

# Inferring Markov Chain for Modeling Order Book Dynamics in High Frequency Environment

Yuan Lung Chang

**Abstract**—In this paper, we propose a Bayesian inference of the Markov chain model class to model dynamics of order book in high frequency trading environment. Accordingly, software program can predict the move of market price for both ask & bid via predictive distribution. A strategy algorithm can be developed for generating, routing & executing orders to gain profit. Experimental result based on security AAPL showed over 98% coverage by 50 transitions from 6561 state space. It further indicated market behavior of short time-frame can be clustered & labeled.

**Index Terms**—Inferring markov chain, bayesian inference, high frequency trading, order book.

## I. INTRODUCTION

Stock exchanges have undergone a transformation. Today's trading, offer, bid, order all go through computer systems; the processes become fully automated. Meanwhile, the trading desk has adopted the same approach as Algorithmic & computer-assisted trading. HFT (High Frequency Trading) is the subset of this trend with the focus on speed & short time-frame. It has been recognized as one of the most significant market structure developments in recent years [1]-[3]. Estimation of HFT typically exceeds 50% of total volume in U.S.-listed equities [1].

Inference of Markov chains has a long and successful history in mathematical statistics [4]. In this paper, we propose a Bayesian inference of the Markov chain model class to model dynamics of order book in high frequency trading environment. This model provides the foundation of high-speed computation for algorithmic trading. Software can predict the change of short time-frame based on the predictive distribution, further generating, routing & executing orders. It's one of the keys to succeed for sophisticated HFT security traders.

In a well regulated market, high frequency doesn't have a designated role with special data privileges [3]. Specifically, the proposed model is verified according to time-frames of 10 milliseconds. Other than millisecond resolution, our model didn't expect public-unknown information. Experimental result based on security AAPL showed over 98% coverage by 50 transitions from total 6561 state space. It further indicated stock market behavior of short time-frame can be clustered & labeled.

## II. INFERRING MARKOV CHAINS

We develop a systematic relation between the data  $D$ , the chosen model  $M$  and the vector of model parameters  $\theta$ . Bayes' theorem expresses the relation as [4]:

$$P(\theta | D, M) = \frac{P(D | \theta, M)P(\theta | M)}{P(D | M)} \quad (1)$$

Accordingly, the object of interest in the inference of model parameters is the posterior probability density  $P(\theta | D, M)$ ; the prior  $P(\theta | M)$  specifies our assumptions regarding the model parameters; the likelihood  $P(D | \theta, M)$  describes the probability of the data given the model; and the evidence (or marginal likelihood)  $P(D | M)$  is the probability of the data given the model.

### A. Markov Chain

We assume  $D$  is a single data set of length  $N$ , which consists of symbols  $s_t$  from a finite alphabet  $A$  as starting point of inference [4],

$$D = s_0 s_1 s_2 s_3 \dots s_{N-1} \quad s_t \in A \quad (2)$$

A generalization of the Markov property from first order model class assumes finite memory & stationary in the data source. It can be further written as only depending on preceding word [4],

$$p(D) = p(s_0) \prod_{t=0}^{N-2} p(s_{t+1} | s_t) \quad (3)$$

The stationary condition for any  $(t, m)$  can be expressed as

$$p(s_t | s_{t-1}) = p(s_{t+m} | s_{t+m-1}) \quad (4)$$

Accordingly, we no longer need to track the position of index [4]. The model parameter of first order Markov chain  $M$  is

$$\theta = \{p(s_t | s_{t-1}) \quad s_t \in A\} \quad (5)$$

With normalization constraint is applied as

$$\sum_{s_t \in A} p(s_t | s_{t-1}) = 1 \quad (6)$$

### B. Bayes' Theorem to Markov Chain

Based on Eq. (3) & Eq. (4), the likelihood results in the form

$$P(D | \theta, M) = \prod_{s_t \in A} \prod_{s_{t-1} \in A} p(s_t | s_{t-1})^{n(s_{t-1}, s_t)} \quad (7)$$

Manuscript received October 15, 2014; revised January 23, 2015.

Yuan Lung Chang is with New York, NY10128 USA (e-mail: jeffrey.chang.dc@gmail.com)

where  $n(s_{t-1}s_t)$  is the number of times the word  $s_{t-1}s_t$  occurs in the sample data  $D$  [4].

