Statistical Inference of Hidden Markov Models on High Frequency Quote Data

AMS 518

Jason Bohne
Applied Mathematics and Statistics Department
Stony Brook University

November 3, 2022

Overview

1. Hidden Markov Models
2. Statistical Inference
3. Data Preparation
4. Numerical Results
5. Conclusion

Model Formulation

- Discrete time stochastic models characterized by unknown states $(X_1 \dots X_n)$
- X is assumed to be a Markov process, with hidden states $S_1 \dots S_k$
- States specify the underlying distribution $(Y_1 \dots Y_n)$
- Observations $(y_1 \dots y_n)$ are samples from the underlying distributions
- Fully specified by transition matrix and distribution of hidden and initial states

$$A = P(X_t | X_{t-1} = S_j)$$
$$B_j = P(y_t | X_t = S_j)$$
$$\pi = P(X_0)$$

Likelihood Function

- Model likelihood function of observations conditional on parameters

$$l(\theta) = P(O|\theta)$$

- Conditional probability of transitioning between state S_i and state S_j at time t

$$\varepsilon_t(i,j) = P(X_{t+1} = S_j, X_t = S_i | O, \theta)$$

- Probability of being in state S_j can be computed by accounting for all states S_i

$$\gamma_j(t) = \sum_{i=1}^{k} \varepsilon_t(i,j)$$

- Utilize above probabilities to determine update procedure when optimizing likelihood

Baum-Welch Algorithm

- Maximum likelihood procedure to determine optimal estimates of A, B_j, π
- Solution to optimization will be the local maximum of likelihood function
- Computationally designed as an Expectation-Maximization algorithm
- Updates estimate each iteration based on reestimation rules

$$\hat{A} = \sum_{t=1}^{n-1} \frac{\varepsilon_t(i,j)}{\gamma_t(i)}$$

$$\hat{B}_j(k) = \frac{\sum_{t=1}^{n-1} \gamma_t(j) I_{O_t=k}}{\sum_{t=1}^{n-1} \gamma_t(j)}$$

$$\hat{\pi} = \gamma_1(i)$$

Viterbi Algorithm

- Utilizes dynamic programming to determine optimal *kth* subsequence of hidden states
- Partitions original model likelihood into sequence of recursive functions
- Problem traced backward in time to determine optimal initial state
- Initial solution iteratively applied to solve the problem forwards in time

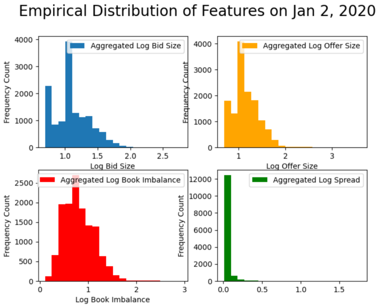
$$l(\hat{\theta}) = l_1(\hat{\theta}) + l_1^2(\hat{\theta}) + \ldots + l_{n-1}^n(\hat{\theta})$$

Data Preparation

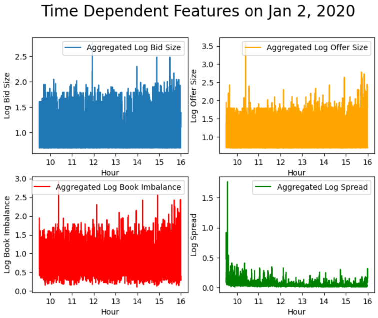
- Original dataset of 1.8 million AAPL top of the book quotes for January 2020
- Top of the book quotes composed of best bid and best offer price and size
- Standard preprocessing to remove invalid observations from dataset
- Generated features relevant to limit orderboook
- Log transformations applied on raw features to lessen effect of outliers
- Transformed features aggregated into one-second bars by taking sample mean

$$F = \{BS_i, OS_i, OB_i = \frac{OS_i}{BS_i}, S_i = OP_i - BP_i\}$$

Exploratory Data Analysis



(a) Empirical Distribution of Features



(b) Time Dependence

Setup

- Fit Hidden Markov model for each feature across single day of high-frequency data
- Assume each day is $i.i.d.$ and repeat procedure over full month of data
- Sample means and standard deviations of parameter estimates are included
- P-values computed for two sample t-test under hypothesis for equal population means

Setup

- Inference problem can be solved as an optimization problem of the model likelihood
- PSG uses Baum-Welch as initial guess and solves via constrained optimization
- Hmmlearn uses Viterbi algorithm to optimize parameters given observations
- Constraints on transition matrix A and initial state distribution π

$$\sum_{i=1}^{k} \sum_{j=1}^{k} A_{ij} = 1 \ \forall A_{ij} \geq 0$$
$$\sum_{i=1}^{k} \pi_i = 1 \ \forall \pi_i \geq 0$$

