

AN INVARIANCE FOR THE LARGE-SAMPLE EMPIRICAL DISTRIBUTION OF WAITING TIME BETWEEN SUCCESSIVE EXTREMES

LO-BIN CHANG^{1,a}, ALOK GOSWAMI^{2,b}, CHII-RUEY HWANG^{3,c}
AND FUSHING HSIEH^{4,d}

¹Department of Applied Mathematics, National Chiao Tung University, Hsinchu, Taiwan 30010.

^aE-mail: lobin_chang@math.nctu.edu.tw

²Theoretical Statistics and Mathematics Unit, I.S.I., 203 B.T. Road, Kolkata 700108, India.

^bE-mail: alok@isical.ac.in

³Institute of Mathematics, Academia Sinica, Taipei, Taiwan 11529.

^cE-mail: crhwang@sinica.edu.tw

⁴Department of Statistics, University of California, Davies, CA, U.S.A.

^dE-mail: fushing@wald.ucdavis.edu

Abstract

For a given set of observations, we consider the waiting times between successive returns to extreme values. Our main result is an invariance theorem that says that, as the size of the data set gets large, the empirical distribution of the waiting time converges with probability one to a geometric distribution, whenever the observations are i.i.d. or, more generally, a realization of any exchangeable joint distribution. We also prove a central limit theorem for these empirical distributions (properly normalized), which leads in natural way to a Donsker-type theorem on $D[0, \infty)$. Our study was motivated by empirical and simulation results, which indicate the existence of an empirical invariance, obtained in connection with our investigation on stock price data. In many of the accepted models for stock prices, our results would apply and suggest such invariance to hold. It may be pointed out that the empirical invariance from the real stock price data was quite far away from the theoretical invariance.

Received March 30, 2012.

AMS Subject Classification: 60F05, 60F15, 60F17, 62E20, 91B24, 91B28, 91G99.

Key words and phrases: Asymptotically i.i.d., Black-Scholes' model, central limit theorem, convergence of empirical distribution, Donsker's theorem, exchangeable random variables, fractional Brownian motion, geometric distribution, hierarchical segmentation, Kolmogorov distribution, Kolmogorov-Smirnov statistic, law of large numbers, Lévy process, sufficient statistic, volatile period.

1. Introduction

The investigation here was motivated by an empirical invariance found in our study of the raw data consisting of the actual trade price of the intraday transactions data (trades and quotes) of companies in S&P500 list from 1998 to 2007 and part of 2008 (see [3]).

The temporal evolution of the price of a stock is given by a stochastic process $\{S_t\}$. For example the Black-Scholes model assumes that $\{S_t\}$ follows a geometric Brownian motion with constant volatility and satisfies the stochastic differential equation

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

where r and σ are positive constants and $\{W_t\}$ is standard Brownian motion. An immediate consequence of this model is that the 'returns' over disjoint time intervals of equal duration are independent and identically distributed normal random variables. More specifically, if S_1, S_2, \dots, S_{n+1} are the prices of a stock at a set of $n + 1$ equally spaced time points, then the random variables $(S_2 - S_1)/S_1, (S_3 - S_2)/S_2, \dots, (S_{n+1} - S_n)/S_n$ are i.i.d. normal random variables.

To allow for possible jumps for stock prices, a more general model is considered where the rate of return is assumed to follow a general Lévy process, resulting in the 'geometric Lévy process' model for evolution of stock prices (see [8] for example). However, the returns over disjoint time intervals of equal duration still remain i.i.d. random variables (though not necessarily normal).

Suppose now that we have available a large set of data on prices of a stock at equal intervals of time and calculate the returns. From this set, say of size n , of return values, suppose we generate a sequence of 0s and 1s that identifies those among the set of return values which constitute the upper (or lower) ten percent among this set. More specifically, a return value is dubbed as 0 if it falls in the upper (or lower) ten percentile of the set of return values and is dubbed as 1 otherwise. This will lead to a n -long sequence of 0s and 1s with (roughly) $0.1n$ many 0s and remaining (roughly) $0.9n$ many 1s. If we now look at the lengths of the successive runs of 1s in this sequence, including the (possibly empty) runs before the first and after the last 0s, then how are they expected to behave? We prove that if

for each n , the successive returns have an exchangeable joint distribution, as under the Black-Scholes model or geometric Lévy process model, then, in the limit, these run lengths follow a common geometric distribution. Moreover, the run-lengths are asymptotically independent. More specifically for any fixed k , the lengths of first k runs of 1s are asymptotically distributed as k i.i.d. geometric random variables (Theorem 2.1).