The prior  $P(\theta|M)$  is adopted to specify assumptions about the model to be inferred before the data is considered. Here we use conjugate priors (the same functional form as posterior distribution). Accordingly, we have

$$P(\theta|M) = \prod_{s_{t-1} \in A} \left\{ \frac{\Gamma(\alpha(s_{t-1}))}{\prod_{s_t \in A} \Gamma(\alpha(s_{t-1}s_t))} \times \delta(1 - \sum_{s_t \in A} p(s_t | s_{t-1})) \times \prod_{s_t \in A} p(s_t | s_{t-1})^{\alpha(s_{t-1}s_t)-1} \right\} \quad (8)$$

where  $\alpha(s_{t-1}s_t)$  is the hyperparameters to prior (must be real & positive);  $\alpha(s_{t-1})$  is defined as  $\sum_{s_t \in A} \alpha(s_{t-1}s_t)$ ; and  $\Gamma(x) = (x-1)!$  is Gamma function [4]. The  $\delta$  function constrains the model parameters to be properly normalized for each  $s_{t-1}$  as [4]

$$\sum_{s_t \in A} p(s_t | s_{t-1}) = 1 \quad (9)$$

Prior's mean & variance can be expressed as [4]

$$E_{prior}[p(s_t | s_{t-1})] = \frac{\alpha(s_{t-1}s_t)}{\alpha(s_{t-1})} \quad (10)$$

$$Var_{prior}[p(s_t | s_{t-1})] = \frac{\alpha(s_{t-1}s_t)(\alpha(s_{t-1}) - \alpha(s_{t-1}s_t))}{\alpha(s_{t-1})^2(1 + \alpha(s_{t-1}))} \quad (11)$$

Evidence  $P(D|M)$  is seen to be a simple normalization term in Bayes' theorem. Formally, its definition is

$$P(D|M) = \int d\theta P(D|\theta, M) P(\theta|M) = \prod_{s_{t-1} \in A} \left\{ \frac{\Gamma(\alpha(s_{t-1}))}{\prod_{s_t \in A} \Gamma(\alpha(s_{t-1}s_t))} \times \frac{\prod_{s_t \in A} \Gamma(n(s_{t-1}s_t) + \alpha(s_{t-1}s_t))}{\Gamma(n(s_{t-1}) + \alpha(s_{t-1}))} \right\} \quad (12)$$

where we see this term can be interpreted as an average of the likelihood (Eq. (7)) over the prior distribution (Eq. (8)) [4].

By knowing likelihood, prior & evidence, posterior distribution becomes [4]

$$P(\theta|D, M) = \prod_{s_{t-1} \in A} \left\{ \frac{\Gamma(n(s_{t-1}) + \alpha(s_{t-1}))}{\prod_{s_t \in A} \Gamma(n(s_{t-1}s_t) + \alpha(s_{t-1}s_t))} \times \delta(1 - \sum_{s_t \in A} p(s_t | s_{t-1})) \times \prod_{s_t \in A} p(s_t | s_{t-1})^{n(s_{t-1}s_t) + \alpha(s_{t-1}s_t) - 1} \right\} \quad (13)$$

Accordingly, we can obtain posterior mean estimate (PME) as [4]

$$E_{post}[p(s_t | s_{t-1})] = \frac{n(s_{t-1}s_t) + \alpha(s_{t-1}s_t)}{n(s_{t-1}) + \alpha(s_{t-1})} \quad (14)$$

$$Var_{post}[p(s_t | s_{t-1})] = \frac{n(s_{t-1}s_t) + \alpha(s_{t-1}s_t)}{(n(s_{t-1}) + \alpha(s_{t-1}))^2} \times \frac{(n(s_{t-1}) + \alpha(s_{t-1})) - (n(s_{t-1}s_t) + \alpha(s_{t-1}s_t))}{(n(s_{t-1}) + \alpha(s_{t-1})) + 1} \quad (15)$$

