

Recovering Latent Age Distributions from Misclassified Observations: A Simulation Study Using EM and Method of Moments

Emilio Otero Johansson*

May 2026

Abstract

Accurate estimation of population age structure is important in wildlife ecology, but age is often measured with error. This thesis studies how to recover a latent age distribution from misclassified observations using statistical deconvolution in a simulation setting. The true population structure is generated from a Leslie matrix model, while measurement error is introduced through a misclassification mechanism.

Two estimators are evaluated: a method-of-moments (MoM) estimator and a maximum likelihood estimator implemented via the Expectation–Maximization (EM) algorithm. Their performance is assessed through Monte Carlo simulations under different levels of measurement error and sample size, using total variation distance and population forecasts. Results are compared to an oracle baseline and to the naive estimator based on misclassified observations.

The results show that under ill-posed conditions the EM estimator reduces bias but has higher variability, while the naive estimator and MoM are more stable but systematically biased. The study shows that even small bias in age structure can lead to large errors in population forecasts, highlighting that improvements in statistical performance do not necessarily translate into improved ecological performance.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: emiliooterojohansson@gmail.com. Supervisor: Martin Sköld.