A Modified Minibatch Sampling Method for Parameter Estimation in Hidden Markov Models using Stochastic Variational Bayes

Reetam Majumder[1]*, **Matthias K. Gobbert**[1], and **Nagaraj K. Neerchal**[1,2]

[1] Department of Mathematics and Statistics, University of Maryland, Baltimore County, USA
[2] Chinmaya Vishwavidyapeeth, Kerala, India

Parameter estimation using stochastic variatonal Bayes (SVB) under a mean field assumption can be carried out by sampling a single data point at each iteration of the optimization algorithm. However, when latent variables are dependent like in hidden Markov models (HMM), a larger sample is required at each iteration to capture that dependence. We describe a minibatch sampling procedure for HMMs where the emission process can be segmented into independent and identically distributed blocks. Instead of sampling a block and using all elements within it, we divide the block into subgroups and sample subgroups from different blocks using simple random sampling with replacement. Simulation results are provided for an HMM for precipitation data, where each block of 90 days represents 3 months of wet season data. SVB based on the proposed sampling method is shown to provide parameter estimates comparable with existing methods.

1 Introduction

Variational Bayes (VB) is an optimization approach to posterior estimation in Bayesian inference [1], and provides an alternative to Markov chain Monte Carlo methods. Most VB applications make a mean field assumption, which specifies the approximate variational posterior of all parameters and latent variables as a product of the distributions of its individual components. Parameters can then be estimated using a VB version of the Expectation Maximization (VBEM) algorithm. However, each iteration of the VBEM requires computing means over the complete data and results in a performance bottleneck for large datasets. Stochastic optimization methods provide us a way around this, and stochastic variational Bayes (SVB) [2] is one such approach which implements VBEM as a stochastic gradient ascent algorithm for each parameter. Instead of computing gradients based on the entire data, SVB uses an unbiased estimate of the gradient at each iteration. The SVB algorithm converges to a local optimum as long as the step sizes for the gradient ascent satisfy the Robbins-Monro conditions [3].

Under the mean field assumption, an unbiased estimator of the gradient can be constructed using a single observation. Now consider a hidden Markov model (HMM) $\{S_t, Y_t\}_{t \geq 0}$ where $\{S_t\}$ is a Markov chain, and conditional on it, $\{Y_t\}$ is a sequence of independent random variables such that the distribution of Y_t depends only on S_t [4]. The mean field assumption does not hold for HMMs, and a single sample point (S_i, Y_i) cannot be used to estimate the transition probabilities of $\{S_t\}$. Consequently, a sample consisting of a sequence of observations is required to estimate the variational posterior of the parameters of $\{S_t\}$. We denote this sequential sample, or minibatch, as y^*. The nature of the dataset dictates the procedure for selecting y^*. If the data consists of a single long chain, [5] proposed subsampling from the chain and buffering the beginning and end with extra observations to preserve the Markov properties of the states. If however the data is seasonal or cyclical in nature that can be represented as N blocks each of size D, then a minibatch is constructed at each optimization iteration by randomly sampling blocks with replacement and selecting all D time points within the block. This approach is discussed in [6]. In both cases, the variational E-step employs the Forward-Backward algorithm [4], and the variational M-step often takes advantage of conjugate priors and provides parameter updates through stochastic gradient ascent.

For the second case where the data admits a block structure, a concern arises from having to run a large number of optimization iterations using a relatively small number of blocks of data. If we want to select a minibatch of D time points from N blocks of data at each iteration, this is equivalent to sampling with replacement from $1, \ldots, N$. The value of N is often not large in practice, and the approach in [6] is not ideal since there might not be a lot of variability within the samples between different iterations. We propose an alternative method which leverages the exchangeability inherent in HMMs.

2 Minibatch Sampling Algorithm for SVB Parameter Estimation in HMMs

We consider daily precipitation data for a single season as our motivating and demonstrative example, which is observed as a semi-coninuous emission process and commonly modeled using HMMs [7]. We note a break in the collection of data between the end of the season in a year and the beginning of the season the next year. Each year's data therefore constitutes blocks which are distributed independently and identically. Further, for N years' data with D days in each year, data for the dth day of every year has the same distribution. In particular, if we think of the D days as coming from C months, the cth month

* Corresponding author: e-mail reetam1@umbc.edu

has the same distribution for all years. With this exchangeability of days between years in mind, we propose the following algorithm for constructing minibatches:

**Algorithm 1** Minibatch sampling in Stochastic Variational Bayes for HMMs
---
1. Divide the D days in each year into C months
1. Draw a sample $s_1, \ldots, s_C$ of size C from $1, \ldots, N$ with replacement
2. If $s_c = i$, the cth month of the ith year provides data for the cth month of the minibatch


Instead of N possible unique minibatches of size D, our approach allows for N^C unique minibatches of size D. An extreme extension of this would be sampling each of the D days from the N different years with replacement; however we found that to be detrimental in estimating the transition probability parameters. The minibatch y^* can now be used for SVB.