Next, we consider the probability histogram generated by the lengths of all the runs of 1s. This will be a histogram of a probability on non-negative integers. We prove a strong law of large numbers (Theorem 2.2) which implies that, under the same hypothesis as in the earlier theorem, this (sample) histogram will converge, with probability one, to that of a geometric distribution. The convergence holds, with probability one, also in total variation norm, in Kolmogorov norm and uniformly (Corollary 2.1). Using results from Chen ([4]), we also obtain an associated central limit theorem (Theorem 2.3) for the (sample) probability histogram. It may be worthwhile to point out in this connection, that by applying a result of Chen ([4], Theorem 2), one may obtain a limit theorem for our probability histogram. However, such a limit theorem is substantially weaker than our Theorem 2.2, in the sense that Chen's result gives convergence only in probability (which is easy in our set-up) and assumes additional conditions.

Another point needs mention here. Under our hypothesis that the returns (over successive equal lengths of times) are exchangeable, it is obvious that the lengths successive runs of 1s, as described above, will also have an exchangeable joint distribution. However, they are not independent, even if the successive returns are assumed to be so. Our Theorem 2.2 shows that the finite sample empirical distribution converges almost surely to an appropriate geometric distribution, as if it is the empirical distribution based on a finite i.i.d. sample from the geometric distribution. While our Theorem 2.1 shows that the successive run lengths are indeed asymptotically i.i.d. geometric, the convergence of the empirical distribution to geometric distribution is not automatic by any means. Further, the de-Finetti decomposition of exchangeable probabilities on infinite product spaces as mixtures of i.i.d. probabilities does not apply here, because, for each n , we only have a finite-dimensional exchangeable distribution. A natural option would be

to try and apply ideas of Diaconis-Freedman for finite exchangeable probabilities (see [7]) where the weak law follows, but our attempts along this line did not prove fruitful for the strong law.

Although we are not going to present simulation results in this paper, remarks might be made according to simulations.

We prove a strong law of large numbers with the limit being an invariance, a geometric distribution. The associated central limit theorem (Theorem 2.3) easily leads to a Donsker-type theorem (Theorem 2.4) for the empirical distribution function on $D[0, \infty)$. This, of course, will characterize the limiting distribution of the corresponding Kolmogorov-Smirnov statistic. As expected, this limiting distribution will not be the usual Kolmogorov distribution, as was already evidenced by our simulations.

As mentioned in the first paragraph, our raw data consisted of the actual trade price of the intraday transactions data (trades and quotes) of companies in S&P500 list from 1998 to 2007 and part of 2008. The return process was analyzed at five-minute, one-minute, and 30-second intervals for a whole year (see [3]). We repeated the procedure mentioned in previous paragraphs and considered the empirical waiting time distributions of hitting a certain percentile. Closeness of distributions was measured by ROC area and Kolmogorov distance. The invariance of these empirical distributions were found across stocks, time units, and years. This empirical invariance turned out to be different from the geometric distribution (see [3], [10]).

On the other hand, simulated data points generated from i.i.d. normal and i.i.d. uniform, regarded as returns, were converted into a sequence of 0s and 1s as above. The histograms generated from the lengths of 1-runs were matched against the histogram of a geometric distribution with the parameter corresponding to the width of the percentile. In all cases, there seemed to be perfect match. Of course, this is only to be expected now, in view of our Theorem 2.2. We would like to point out, however, that originally it was this observed invariance in our simulation results that inspired us to formulate and prove the main theoretical results described in this work.

Our theoretical results when contrasted with our above-mentioned analysis based on real data on prices of various stocks would raise questions on the validity of a large class of widely accepted models on stock-prices. In fact, any model that merely implies exchangeability of successive returns

would be put to question. In recent times, models that incorporate long range dependence, have been proposed; these models use fractional Brownian motion instead of Brownian motion (see [11]). Since exchangeability of successive returns no longer hold for these models (except in two special cases), our invariance results do not apply. However, simulations indicate that a law of large numbers holds but with a different limit.

