

Properties of the Dirichlet-multinomial distribution

When fitting compositional data in assessments

Goal

- Evaluate the relative performance of three alternative composition likelihood models:
 - Dirichlet
 - Linear Dirichlet-multinomial
 - Saturated Dirichlet-multinomial
- Support better selection of composition likelihoods in integrated fishery stock assessment models.
- Use function-level diagnostics to assess estimator quality across appropriate regions of the probability simplex for observed composition data.

Problem

- Traditional modeling of age or size composition data often fails to account for:
 - **Overdispersion** beyond the multinomial assumption.
 - **Residual correlation** between composition bins.
 - **Intracluster correlation** among composition data samples.
- This can lead to biased estimates, underestimated uncertainty, and poor model fit in stock assessments.

Solution

- Implement a **functional analysis framework in C++** to:
 - Sample observed proportions from regions of the **probability simplex**.
 - Evaluate each model using:
 - Log-likelihood values
 - Stochasticity and variance metrics
 - Sensitivity diagnostics from **gradients and Hessians**
 - Residual correlation structure
 - Effective sample size estimation
 - Bias in recovered proportions
- Provide **practical guidance** on the selection of composition model likelihoods based on robustness, informativeness, and computational stability.