## 3 Simulation Study for Daily Precipitation at a Single Location

Our simulation study has the same setup as in [7]. We compared the computational cost and accuracy of the old and our new approach for minibatch selection. The HMM for precipitation is described as $\{S_t, R_t, Y_t\}_{t \geq 0}$, where Y_t has a semi continuous distribution and R_t is the corresponding mixture distribution. The synthetic data comprises 90 days of data for 20 years. The $D = 90$ days are divided into $C = 3$ months with each month being 30 days. The parameter space is given by $\Theta = (\alpha, \zeta, \Lambda)$ corresponding to the 3 variables, where α, ζ and Λ are matrices. We constructed minibatches as described in Algorithm 1 using samples of size 3. The difference in computational time between the new and old methods was below the measurement threshold, pointing to no additional computational cost. To compare the accuracy from the 2 methods, we computed 1000
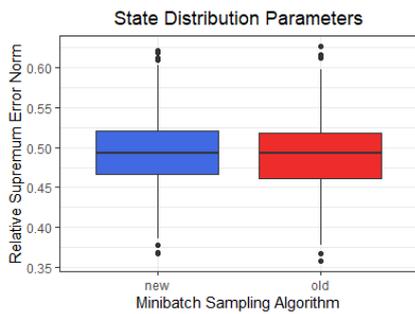
Fig. 1: Distribution of the relative supremum error norms when estimating α based on the old and new method.
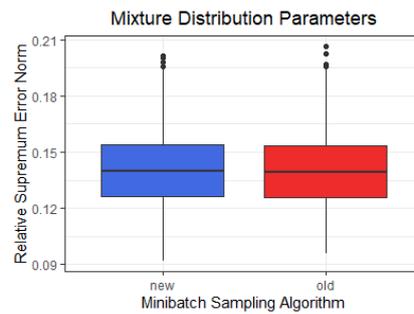
Fig. 2: Distribution of the relative supremum error norms when estimating ζ based on the old and new method.
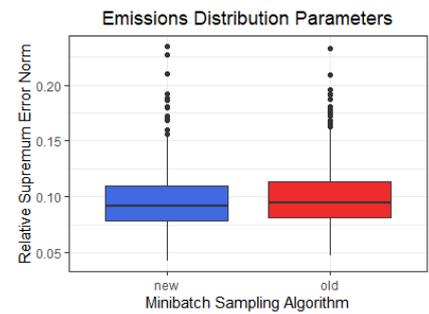
Fig. 3: Distribution of the relative supremum error norms when estimating Λ based on the old and new method.

independent estimates $\hat{\Theta} = (\hat{\alpha}, \hat{\zeta}, \hat{\Lambda})$, where each estimate is based on 1000 iterations of the SVB. The relative supremum error norms $\frac{\|\hat{\alpha} - \alpha\|_\infty}{\|\alpha\|_\infty}$, $\frac{\|\hat{\zeta} - \zeta\|_\infty}{\|\zeta\|_\infty}$, and $\frac{\|\hat{\Lambda} - \Lambda\|_\infty}{\|\Lambda\|_\infty}$ are calculated at the end of the SVB for both methods. Figures 1–3 show the distribution of the relative supremum error norms for the 1000 estimates, with the old minibatch algorithm represented in red and the new one in blue. Both methods provide very similar estimates of α and ζ, but SVB has does not perform well in estimating the Markov process parameters based on either minibatch. Algorithm 1 does provide better estimates for Λ due to its conditional independence given the state. Our work seems to indicate that there is merit in devising a customized minibatch sampling method that takes advantage of the data structure to strike a balance between greater variability in our samples for SVB and the need to reproduce the estimates obtained from the entire data. Future work will focus on the quality of the estimates produced by different SVB minibatch algorithms compared to VB estimates from the entire data.

References

[1] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, Mach. Learn. **37**(2), 183–233 (1999).
[2] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, J. Mach. Learn. Res. **14**(4), 1303–1347 (2013).
[3] H. Robbins and S. Monro, Ann. Math. Stat. **22**(3), 400 – 407 (1951).
[4] L. R. Rabiner, Proc. IEEE **77**(2), 257–286 (1989).
[5] N. Foti, J. Xu, D. Laird, and E. Fox, Stochastic variational inference for hidden Markov models, in: NIPS, (2014).
[6] M. J. Johnson and A. S. Willsky, Stochastic variational inference for Bayesian time series models, in: ICML, (2014), pp. 1854–1862.
[7] R. Majumder, M. K. Gobbert, A. Mehta, and N. K. Neerchal, Variational Bayes estimation of hidden Markov models for daily precipitation with semi-continuous emissions, Tech. Rep. HPCF–2021–8, University of Maryland, Baltimore County, 2021.