# Stochastic EM - A closely related cousin to MCEM

Wakeel Adekunle Kasali
wakeel.kasali@stat.ubc.ca

Fall 2024

## Topics in Probability (STAT 547C)

## Project Outline

## 1 Description

The Stochastic Expectation Maximization (SEM) algorithm is a variant of the traditional Expectation Maximization (EM) algorithm. It introduces stochasticity to manage complex data settings. This modification addresses cases where exact computation of the expected complete-data log-likelihood is computationally prohibitive. In the standard EM algorithm, each iteration has two main steps. First is the E-step. This step calculates

$$\mathbb{E}_{X|Y,\theta^{(k)}}\left[\ell_c(\theta; Y, X)\right].$$

This represents the expected value of the complete-data log-likelihood. It is based on the observed data $Y$ and the current parameter estimate $\theta^{(k)}$. Next is the M-step, which maximizes this expectation with respect to $\theta$. In SEM, however, the E-step does not compute the full expectation. Instead, a sample is drawn from the conditional distribution $X|Y, \theta^{(k)}$. This sample approximates the expected complete-data log-likelihood, lowering computational demands [1, 2].

The approximation replaces the deterministic E-step with a stochastic version. Each sample introduces randomness to the iterative process. Let the complete-data log-likelihood be $\ell_c(\theta; Y, X)$. The SEM algorithm approximates the E-step with $\ell_c(\theta; Y, X^{(k)})$, where $X^{(k)} \sim P(X|Y, \theta^{(k)})$. This method is similar to the Monte Carlo EM (MCEM) algorithm. In MCEM, a Monte Carlo sample estimates the expectation in the E-step. However, SEM uses only a single sample per iteration, not a large sample, to achieve convergence. This creates a stochastic approximation of the EM objective function. Effectively, it forms a Markov chain that converges to the target maximum likelihood estimator (MLE) under suitable conditions.

This project aims to analyze the effectiveness and computational trade-offs of SEM. It compares SEM with EM and MCEM in various simulation settings. The focus is on cases where computing the full E-step is infeasible. By applying SEM in these scenarios, the project will explore its strengths and limitations.

## 2 Literature and Technical Aspects

SEM's convergence relies on the principles of stochastic approximation. Specifically, it follows the approach of [3]. The iterative sampling in SEM forms a Markov chain. Under suitable conditions, this chain converges to the Maximum Likelihood Estimator (MLE). Unlike EM, which converges deterministically, SEM includes a stochastic component. This component introduces variability in each iteration.

The convergence properties of SEM are based on stochastic approximation theory. This theory provides conditions for the sequence $\theta^{(k)}$ to converge to the MLE of the observed data likelihood. Each E-step in SEM introduces randomness, generating iterates that vary around the target MLE. However, this variability decreases with more iterations and proper tuning of sampling parameters. Unlike the deterministic convergence of EM or stable approximations in MCEM, SEM's convergence includes a probabilistic component. Therefore, step size and sample variability require careful adjustment. Convergence rates can also depend on the "fraction of missing information." This fraction measures information loss about $\theta$ due to unobserved data, denoted as $I_m(\theta)/I_o(\theta)$, where $I_m$ and $I_o$ are the information matrices for missing and observed data, respectively [4].

SEM is implemented using various statistical programming tools. R packages like `mclust` and `mixtools` support EM and related algorithms. These packages are suitable for adapting to SEM's stochastic E-step. Additionally, `dplyr` from the Tidyverse collection aids data manipulation in iterative sampling.

At this moment, we are six weeks away from the due date. Therefore, the following is the weekly plan to carry out the project.

# 3 Plan

1. Week 1: Further literature review.
2. Week 2: Start of coding development.
3. Week 3: Continuation of coding.
4. Week 4: Expected end of coding phase and the start of the final draft.
5. Week 5: Final eraser finishing.
6. Week 6: Review and final details.

# References


[1] M. Gu and F. Kong, "A stochastic approximation algorithm with markov chain monte-carlo method for incomplete data estimation problems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 13, p. 7270—7274, June 1998. [Online]. Available: https://europepmc.org/articles/PMC22587

[2] B. Delyon, M. Lavielle, and E. Moulines, "Convergence of a stochastic approximation version of the em algorithm," *The Annals of Statistics*, vol. 27, no. 1, pp. 94–128, 1999. [Online]. Available: http://www.jstor.org/stable/120120

[3] H. E. Robbins, "A stochastic approximation method," *Annals of Mathematical Statistics*, vol. 22, pp. 400–407, 1951. [Online]. Available: https://api.semanticscholar.org/CorpusID:16945044

[4] X.-L. Meng and D. B. Rubin, "On the global and componentwise rates of convergence of the em algorithm," *Linear Algebra and its Applications*, vol. 199, pp. 413–425, 1994, special Issue Honoring Ingram Olkin. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0024379594903638