One other point that we like to note here is that, in the above, we have only described the first layer of coding in what is known as hierarchical segmentation of time-series data (see [10]). The data were coded into a binary 0 – 1 sequence identifying times of occurrences of “rare events” (namely, occurrence of extreme values). One can now think of the lengths of successive 1-runs (waiting times between occurrence of rare events) as the data, which is only exchangeable but not i.i.d., and code these further into a “second layer” of 0 – 1 sequence in exactly the same way as before. An occurrence of a 0 at this level corresponds to times of very long waiting between rare events at the earlier level. A long 1-run at this level corresponds to a long wait between two long waits for extreme events at the earlier level. By assuming only exchangeability (instead of i.i.d.) of the underlying random variables, our results would imply that at every level of coding, the same limiting result for the histograms generated by the lengths of 1-runs is valid.

For possible applications to finance, one may refer to ([3], [9], [10]). For example, volatile period may be defined hierarchically up to the third level using the real data: if the length of runs of 1s falls in say upper ten percentile, denote that period 0*, otherwise 1*; repeat the same procedure for the length of runs of 1*s and denote the period in the upper ten percentile 0[@]. One may regard 0[@]s as the volatile periods. Using this segmentation, the dynamics and trading strategy are studied in ([10]).

2. Statements of Main Results

Consider n objects arranged in a row. Suppose m of the objects are selected at random, with each of the $\binom{n}{m}$ possible selections having the same probability. If we describe a particular selection by dubbing each selected object as a 0 and each unselected object as a 1, each selection gives an n -long binary sequence with m many 0s and $n - m$ many 1s. Let now Y_1^n, \dots, Y_{m+1}^n denote the lengths of the $m + 1$ runs of 1s thus obtained. We

include the 1-runs, possibly of zero length, before the first 0 and after the last 0. This gives a sequence of $m + 1$ non-negative integer valued random variables, which are clearly not independent, because $Y_1^n + \cdots + Y_{m+1}^n = n - m$. Thus, the possible values of the random vector $(Y_1^n, \dots, Y_{m+1}^n)$ are vectors (l_1, \dots, l_{m+1}) of nonnegative integers with $l_1 + \cdots + l_{m+1} = n - m$ and, for each such vector,

$$P(Y_1^n = l_1, \dots, Y_{m+1}^n = l_{m+1}) = \frac{1}{\binom{n}{m}},$$

since the event on the right hand side corresponds precisely to selecting the $(l_1 + 1)$ th, $(l_1 + l_2 + 2)$ th, \dots , and the $(l_1 + \cdots + l_m + m)$ th objects among the n objects. It is an easy consequence of this that the random variables Y_1^n, \dots, Y_{m+1}^n are exchangeable. This is because, for any permutation π on $\{1, \dots, m + 1\}$, the event $\{Y_{\pi(1)}^n = l_1, \dots, Y_{\pi(m+1)}^n = l_{m+1}\}$ is the same as $\{Y_1^n = l_{\pi^{-1}(1)}, \dots, Y_{m+1}^n = l_{\pi^{-1}(m+1)}\}$, and this last event has the same probability as the event $\{Y_1^n = l_1, \dots, Y_{m+1}^n = l_{m+1}\}$, both equal to $1/\binom{n}{m}$.

If now $n \rightarrow \infty$, we get a triangular array where each row consists of a finite sequence of random variables that are exchangeable but not independent. We show that if n and m both go to infinity in such a way that $m/n \rightarrow p$ for some $p \in (0, 1)$, then the random variables become asymptotically independent. Moreover, the limiting common distribution is geometric with parameter p .

Theorem 2.1. *If $n \rightarrow \infty$ and $m \rightarrow \infty$ in such a way that $m/n \rightarrow p \in (0, 1)$, then, for any $k \geq 1$,*

$$(Y_1^n, \dots, Y_k^n) \xrightarrow{d} (Y_1, \dots, Y_k),$$

where Y_1, \dots, Y_k are independent and identically distributed random variables having the geometric distribution with parameter p .

\xrightarrow{d} denotes convergence in distribution. A particular consequence of the above theorem will be that the sequence Y_k^n , for each fixed k , converges in distribution to a geometric random variable with parameter p .

We now consider a slightly different problem. For each n , we consider the probability histogram generated by the random variables Y_1^n, \dots, Y_{m+1}^n . We

get a (random) probability distribution on non-negative integers, given by the probability mass functions

$$\theta_n(l)(\omega) = \frac{1}{m+1} \sum_{i=1}^{m+1} \mathbf{1}_{\{Y_i^n(\omega)=l\}}, \quad l = 0, 1, \dots .$$

The next theorem says that, under the same hypothesis as in Theorem 2.1, these probability distributions converge, with probability 1, to the geometric distribution with parameter p . In other words, the empirical distribution from each row of the triangular array converges almost surely to the geometric distribution.

