

Improving Catch Estimation Methods in Sparsely Sampled Mixed-Stock Fisheries.

Nick Grunloh, Edward Dick, Don Pearson, John Field, Marc Mangel

Abstract

Introduction

Context

Data

- Collection issues
 - funding => nature of sparcity
- Lay down goal modeling goal
 - mean
 - uncertainty

Methods

Data Generating Model

Something something heirarchical poisson model. Something something (Shelton, 2012).

For the purposes of accurately modeling not only species composition means, but also higher moments of the data, such as species composition variances, it is necessary to recognize model limitations with respect to over-dispersed data. Among the simplest models for count data are the poisson and binomial models. Both models are typically specified with a single degree of freedom for modeling all of the moments of the data, and thus they rely heavily on their respective data generating processes to accurately represent higher moments in the data. McCullagh and Nelder (1989, pg. 124) commiserate about the prevalence of

over-dispersed data in cluster sampling, and explain the numerous ways in which cluster sampling may result in over-dispersion.

Extending the Poisson and binomial models to deal with over-dispersion, typically involves adding additional parameters for the purpose of modeling higher moments of the data. The negative binomial (NB) distribution is often used as an over-dispersed extension of the poisson model, since it can be expressly written as an infite mixture of poisson distributions. While the beta-binomial model is typically used to as an over-dispersed extension of the binomial model.

An Example

To discern between these models we consider a small scale example of the Poisson, binomial, negative binomial, and beta-binomial models fit to the port sampling integer weight data from market category 250, in the Monterey port complex trawl fishery in 1990. (*anywillwork*) This stratum was visited 38 times by port samplers, collecting a total of 67 cluster samples, resulting in 344 model observations across 21 (*atleast;URCK*) unique species. Each of the above models are fit to these data. The predictive species composition distributions from each model are visualized in Figure (1) as 95% Highest Density Intervals (HDI), plotted on top of the predictive means for each model and the observed species compositions from the data in Figure(1). For brevity we only consider the most prevalent six species in this example (CLPR, BCAC, WDOW, BLGL, ARRA, BANK). Additionally, the MSE, DIC, WAIC, and marginal likelihood ratio model probabilities are computed for each model as measures of model fit as seen in Table(1).

	Poisson	Binomial	NB	BB
MSE	0.05286	0.05683	0.05188	0.05170
DIC	5675.25	6759.86	1301.51	1261.00
WAIC	5840.56	6939.74	1302.19	1261.30
$pr(M y)$	≈ 0	≈ 0	$< 10^{-10}$	≈ 1

The large spread of the observed species compositions seen in Figure(1) demonstrate the degree of overdispersion present in port sampling data. The Poisson and binomial models attempt to model both the mean and residual variances with a single parameter for each species. This can tend to biase mean estimates in overdispersed stratum toward larger values, for the sake of estimating larger residual variances. Similarly variance estimates may tend to be biased toward smaller values for the sake of estimating smaller means. In contrast, the negative binomial and beta-binomial models estimate an additional parameter which disentangles the mean and residual variance estimates. Thus the negative binomial and beta-binomial models are able to produce more accurate estimates of both the mean and residual variance.

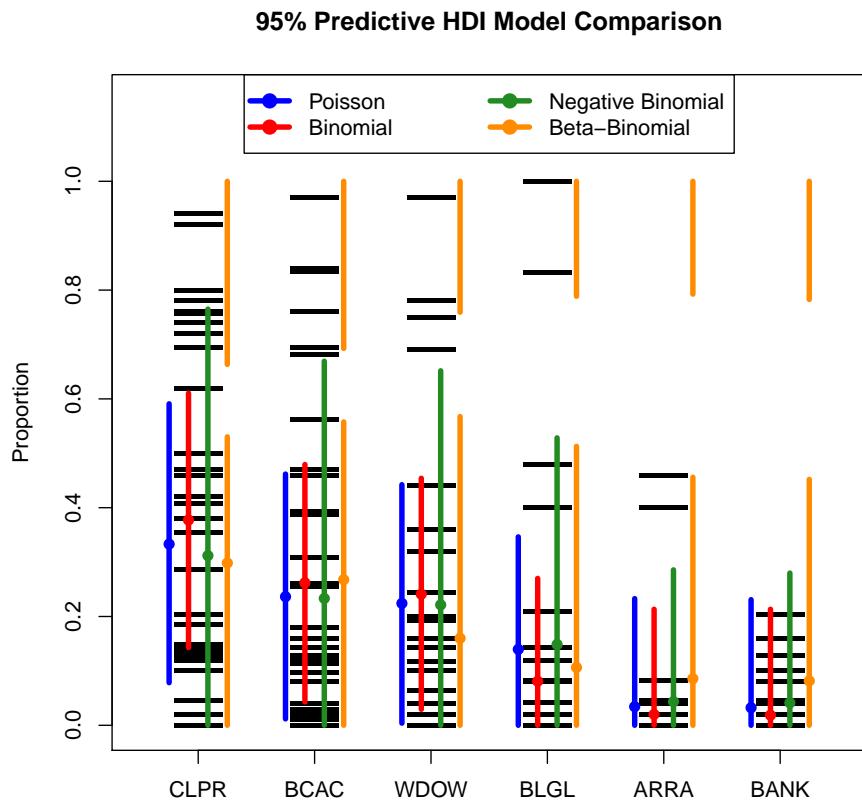


Figure 1: overdisp box

All of the measures in Table(1) consistently agree that the negative binomial and beta-binomial models out perform the overdispersed Poisson and binomial models. Furthermore, all of the metrics in Table(1) are also able to discern that the beta-binomial model outperforms the negative binomial model. Depending on the users value system toward model selection (ex. predictive or inferential), the support for the beta-binomial model over the negative binomial model may vary, but it is worth noting that all of the measures considered here tell the same story indicating preference for beta-binomial model.

