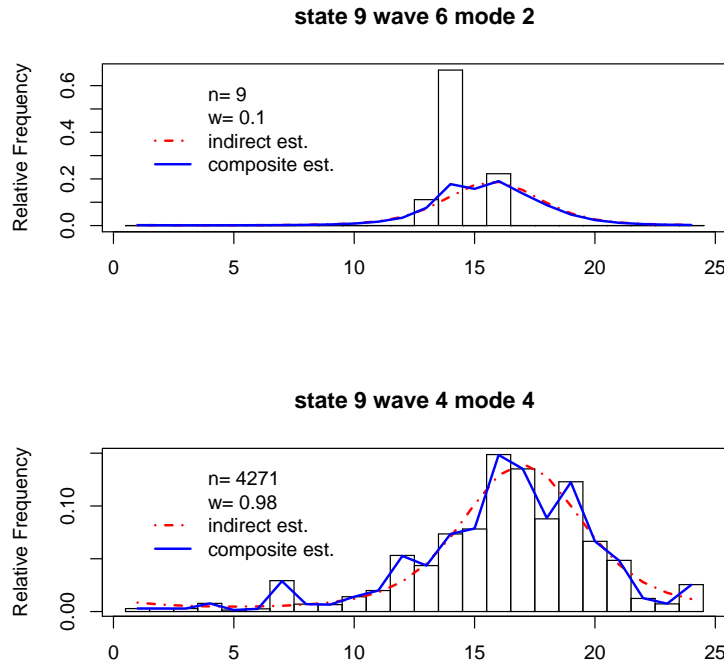


Figure 2: Estimators of the fraction of daily departures in the state of Connecticut. Histogram is the direct estimator. Dashed line is the posterior mean from the model. Piecewise linear blue curve is the composite estimator.

We estimate the t_0 -averaged $MSE(\hat{F}_{ijk}^M(t_0))$ in w_{ijk}^{opt} by

$$\widehat{MSE}(F^M) = \frac{1}{n} \sum_i \sum_j \sum_k \sum_t \left((\hat{F}_{ijk}^M(t) - \hat{F}_{ijk}^D(t))^2 - \frac{\hat{F}_{ijk}^D(t) (1 - \hat{F}_{ijk}^D(t))}{n_{ijk}} \right).$$

Replacing $\text{Var}(\hat{F}_{ijk}^D(t_0))$ in w_{ijk}^{opt} by the upper bound $0.25/n_{ijk}$, we obtain the “adaptive weights”

$$\hat{w}_{ijk} = \frac{\widehat{MSE}(F^M)}{\widehat{MSE}(F^M) + 0.25/n_{ijk}}.$$

Finally, we plug these in to obtain the composite estimator

$$\hat{F}_{ijk}^C(t_0) = \hat{w}_{ijk} \hat{F}_{ijk}^D(t_0) + (1 - \hat{w}_{ijk}) \hat{F}_{ijk}^M(t_0).$$

Figure 2 shows examples of the direct, model-based, and composite estimators for both a small sample ($n_{ijk} = 9$) and large sample ($n_{ijk} = 4271$) case. As in the top plot, when the sample size is small, our procedure selects a small value for w_{ijk} and the composite estimator is close to the model. In contrast, when the sample size is large, w_{ijk} is very close to 1 and the composite estimator is very close to the direct estimator (see bottom plot). In particular, it allows large deviations from the unimodal and symmetric shape imposed by the projected normal distribution. Hence, the composite estimator appears successful in responding to the relative precision of the direct and the model-based estimators.

Hierarchical modeling with Dirichlet mixtures

While the compositing methodology described above is relatively simple and quite effective, it would be nice to allow the model machinery to construct the composite estimator automatically, as

the appropriate posterior mean under the model. One method that we are currently exploring uses mixtures of projected normal distributions for circular density estimation. The number of mixture components is unknown a priori and is to be inferred from the data. The clustering property of Dirichlet processes provides a nonparametric prior for the number of mixture components. Density estimation is then done by modeling the data as a sample from mixtures of projected normal distributions. This allows for direct inference on uncertainty about density estimates, assessment of modality, and inference on the number of components. Finally, groups of data will be used to model the parameters of the base measure of the Dirichlet processes.