Theorem 2.2. *If $n \rightarrow \infty$ and $m \rightarrow \infty$ in such a way that $m/n \rightarrow p \in (0, 1)$, then,*

$$P\left(\lim_{n \rightarrow \infty} \theta_n(l) = p(1-p)^l, \quad l = 0, 1, \dots\right) = 1.$$

Using Scheffé's Theorem and the fact that all the distributions involved are discrete probability distributions concentrated on non-negative integers, it is easy to deduce that if we denoting the empirical distribution generated by the random variables in the n th row by P_n , then one has

Corollary 2.1. *Under the same hypothesis as in Theorem 2.1, the distributions P_n converge, with probability 1, to the geometric distribution with parameter p , in total variation as well as in Kolmogorov distance. Moreover, the convergence $\theta_n(l) \rightarrow p(1-p)^l$ holds uniformly in l with probability 1.*

The next result is a central limit theorem for the empirical probability mass functions $\theta_n(l)$, $l = 0, 1, \dots$. It says that for distinct non-negative integers l_1, \dots, l_k , the vector $(\theta_n(l_1), \dots, \theta_n(l_k))$ is asymptotically normal. This will easily imply asymptotic normality of all the finite-dimensional distributions of the empirical distribution function. It would be quite natural to push this a little more to get weak convergence of the (properly normalized) empirical distribution in $D([0, \infty))$ to a Gaussian process, which is described later. To state the central limit theorem, it will be convenient to introduce some notations. Let $\theta(\cdot)$, μ and σ^2 denote respectively the probability mass function, the mean and the variance of the geometric distribution with parameter p ; thus, $\theta(l) = p(1-p)^l$, $l = 0, 1, \dots$, $\mu = (1-p)/p$ and $\sigma^2 = (1-p)/p^2$. Since $p = \mu/(1+\mu)$, one can also

think of μ as parameterizing the geometric distribution. For a given set of distinct non-negative integers l_1, \dots, l_k , let $\alpha_j = pl_j - (1 - p)$, $j = 1, \dots, k$ and $\beta_{ij} = \delta_{ij}\theta(l_i) - \theta(l_i)\theta(l_j)[1 + \alpha_i\alpha_j/(1 - p)]$, $i, j = 1, \dots, k$, where δ_{ij} denotes the Kronecker delta. The following result can then be derived from a result due to Chen ([4, Thm. 6]).

Theorem 2.3. *If $n \rightarrow \infty$ and $m \rightarrow \infty$ in such a way that $m/n \rightarrow p \in (0, 1)$, then the vector $\sqrt{m+1}(\theta_n(l_1) - \theta(l_1), \dots, \theta_n(l_k) - \theta(l_k))$ converges in distribution to a zero-mean k -dimensional Gaussian distribution with dispersion matrix $\mathbf{D} = ((\beta_{ij}))$.*

As mentioned, the above central limit theorem easily leads to a Donsker-type theorem for the empirical distribution function on $[0, \infty)$. To be precise, let $F_n(t)$, $t \in [0, \infty)$ denote the empirical (cumulative) distribution function based on the random variables Y_1^n, \dots, Y_{m+1}^n , that is

$$F_n(t)(\omega) = \frac{1}{m+1} \sum_{i=1}^{m+1} \mathbf{1}_{\{Y_i^n(\omega) \leq t\}}, \quad t \in [0, \infty).$$

Denote the (cumulative) distribution function of the geometric distribution (with parameter p) by F , that is, $F(t) = \sum_{l=0}^{[t]} \theta(l) = 1 - (1 - p)^{[t]+1}$, $t \in [0, \infty)$. Also, let G denote the function defined by $G(t) = \sum_{l=0}^{[t]} l\theta(l)$, $t \in [0, \infty)$. We then have the following result, as an easy consequence of Theorem 2.3 above.