The split beta-binomial interval seen in Figure(1) is the consequence of confining a large amount of residual variability to the unit interval. The beta-binomial is the only model considered here, which includes such a large degree of variability and thus it is the only model that produces predictive species composition distributions of the sort. Figure(2) shows the beta-binomial predictive distributions as a violin plot demonstrating how the beta-binomial model arranges predictive density over the unit interval. The intervals in Figure(1) are the smallest possible regions on these densities so that the intervals contain 95% of the predictive density. For the cases of Aurora and Bank rockfish, the empty upper regions seen in Figure(1) are understandable in terms of the relatively low density region of the posterior seen in Figure(2).

Operationalized Model

See proposals...

- maths stuffs
 - mean function
 - variance; introduce ρ
- justify linear predictor/transistion to priors

Heirarchical Priors

Prediction

Model Exploration & Averaging

Results

- General Products
- Degree of smoothing (heirarchical parameters)
- Posterior v. Current
 - Report degree of similarity
- Prediction v. Data
 - Report predictive accuracy

Beta–Binomial Posterior Predictive Species Compositions

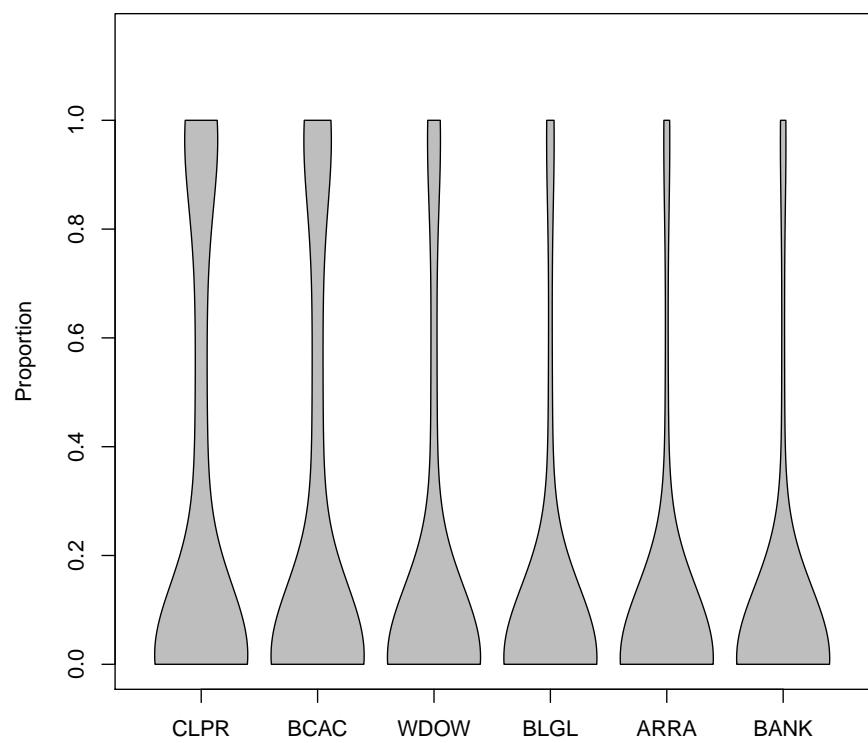


Figure 2: overdisp vio

Conclusions

- General Math/Science
- Database Stuff
- Looking Forward
 - forecasting/hindcasting
 - * simple
 - * timeseries models
 - more computation faster
 - * broader model exploration
 - * broader spatial expansion

References

- [1] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.
- [2] Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382-401.
- [3] McCullagh P. & Nelder, J.A. (1989). Generalized Linear Models, 2nd ed. London: Chapman and Hall.
- [4] Pearson, D.E., and Erwin, B. (1997). Documentation of California's commercial market sampling data entry and expansion programs. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-240.
- [5] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [6] Rue H., Martino S., Lindgren F., Simpson D., Riebler A. (2013). R-INLA: Approximate Bayesian Inference using Integrated Nested Laplace Approximations. Trondheim, Norway. URL <http://www.r-inla.org/>.
- [7] Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2), 319-392.
- [8] Sen, A.R. (1984). Sampling commercial rockfish landings in California. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-45.
- [9] Sen AR. (1986). Methodological problems in sampling commercial rockfish landings. *Fish Bull.* 84: 409-421 .
- [10] Shelton, A. O., Dick, E. J., Pearson, D. E., Ralston, S., & Mangel, M. (2012). Estimating species composition and quantifying uncertainty in multispecies

fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data. Canadian Journal of Fisheries and Aquatic Sciences, 69(2), 231-246.