Suppose the departure times come from the following mixture model:

$$T_{ijkt} \mid \boldsymbol{\mu}_{ijkt} \sim \mathcal{PN}_2(\boldsymbol{\mu}_{ijkt}, \mathbf{I}_2),$$

which implies that $\mathbf{X}_{ijkt} \mid \boldsymbol{\mu}_{ijkt} \sim \mathcal{N}_2(\boldsymbol{\mu}_{ijkt}, \mathbf{I}_2)$ as before. If the normal means $\boldsymbol{\mu}_{ijkt}$ come from some uncertain prior distribution $G(\cdot)$ on $\mathbb{R} \times \mathbb{R}$ and $G(\cdot)$ is modeled as a Dirichlet process, then the \mathbf{X} come from a Dirichlet mixture of normals (see [1]) and hence the T come from a Dirichlet mixture of projected normals.

The Dirichlet process $\mathcal{DP}(\alpha_0, G_0)$ (see [2]) is a distribution over distributions. It has two parameters, a scaling parameter $\alpha > 0$ and a base distribution G_0 . An explicit representation of a draw from a Dirichlet process was given by [7], who showed that if $G \sim \mathcal{DP}(\alpha_0, G_0)$, then with probability one:

$$G = \sum_{k=1}^{\infty} \beta_k \delta_{\boldsymbol{\mu}_k},$$

where the $\boldsymbol{\mu}_k$ are independent random variables distributed according to G_0 and $\delta_{\boldsymbol{\mu}_k}$ is an atom at $\boldsymbol{\mu}_k$, and where the “stick-breaking” weights β_k are also random and depend on the parameter α_0 . This representation of the Dirichlet process shows that draws from it are discrete. The discreteness of the Dirichlet process makes it suitable for the problem of placing priors on mixture components.

We consider the following extended hierarchical model:

$$\begin{aligned} T_{ijkt} \mid \boldsymbol{\mu}_{ijkt} &\sim \mathcal{PN}_2(\boldsymbol{\mu}_{ijkt}, \mathbf{I}_2) \\ \boldsymbol{\mu}_{ijkt} \mid G_{ijk} &\sim G_{ijk} \\ G_{ijk} &\sim \mathcal{DP}(\alpha_0, G_0(\mathbf{m}_{ijk})) \\ G_0 &= \mathcal{N}_2(\mathbf{m}_{ijk}, \mathbf{I}_2) \\ \mathbf{m}_{ijk} &= \mathbf{m}_0 + \mathbf{s}_i + \mathbf{w}_j + \mathbf{m}_k \\ \alpha_0 &\sim \text{Gamma}(a, b) \end{aligned}$$

where the effects $\mathbf{s}_i, \mathbf{w}_j, \mathbf{m}_k$ and the overall mean \mathbf{m}_0 are all distributed as bivariate normals with their own means and variances.

While our experimentation with this model structure is ongoing, preliminary results are encouraging. Figure 3 shows results for the estimated mode effect in the simplified model

$$\mathbf{m}_{ijk} = \mathbf{m}_0 + \mathbf{m}_k,$$

for a given state (state 33, New Hampshire) and Figure 4 shows composite estimators by wave and mode for various waves and modes in the same state. The composite estimators seem to do an exceptionally good job of approximating the direct estimates when data are rich, and reverting to the overall mode effect when data are sparse. The only disadvantage of the Dirichlet mixture approach at this point is its computational expense, making it currently unfeasible for implementation by NOAA. We are currently investigating approaches to reduce the computational burden of this approach.

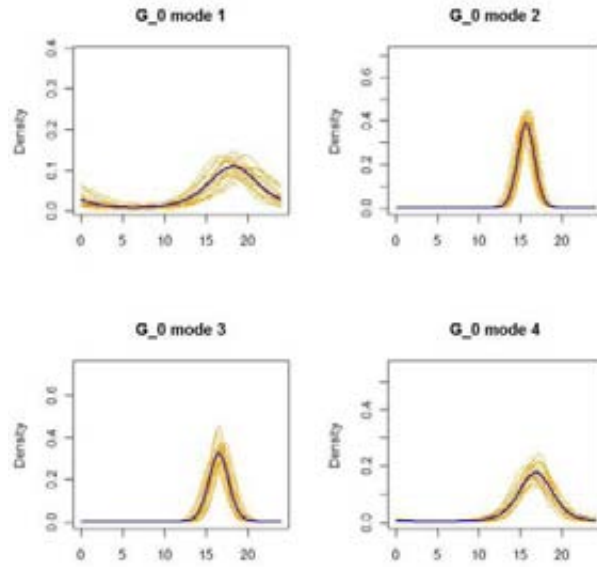


Figure 3: Results from the Dirichlet mixture model for mode effects. The blue curve in each panel is the posterior mean estimate of the mode effect, computed by averaging all Gibbs iterates. Orange curves are 20 Gibbs iterates sampled at random from all iterates, to show variability in the posterior distribution. Departures by anglers fishing from shore (mode = 1) are more diffuse across time than departures by anglers fishing from boats (modes 2, 3, 4).