PSG Results

Features	a_{11}	a_{12}	a_{21}	a_{22}	mu_1	si_1	mu_2	si_2
BS	0.7756	0.2243	0.0573	0.9426	1.0700	0.2195	1.4669	0.3708
OS_i	0.7998	0.2001	0.1048	0.8951	1.1015	0.2153	1.4351	0.3566
OB_i	0.8594	0.1405	0.1046	0.8953	0.6567	0.2092	1.0311	0.3188
S_i	0.8062	0.1937	0.1335	0.8664	0.0606	0.0208	0.1818	0.1225

Table: Sample Mean of PSG Parameter Estimates

Features	a_{11}	a_{12}	a_{21}	a_{22}	mu_1	si_1	mu_2	si_2
BS	0.2156	0.2156	0.0412	0.0412	0.0982	0.0236	0.2063	0.1694
OS_i	0.2660	0.2660	0.2182	0.2182	0.1072	0.0609	0.2886	0.1951
OB_i	0.0842	0.0842	0.0663	0.0663	0.0668	0.0219	0.1297	0.1100
S_i	0.1056	0.1056	0.0988	0.0988	0.0671	0.0340	0.1334	0.0639

Table: Sample Standard Deviation of PSG Parameter Estimates

Hmmlearn Results

Features	a_{11}	a_{12}	a_{21}	a_{22}	mu_1	si_1	mu_2	si_2
BS_i	0.9201	0.0798	0.1131	0.8868	1.0225	0.0478	1.3749	0.1196
OS_i	0.6505	0.3494	0.3655	0.6344	1.0756	0.0332	1.262011	0.0992
OB_i	0.8429	0.1570	0.1583	0.8416	0.6687	0.0473	1.0231	0.1143
S_i	0.8857	0.1142	0.2120	0.7879	0.0606	0.0015	0.1822	0.0189

Table: Sample Mean of Hmmlearn Parameter Estimates

Features	a_{11}	a_{12}	a_{21}	a_{22}	mu_1	si_1	mu_2	si_2
BS_i	0.0893	0.0893	0.1307	0.1307	0.0916	0.0180	0.2086	0.1454
OS_i	0.3884	0.3884	0.3933	0.3933	0.0845	0.0278	0.0419	0.1108
OB_i	0.1705	0.1705	0.1811	0.1811	0.0738	0.0166	0.1376	0.1170
S_i	0.0764	0.0764	0.1097	0.1097	0.0671	0.0060	0.1333	0.0225

Table: Sample Standard Deviation of Hmmlearn Parameter Estimates

Algorithm Comparison

- Would not reject hypothesis that population means are equal at 10% significance
- Supports argument that both methods parameterize underlying distribution similarly

Features	p_1	p_2
BS_i	0.4092	0.4100
OS_i	0.3982	0.4034
OB_i	0.5034	0.4864
S_i	0.5124	0.7450

Table: Average p-values for two sample t-test across implementations

Conclusion

- Both methods estimate true distribution supported with two-sample t-test
- Advantage as allows for greater flexibility in choice of algorithm to solve optimization
- Disadvantage is ambiguity in the choice of hyperparameters in model specification
- Further investigation required to determine optimal number of states for each feature
- Extension to multivariate observations composed of features mentioned

Code Repository

To access the data and scripts to reproduce examples, refer to the Github,
https://github.com/jasonbohne123/HMM_LOB_Inference

References

A tutorial on hidden Markov models and selected applications in speech recognition
Rabiner, L.R. *Proceedings of the IEEE* 1989
https://ieeexplore.ieee.org/document/18626

Case Study: Maximization of Log-Likelihood in Hidden Markov Model
Stan Uryasev 2022

Statistical Inference for Probabilistic Functions of Finite State Markov Chains
Leonard E. Baum and Ted Petrie. *The Annals of Mathematical Statistics* 1966

Error bounds for convolutional codes and an asymptotically optimum decoding algorithm
Viterbi, A. *IEEE Transactions on Information Theory* 1967,
https://ieeexplore.ieee.org/document/1054010

Unsupervised learning and inference of Hidden Markov Models
Hmmlearn Developers 2022,
https://github.com/hmmlearn/hmmlearn

Portfolio Safeguard, American Optimal Decisions, Inc. 2022,
http://www.aorda.com/