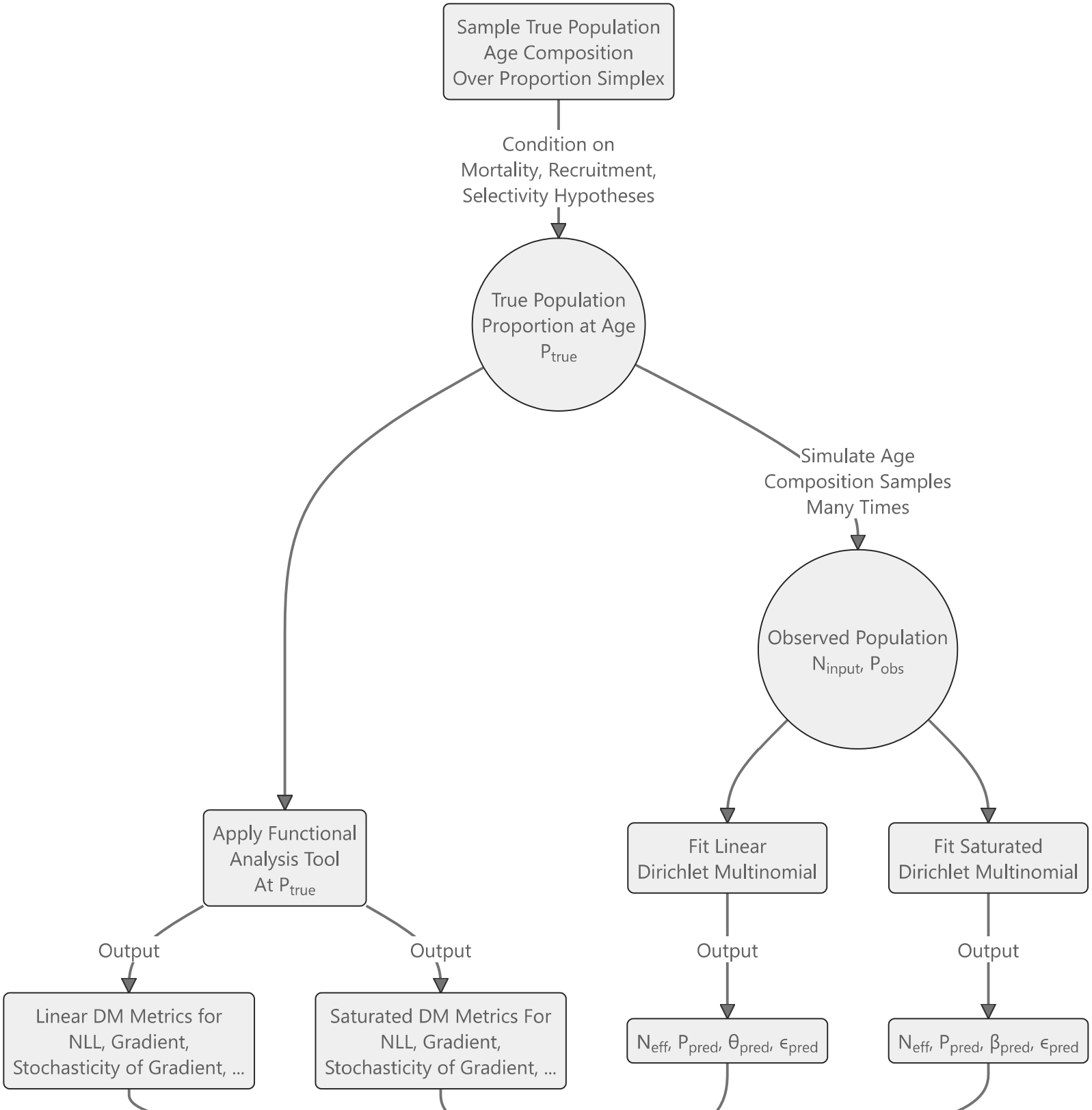
Which Dirichlet-Multinomial Parameterization - Linear or Saturated - Is Better for Fitting Compositional Data in Stock Assessment Applications?

An Algorithmic Approach to Evaluate Relative Performance

Here's an initial algorithm to evaluate relative performance of the linear and saturated Dirichlet-multinomial likelihoods for fitting age composition data over the set of possible proportions at age at a given time.

- Sample the true age composition from the proportion simplex grid
  - To Be Determined ~ Condition the acceptance of the sample on the fishery system characteristics, like fishery selectivity, total mortality, and recruitment patterns
- Apply the functional analysis tool to analyze the likelihood surface in the neighborhood of the true age composition point
- Simulate-Estimate age compositions under each Dirichlet-multinomial distribution
  - Simulate observed age composition(s) with error for fixed sample size(s)
  - Estimate predicted age compositions with both distributions
- Output performance metrics from functional analysis for both distributions
- Output predicted age compositions and goodness-of-fit metrics for both distributions
- Repeat for the next grid point in the simplex

The following diagram illustrates the workflow for simulating and analyzing age composition data. Starting from a true population distribution of proportions-at-age (P_{true}), simulated age composition samples are generated to represent observed proportions-at-age (P_{obs}). These observations are then analyzed using a functional analysis tool and two estimation models: a **linear Dirichlet-multinomial model** and a **saturated Dirichlet-multinomial model**, each producing predicted proportions (P_{pred}), dispersion parameters (θ_{pred}, β_{pred}), effective sample sizes (N_{eff}) and other performance metrics.


Sample True Population Age Composition Over Proportion Simplex
Condition on Mortality, Recruitment, Selectivity Hypotheses
True Population Proportion at Age P_{true}
Simulate Age Composition Samples Many Times
Observed Population N_{input}, P_{obs}
Apply Functional Analysis Tool At P_{true}
Fit Linear Dirichlet Multinomial
Fit Saturated Dirichlet Multinomial
Output
Output
Output
Output
Linear DM Metrics for NLL, Gradient, Stochasticity of Gradient, ...
Saturated DM Metrics For NLL, Gradient, Stochasticity of Gradient, ...
N_{eff}, P_{pred}, θ_{pred}, ϵ_{pred}
N_{eff}, P_{pred}, β_{pred}, ϵ_{pred}


Synthesize and Summarize Results