### C. Predictive Distribution

After we have an inferred model, a common task is to estimate the probability of a new observation  $D^{(new)}$ . This can be implemented from likelihood as [4]

$$P(D^{(new)} | \theta, M) = \prod_{s_{t-1} \in A} \prod_{s_t \in A} p(s_t | s_{t-1})^{m(s_{t-1}s_t)} \quad (16)$$

where  $m(s_{t-1}s_t)$  is the number of times the word  $s_{t-1}s_t$  occurs in the sample data  $D^{(new)}$ . The respected posterior distribution as [4]

$$P(D^{(new)} | D, M) = \int d\theta P(D^{(new)} | \theta, M) P(\theta | D, M) \quad (17)$$

After following the same methodology to calculate evidence, posterior distribution can be obtained

$$P(D^{(new)} | D, M) = \prod_{s_{t-1} \in A} \left\{ \frac{\Gamma(n(s_{t-1}) + \alpha(s_{t-1}))}{\prod_{s_t \in A} \Gamma(n(s_{t-1}s_t) + \alpha(s_{t-1}s_t))} \times \frac{\prod_{s_t \in A} \Gamma(n(s_{t-1}s_t) + m(s_{t-1}s_t) + \alpha(s_{t-1}s_t))}{\Gamma(n(s_{t-1}) + m(s_{t-1}) + \alpha(s_{t-1}))} \right\} \quad (18)$$

## III. MODELING ORDER BOOK DYNAMICS

We consider a financial asset traded in an order-driven market. Accordingly, market participants can post two types of buy/sell orders: market order or limit order. When market order arrives to security exchange, it is matched with the best available price in the limit order book and a trade occurs. The quantities available in the limit order book will be updated thereafter. A limit order is an order intended to trade a certain amount of a security at a given price. Typically, limit orders are posted to an electronic trading system and orders' states are summarized at each price level: this is known as the limit order book. For the lowest price of an outstanding limit sell order, it is called ask price; bid price is referred to the highest buy price from the order book [5], [6].

A limit order from the order book can be executed or canceled. The execution of a limit order can be very quickly, just it may take a long time till participates are willing to trade based on the requested price [5], [6].

For traders, the change of ask & bid price represents the opportunity to gain profit. If trader's buy order was filled with bid price lower than current ask price, an executed sell order based on current market price means capital gain (from the price gap between sell & buy orders). Alternatively, if trader's security was borrowed, a current buy order filled with lower price than previously executed sell order is signifying another type of capital gain: After returning the security, the price gap between sell & buy orders is the margin. A high-precision model for order book dynamics can be utilized to predict the move for both ask & bid prices. A trading algorithm can be further developed.

#### A. Market Structure

We describe the order book and security trading with four  $Z^+$  discrete time Markov processes in the sequence of ask process, bid process, executed volume process and filled price process (i.e. executed order) as below

$$Ask(t) = \{Ask_1(t), Ask_2(t), \dots, Ask_{n(ask,t)}(t)\} \quad (19)$$

$$Bid(t) = \{Bid_1(t), Bid_2(t), \dots, Bid_{n(bid,t)}(t)\} \quad (20)$$

$$orderVol(t) = \{Vol_1(t), Vol_2(t), \dots, Vol_{n(order,t)}(t)\} \quad (21)$$

$$orderPrice(t) = \{Price_1(t), Price_2(t), \dots, Price_{n(order,t)}(t)\} \quad (22)$$

Accordingly,  $Ask_k(t)$  is the ask price from participant  $k$  at time  $t$ ;  $n(ask,t)$  stands for the total number of asks at time  $t$ . Similarly,  $Bid_l(t)$  is the bid price from participant  $l$  at time  $t$ ;  $n(bid,t)$  stands for the total number of bids at time  $t$ ;  $orderVol_m(t)$  and  $orderPrice_m(t)$  is the traded volume & price for participant  $m$  at time  $t$ ; we use  $n(order,t)$  to represent the total number of trades at time  $t$ .