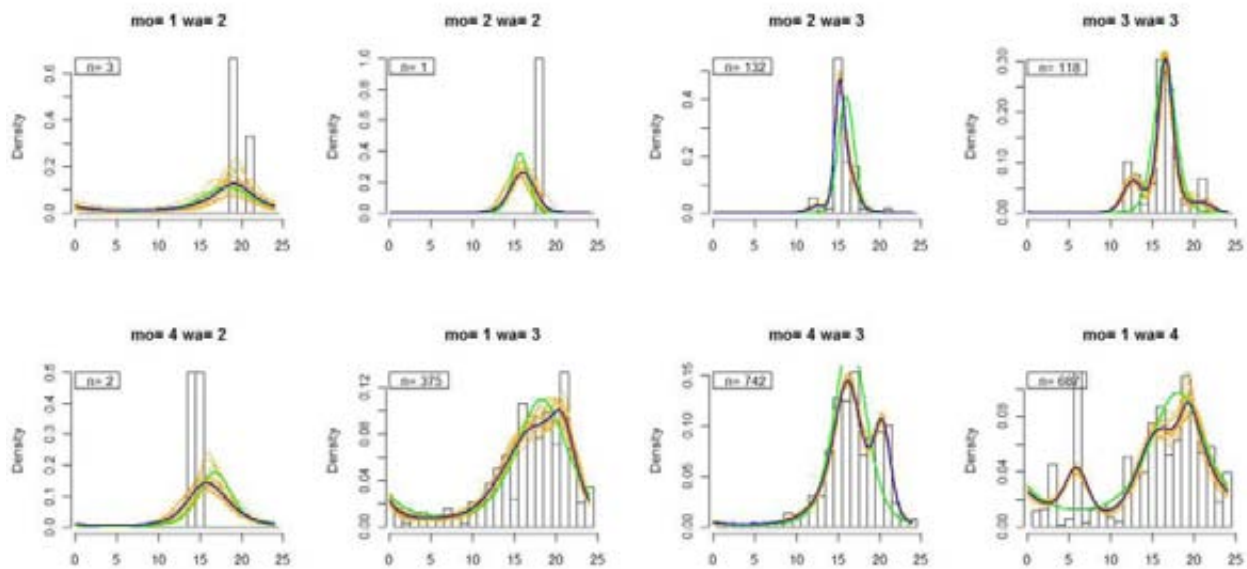


Figure 4: Results from the Dirichlet mixture model for various waves and modes in the state of New Hampshire. Histogram is the direct estimator. The blue curve is the composite estimator for the given domain (posterior mean computed as the sample mean of all Gibbs iterates). Orange curves are 20 Gibbs iterates sampled at random from all iterates, to show variability in the posterior distribution. Green curves show the overall (non-state-specific) mode effects (as in Figure 3).

REFERENCES (RÉFÉRENCES)

- [1] M.D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588, 1995.
- [2] T.S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.
- [3] A. E. Gelfand, S. K. Sahu, and B. P. Carlin. Efficient parametrizations for normal linear mixed models. *Biometrika*, 82:479–488, 1995.
- [4] M. Ghosh and J. N. K. Rao. Small area estimation: an appraisal. *Statistical Science*, 9:55–93, 1994.
- [5] G. Nuñez-Antonio and E. Gutiérrez-Peña. A Bayesian analysis of directional data using the projected normal distribution. *Journal of Applied Statistics*, 32(10):995–1001, 2005.
- [6] B. Presnell, S. P. Morrison, and R. C. Littell. Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, 93(443):1068–1077, 1998.
- [7] J. Sethuraman. A constructive definition of Dirichlet measures. *Statistica Sinica*, 1994.
- [8] P.J. Sullivan, F.J. Breidt, R.B. Ditton, B.A. Knuth, B.M. Leaman, V.M. O’Connell, G.R. Parsons, K.H. Pollock, S.J. Smith, and S.L. Stokes. *Review of Recreational Fisheries Survey Methods*. National Academies Press, Washington, DC, 2006.
- [9] S. K. Vines, W. R. Gilks, and P. Wild. Fitting Bayesian multiple random effects models. *Statistics and Computing*, 6:337–346, 1996.