# 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

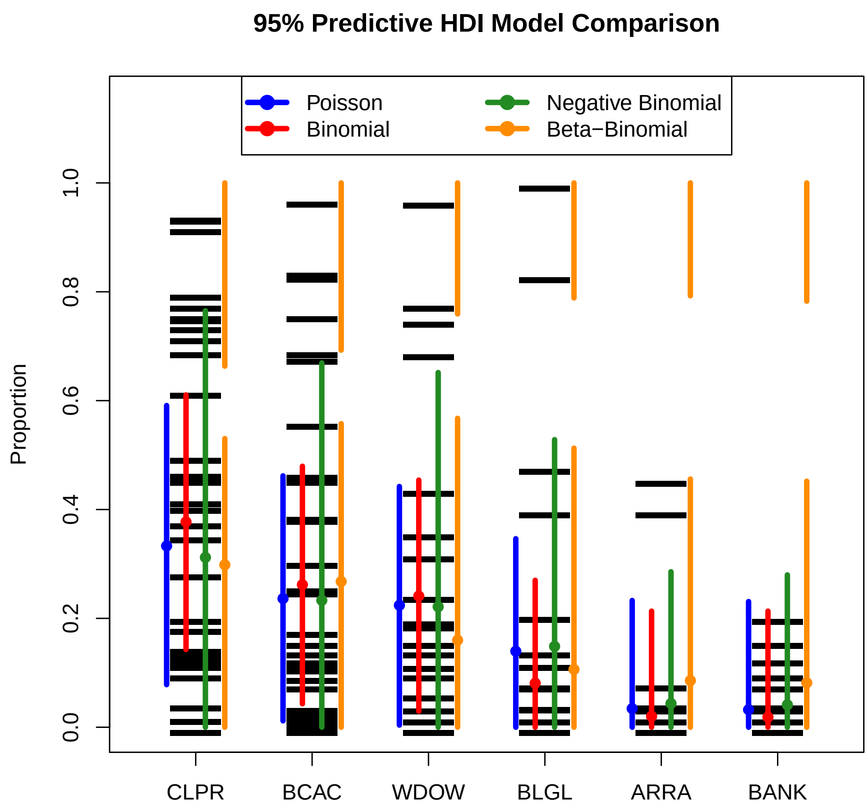
Something something heirarchical poisson model. Something something (Shelton, 2012).

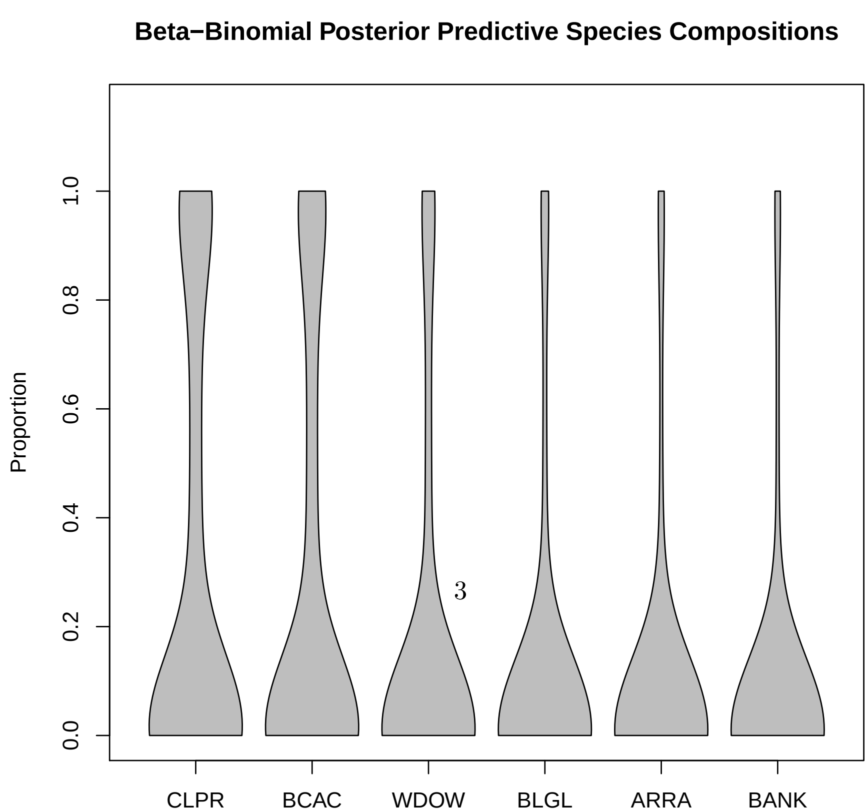
For the purposes of accuratly modeling not only species composition means, but also higher moments of the data, such as species composition variances, it is neccisary to recognize model limitations with respect to over-dispersed data. Amoung the simplest models for count data are the poisson and binomial models. Both models are typically specificed 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 accuratly 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 numerious 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 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 Montery 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. For brevity we only consider for the top six species here (CLPR, BCAC, WDOW, BLGL, ARRA, BANK). These data are serially fit by each of the Poisson, binomial, negative binomial, and beta-binomial models, predictive species compositions are then derived from each model, plotted along side the observed species compositions, and metrics of model fits are compared.


95% Predictive HDI Model Comparison
Poisson
Binomial
Negative Binomial
Beta−Binomial
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
CLPR
BCAC
WDOW
BLGL
ARRA
BANK


Beta−Binomial Posterior Predictive Species Compositions
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
CLPR
BCAC
WDOW
BLGL
ARRA
BANK

| | Poisson | Binomial | Negative-Binomial | Beta-binomial |
|---|---|---|---|---|
| DIC WAIC | 5675.25 | 6759.86 | 1301.51 1302.19 | 1261.00 |
| log $p(y\|M)$ | 5840.56 | 6939.74 | -688.19 | 1261.30 |
| | -2864.01 | -3406.01 | | -650.49 |

- describe picture
- notice overdispersion
- impropper variance model biases mean, beta-binomial is flexible to disentagles these effects.
- maths stuffs
  - mean function
  - variance; introduct $\rho$
- 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

## Conclusions

- General Math/Science
- Database Stuff
- Looking Forward
  - forcasting/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.