Limit order will be valid & stay in the order book till execution or cancelation.  $Ask_{valid}(t)$  &  $Bid_{valid}(t)$  are used to indicate the still valid asks & bids at time  $t$  (they were created before time  $t$ ). Therefore, the best ask & bid prices at time  $t$  are then defined as

$$topAsk(t) = \min\{Ask_1(t), Ask_2(t), \dots, Ask_{n(ask,t)}(t), Ask_{valid}(t)\} \quad (23)$$

$$topBid(t) = \max\{Bid_1(t), Bid_2(t), \dots, Bid_{n(bid,t)}(t), Bid_{valid}(t)\} \quad (24)$$

We must have  $topAsk(t) > topBid(t)$ , otherwise, a free-lunch situation exists. The best ask & bid prices sometimes are referred to NBBO (National Best Bid & Offer). For our model, we are also interested to know spread & mid-price at time  $t$  as

$$bookMid(t) = (topAsk(t) + topBid(t)) / 2 \quad (25)$$

$$spread(t) = topAsk(t) - topBid(t) \quad (26)$$

We also want to know the highest & lowest traded prices

$$barHi(t) = \max\{Price_1(t), Price_2(t), \dots, Price_{n(order,t)}(t)\} \quad (27)$$

$$barLo(t) = \min\{Price_1(t), Price_2(t), \dots, Price_{n(order,t)}(t)\} \quad (28)$$

In addition to that, their mid-price and total volume are also considered

$$barMid(t) = (barHi(t) + barLo(t)) / 2 \quad (29)$$

$$vol(t) = \sum_{i=1..n(order,t)} Vol_i(t) \quad (30)$$

#### B. Inference of Markov Chain

We first define the normalization function  $f(x,t)$  to ensure finite memory & stationary as

$$f(x,t) = \begin{cases} 1 & \text{if } x(t) > x(t-1) \\ 0 & \text{if } x(t) = x(t-1) \\ -1 & \text{if } x(t) < x(t-1) \end{cases} \quad (31)$$

Accordingly, a join operation to represent the at time  $t$  is created as

$$state(t) = \begin{bmatrix} f(barHi,t) \\ f(barLo,t) \\ f(topAsk,t) \\ f(topBid,t) \\ f(barMid,t) \\ f(bookMid,t) \\ f(spread,t) \\ f(vol,t) \end{bmatrix}^T \quad (32)$$

where  $^T$  is the transpose operation.

For every element in state, there are 3 possible outcomes (i.e. 1, 0 & -1). Totally, we have 6561 possible states ( $3^8 = 6561$ ). Here, we define Markov chain transition matrix as

$$MC = \begin{bmatrix} P_{0,0} & P_{1,0} & P_{z,0} \\ P_{0,1} & P_{1,1} & \\ P_{0,z} & P_{1,z} & P_{z,z} \end{bmatrix} \quad (33)$$

where  $p_{k,j}$  represents the transition number from state  $k$  to state  $j$  and  $z$  indicates the state space.

We consider Eq. (33) Markov chain transition matrix  $MC$  as model  $M$  in inference framework. Without loss of

generality, we can assume a high-frequency data  $D$  being described as

$$D = \left\{ D(t) \quad \text{where } D(t) = \begin{bmatrix} \text{barHi}(t) \\ \text{barLo}(t) \\ \text{topAsk}(t) \\ \text{topBid}(t) \\ \text{barMid}(t) \\ \text{bookMid}(t) \\ \text{spread}(t) \\ \text{vol}(t) \end{bmatrix} \right\}^p \quad (34)$$

In order to map  $D(t)$  to a Markov state, we first apply normalization function  $f(x,t)$  (Eq. (31)). Thereafter, a hash mapping  $\text{hash}(t)$  can be considered as

$$\text{hash}(t) = (\text{state}(t) + \begin{bmatrix} \text{offset}_1 \\ \text{offset}_2 \\ \text{offset}_3 \\ \text{offset}_4 \\ \text{offset}_5 \\ \text{offset}_6 \\ \text{offset}_7 \\ \text{offset}_8 \end{bmatrix})^p \begin{bmatrix} Y^0 \\ Y^1 \\ Y^2 \\ Y^3 \\ Y^4 \\ Y^5 \\ Y^6 \\ Y^7 \end{bmatrix} \quad (35)$$

where  $\text{offset}_k$  and  $Y^j$  are element  $k$ 's offset and element  $j$ 's order. Here, we can set  $\text{offset}_k = 1$  for all  $k$  and  $Y^j = 3^j$  for all  $j$ . This hash will guarantee a unique positive integer for every state.