Theorem 2.4. *If $n \rightarrow \infty$ and $m \rightarrow \infty$ in such a way that $m/n \rightarrow p \in (0, 1)$, then the sequence $\{\sqrt{m+1}(F_n(t) - F(t)), t \in [0, \infty)\}$ of $D[0, \infty)$ -valued processes converges weakly in the Skorohod topology, to a zero-mean Gaussian process with covariance kernel given by*

$$\begin{aligned} \Sigma(s, t) = & F([s \wedge t]) - F([s])F([t]) - (1 - p)^{-1}\{(1 - p)^2 F([s])F([t]) \\ & + p^2 G([s])G([t]) - p(1 - p)(F([s])G([t]) + G([s])F([t]))\}, \end{aligned}$$

where, for a real number t , $[t]$ denotes the greatest integer less than or equal to t .

In the results stated above the basic set up has been that we have a set of n objects from which m are selected at random (with equal probability for each selection). We then have some limiting results as n and m become

large with $m \sim np$. We now want to go on to describe how this is connected with the data on prices of a stock. This is precisely where the hierarchical segmentation of the stock price time series comes into play. Given prices of a stock at equal intervals of time, if we consider the times of occurrences of extreme values for the returns over successive time intervals, we end up selecting a certain subset of a fixed proportion from the set of all time points. Our result simply says that if under the assumed model for a stock price, the returns over successive time intervals have an exchangeable joint distribution, then all selections are equally likely. This should be obvious. We elaborate it only for the sake of completeness.

Let $\alpha, \beta \geq 0$ with $0 < \alpha + \beta < 1$. From a data set consisting n points (x_1, x_2, \dots, x_n) , we want to choose those that form the lower 100α -percentile and those that form the upper 100β -percentile. To avoid trivialities, let us assume the size n of the data set is strictly larger than $n > (1 - \alpha - \beta)^{-1}$. In case the data points are all distinct, we have an unambiguous choice. Indeed, we may arrange the data points in the (strictly) decreasing order as $x_{(1)} < x_{(2)} < \dots < x_{(n)}$. If now k and l are integers satisfying

$$\frac{k}{n} \leq \alpha < \frac{k+1}{n} \leq \frac{l-1}{n} < 1 - \beta \leq \frac{l}{n},$$

then $(x_{(1)}, \dots, x_{(k)})$ will form the lower 100α -percentile and $(x_{(l)}, \dots, x_{(n)})$ will form the upper 100β -percentile. In case the data points are not all distinct, we may have more than one possible choices for the k among the n data points that form the lower 100α -percentile or for the $n - l + 1$ that form the upper 100β -percentile. In such cases, our prescription is to pick one among the possible choices with equal probability for each.

Thus, we will always end up selecting exactly $k + n - l + 1$ from the n data points with k of them forming the lower 100α -percentile and remaining $n - l + 1$ forming the upper 100β -percentile. The next theorem considers the case when the data points consist of n random variables with an exchangeable joint distribution. This result, which is too obvious to require a formal proof, provides the required connecting link between stock price data and the limiting results in the earlier theorems.

Theorem 2.5. *If X_1, \dots, X_n are random variables with an exchangeable joint distribution, then any one of the $\binom{n}{k+n-l+1}$ possible choices can occur*

with equal probability as the set of points constituting the lower 100α - and upper 100β -percentiles.

3. Proofs

Proof of Theorem 2.1. The proof is straightforward. We have to prove that, for every choice of k non-negative integers l_1, \dots, l_k ,

$$P(Y_1^n = l_1, \dots, Y_k^n = l_k) \longrightarrow \prod_{i=1}^k [p(1-p)^{l_i}].$$

The left hand side clearly equals $\binom{n-l_1-\dots-l_k-k}{m-k} / \binom{n}{m}$. Denoting $s_1 = l_1 + 1, s_2 = l_1 + l_2 + 2, \dots, s_k = l_1 + \dots + l_k + k$, this last expression can be written as $\prod_{i=1}^k \binom{n-s_i}{m-i} / \binom{n-s_i+l_i+1}{m-i+1}$. It, therefore, suffices for us to prove that, for each $i = 1, 2, \dots, k$,

$$\binom{n-s_i}{m-i} / \binom{n-s_i+l_i+1}{m-i+1} \longrightarrow p(1-p)^{l_i}.$$

But this follows by using Lemma 3.1 below with $\tilde{n} = n - s_i, \tilde{m} = m - i, \tilde{l} = l_i$ and by noting that $m/n \rightarrow p$ implies $(m-i)/(n-s_i) \rightarrow p$. \square