We only consider first order model,  $p_{k,j}$  can be realized as follow

$$p_{k,j} = \sum_i \text{total}(k, j) \quad (36)$$

where  $\text{hash}(t-1) = k$  and  $\text{hash}(t) = j$

Without loss of generality, Eq. (5) can be rewritten as

$$\theta = \{p(s_t | s_{t-1}) \quad s_t \in \mathbf{A}\} = \{\text{norm}(p_{k,j})\} \quad (37)$$

$$\text{norm}(p_{k,j}) = \frac{p_{k,j}}{\sum_i p_{k,i}} \quad (38)$$

Combining Eq. (33), Eq. (34) and Eq. (37), an inference Markov chain is obtained with

- 1) High-frequency & time-series data  $D = \{D(t)\}$
- 2) Markov chain transition matrix  $MC$  as chosen model
- 3) The vector of model parameters  $\theta$

Thereafter, online & real-time trading software can adopt predictive distribution to predict the change of bid & ask prices.

#### IV. IMPLEMENTATION

The proposed model is realized in two packages `hftCsvParser` and `hftMarkovChain`. We use Python language with scientific computing package `numpy`.

Package `hftCsvParser` was mainly developed for data processing. Similar to real-world applications, HFT data feed can be noisy. To provide accurate result, it's necessary to clean out noise before fitting. We implemented multi-stage processing: `csv raw input`, `pre data`, and `parsed data`.

The proposed inference Markov chain is developed by package `hftMarkovChain`. Other than model building, it also included serialization & deserialization based on sparse for storage & computing efficiency. The direct dump of transition matrix  $MC$  is over 1GB (Gigabyte); the sparse-based serialization is 40KB (kilobyte) in our experimentation.

#### V. EXPERIMENTAL RESULTS

Experimentation is conducted based on security AAPL with 43,710 time-series records. The period of time is ranged from 19:19:08.12 GMT to 19:26:25.07 GMT on May 18th, 2012 with inputs `topAsk`, `topBid`, `barHi`, `barLo`, and `vol`. The model building time will depend on machine. In our lab, Ubuntu 12.10 can finish fitting within 2 minutes.

Fig. 1 is the scatter plot of inference Markov chain transition matrix. Horizontal axis represents the start state; the end state is vertical axis. Each dot indicates actually happened transition from start state to end state. Theoretically, there are over 43 million possible transitions (6561 states to 6561 states, i.e. 43,046,721 possible transitions.) From Fig. 1, we see limited number of transitions and recognizable clusters. Specifically, with looking at the top 50 transitions (based on appeared numbers), its sparsity can be further acknowledged as illustrated in Fig. 2.

Fig. 3 is another chart for transition matrix. Its vertical axis stands for the happening times for transition  $i$ ; every possible transition has been labeled on the horizontal axis. 521 transitions have actually happened (less than 0.01% of all possible transitions.) By skipping the highest transition, the rest 49 transitions have similar range (between 10 and 380) as displayed in Fig. 4. If we look at the coverage percentage, top 10 transitions is over 95.5%; top 20 can cover 97%; over 98% has been included by top 50 transitions as in Table I.

As a result, over 98% coverage can be achieved by 50 transitions. Since the number of nodes can't be higher than number of transitions, 50 Markov nodes can model the dynamic of order book. It further indicates stock market behavior of short time-frame can be possibly clustered & labeled.

#### VI. CONCLUSION

In this paper, we propose a Bayesian inference of the Markov chain model class to model dynamics of order book for high frequency trading environment. Our real-world experimental result indicates the change behavior of short time-frame is in limited variety: 50 states can cover over 98% of transitions; their happening numbers are also mostly in the similar range (only one exception.) Consequently, this model

can be utilized as the foundation for algorithmic trading; the predictive distribution can be adopted to predict the move of order book. Profit strategy can be developed accordingly.

Possible future works of this research include further large-scale fitting of other securities to understand the difference of models; holdout methods for performance benchmarking; multi-order inference Markov chain.

TABLE I: THE PERCENTAGE OF TOP TRANSITION STATES

	Top 10	Top 20	Top 50	All (522)
Percentage	95.537%	97.066%	98.038%	100%

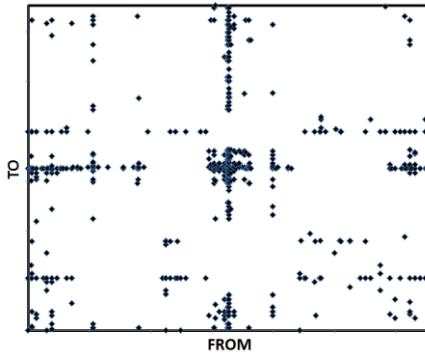


Fig. 1. Scatter plot of inference Markov chain transition matrix.

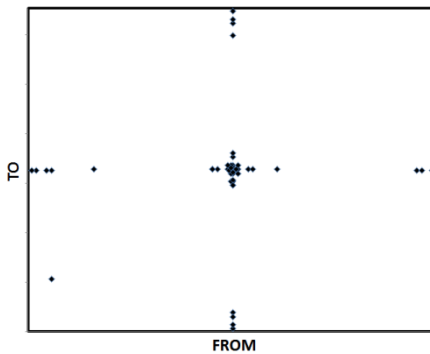


Fig. 2. Scatter plot of inference Markov chain transition matrix (50 transitions).

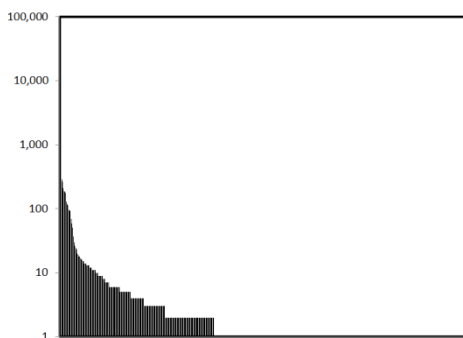


Fig. 3. Number of transitions in Markov chain transition matrix.

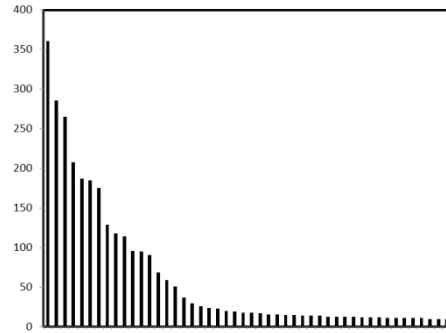


Fig. 4. Number of top 2 to 50 transitions in Markov chain transition matrix.

ACKNOWLEDGMENT

Y. L. Chang would like to thank his family A. Huang, L. Chan and J. Chang to support his research.

REFERENCES

- [1] Staff of the Division of Trading & Markets, U.S. Securities & Exchange Commission. (March 2014). Equity Market Structure Literature Review Part II: High Frequency Trading. [Online]. Available: [http://www.sec.gov/marketstructure/research/hft\\_lit\\_review\\_march\\_2014.pdf](http://www.sec.gov/marketstructure/research/hft_lit_review_march_2014.pdf)
- [2] C. M. Jones. (March 2013). What do we know about high-frequency trading? *Columbia Business School Research Paper No. 13-11*. [Online]. Available: <http://online.wsj.com/public/resources/documents/HFT0324.pdf>
- [3] J. Brogaard, T. Hendershoot, and R. Riordan. High frequency trading & price discovery. *European Central Bank, Working Paper Series. No. 1602/Nov 2013*. [Online]. Available: <http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1602.pdf>
- [4] C. C. Strelhoff, J. P. Crutchfield, and A. W. Hübler, "Inferring markov chains: Bayesian estimation, model comparison, entropy rate, and out-of-class modeling," *Physical Review*, E76 011106, 12 July 2007
- [5] H. Huang and A. N. Kercheval, "A generalized birth-death stochastic model for high-frequency order book dynamics," *Quantitative Finance*, vol. 12, issue 4, pp. 547-557, Feb. 2011.
- [6] R. Cont, S. Stoikov, and R. Talreja, "A stochastic model for order book dynamics," *Operations Research*, vol. 58, issue 3, pp. 549-563, Feb 2010.



**Yuan Lung Chang** received his Ph.D. degree in computer engineering and science from Yuan Ze University, Taiwan, in 2007. He was a paper reviewer for IEEE Transactions on Vehicle Technologies and IEEE Wireless Conference with 3 published papers. His current research interests are in machine learning, adaptive filter, finance engineering, and parallel computing. He has been working near 3 years for computer-assisted trading system.