Lemma 3.1. *If $\tilde{m} \rightarrow \infty, \tilde{n} \rightarrow \infty$ in such a way that $\tilde{m}/\tilde{n} \rightarrow p \in (0, 1)$, then for any $\tilde{l} = 0, 1, 2, \dots$, one has $\binom{\tilde{n}}{\tilde{m}} / \binom{\tilde{n}+\tilde{l}+1}{\tilde{m}+1} \longrightarrow p(1-p)^{\tilde{l}}$.*

Proof of Theorem 3.1. Easy simplification gives

$$\binom{\tilde{n}}{\tilde{m}} / \binom{\tilde{n}+\tilde{l}+1}{\tilde{m}+1} = \frac{\tilde{m}+1}{\tilde{n}+\tilde{l}+1} \cdot \frac{\tilde{n}!(\tilde{n}-\tilde{m}+\tilde{l})!}{(\tilde{n}-\tilde{m})!(\tilde{n}+\tilde{l})!}.$$

As $n \rightarrow \infty$ (and hence $m \rightarrow \infty$), an application of Stirling's approximations for all the factorials and simplification gives

$$\frac{\tilde{n}!(\tilde{n}-\tilde{m}+\tilde{l})!}{(\tilde{n}-\tilde{m})!(\tilde{n}+\tilde{l})!} \sim \frac{\tilde{n}^{\tilde{n}+\frac{1}{2}}}{(\tilde{n}+\tilde{l})^{\tilde{n}+\frac{1}{2}}} \cdot \frac{(\tilde{n}-\tilde{m}+\tilde{l})^{\tilde{n}-\tilde{m}+\frac{1}{2}}}{(\tilde{n}-\tilde{m})^{\tilde{n}-\tilde{m}+\frac{1}{2}}} \cdot \frac{(\tilde{n}-\tilde{m}+\tilde{l})^{\tilde{l}}}{(\tilde{n}+\tilde{l})^{\tilde{l}}}.$$

The three factors above converge to $e^{-\tilde{l}}, e^{\tilde{l}}$ and $(1-p)^{\tilde{l}}$ respectively. These, together with the fact that $\frac{\tilde{m}+1}{\tilde{n}+\tilde{l}+1} \longrightarrow p$ proves the result. \square

Before proceeding with the proof of Theorem 2.2, we make a couple of observations.

First of all, it is enough to prove that for each $l = 0, 1, 2, \dots$,

$$P\left(\lim_{n \rightarrow \infty} \theta_n(l) = p(1-p)^l\right) = 1.$$

In fact, it is enough to do this only for $l = 1, 2, \dots$, since the case $l = 0$ will then automatically follow.

Secondly, since $E(\mathbf{1}_{\{Y_i^n=l\}}) = P(Y_i^n = l) \rightarrow p(1-p)^l$ by Theorem 2.1, it is enough to prove that

$$\frac{1}{m+1} \sum_{i=1}^{m+1} \left[\mathbf{1}_{\{Y_i^n=l\}} - E(\mathbf{1}_{\{Y_i^n=l\}}) \right] \rightarrow 0, \text{ with probability 1.}$$

This, of course, is just a strong law of large numbers for the triangular array of random variables $\{\mathbf{1}_{\{Y_i^n=l\}}, 1 \leq i \leq m+1, n \geq 1\}$. Unfortunately, there does not seem to be a whole lot of literature on strong law of large numbers for triangular arrays. Teicher ([13]) had the first important paper on almost sure convergence for double arrays. Andrews ([1]), Davidson ([5]) and De Jong ([6]) had results on weak law for mixingale sequences, which can be easily be generalised for triangular arrays. De Jong ([6]) proved a strong law for mixingale triangular arrays. However, none of these results apply in our situation. We prove the result directly by showing that, for any $\epsilon > 0$,

$$(1) \quad \sum_{n=1}^{\infty} P\left(\left|\frac{1}{m+1} \sum_{i=1}^{m+1} \left[\mathbf{1}_{\{Y_i^n=l\}} - E(\mathbf{1}_{\{Y_i^n=l\}}) \right]\right| > \epsilon\right) < \infty,$$

from which the desired almost sure convergence will follow by using Borel-Cantelli Lemma.

Proof of Theorem 2.2. To prove the convergence of the series in (1), we simply prove that

$$(2) \quad \sum_{n=1}^{\infty} \frac{1}{(m+1)^4} E\left|\sum_{i=1}^{m+1} \left[\mathbf{1}_{\{Y_i^n=l\}} - E(\mathbf{1}_{\{Y_i^n=l\}}) \right]\right|^4 < \infty,$$

and then Chebyshev inequality will take care.

To compute the fourth moments in the summands in (2), we introduce some notations. Let $Z_i^n = \mathbf{1}_{\{Y_i^n=l\}}$. Then $E(Z_i^n) = P(Y_i^n = l)$, which by the exchangeability of the random variables $\{Y_i^n, 1 \leq i \leq m+1\}$ is also equal to $P(Y_1^n = l) = p_1$, say.

Further, exchangeability of $\{Y_i^n, 1 \leq i \leq m+1\}$ implies exchangeability of $\{Z_i^n, 1 \leq i \leq m+1\}$ as well, from which it is easy to see that

$$E \left| \sum_{i=1}^{m+1} \left[\mathbf{1}_{\{Y_i^n=l\}} - E(\mathbf{1}_{\{Y_i^n=l\}}) \right] \right|^4 = E \left| \sum_{i=1}^{m+1} (Z_i^n - p_1) \right|^4 = S_1 + S_2 + S_3 + S_4,$$

where

$$S_1 = (m+1)E(Z_1^n - p_1)^4,$$

$$S_2 = m(m+1)E \left[(Z_1^n - p_1)^3 (Z_2^n - p_1) \right] \\ + \frac{m(m+1)}{2} E \left[(Z_1^n - p_1)^2 (Z_2^n - p_1)^2 \right],$$

$$S_3 = \frac{m(m-1)(m+1)}{2} E \left[(Z_1^n - p_1)^2 (Z_2^n - p_1) (Z_3^n - p_1) \right],$$

and

$$S_4 = \binom{m+1}{4} E \left[(Z_1^n - p_1) (Z_2^n - p_1) (Z_3^n - p_1) (Z_4^n - p_1) \right].$$

Thus the series in (2) becomes $\sum_{n=1}^{\infty} \frac{1}{(m+1)^4} [S_1 + S_2 + S_3 + S_4]$ and we will have to prove convergence of this.

Noting that $|Z_i^n - p_1| \leq 1$ and $m \sim np$ with $p \in (0, 1)$, it is clear that both the series

$$\sum_{n=1}^{\infty} \frac{S_1}{(m+1)^4} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{S_2}{(m+1)^4}$$

are convergent.

To show convergence of the two series

$$\sum_{n=1}^{\infty} \frac{S_3}{(m+1)^4} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{S_4}{(m+1)^4},$$

we introduce some notations. In analogy with the notation $p_1 = P(Z_1^n =$

1) = p_1 , let us denote $p_0 = P(Z_i^n = 0) = 1 - p_1$. Similarly, let

$$\begin{aligned} p_{11} &= P(Z_1^n = 1, Z_2^n = 1), \quad p_{10} = P(Z_1^n = 1, Z_2^n = 0), \\ p_{01} &= P(Z_1^n = 0, Z_2^n = 1), \quad p_{00} = P(Z_1^n = 0, Z_2^n = 0). \end{aligned}$$

Using exchangeability and elementary properties of probability, one can express the last three in terms of p_1 and p_{11} . Indeed,

$$p_{10} = p_{01} = p_1 - p_{11},$$

and

$$p_{00} = 1 - 2p_1 + p_{11}.$$

We can likewise define p_{ijk} and p_{ijkh} for all $i, j, k, h \in \{0, 1\}$ and express all the p_{ijk} in terms of p_1, p_{11}, p_{111} and all the p_{ijkh} in terms of $p_1, p_{11}, p_{111}, p_{1111}$. Indeed,

$$\begin{aligned} p_{110} &= p_{101} = p_{011} = p_{11} - p_{111}, \\ p_{100} &= p_{010} = p_{001} = p_1 - 2p_{11} + p_{111}, \\ p_{000} &= 1 - 3p_1 + 3p_{11} - p_{111}; \end{aligned}$$

and

$$\begin{aligned} p_{1110} &= p_{1101} = p_{1011} = p_{0111} = p_{111} - p_{1111}, \\ p_{1100} &= p_{1010} = p_{1001} = p_{0011} = p_{0101} = p_{0110} = p_{11} - 2p_{111} + p_{1111}, \\ p_{1000} &= p_{0100} = p_{0010} = p_{0001} = p_1 - 3p_{11} + 3p_{111} - p_{1111}, \\ p_{0000} &= 1 - 4p_1 + 6p_{11} - 4p_{111} + p_{1111}. \end{aligned}$$

In the expression

$$E[(Z_1^n - p_1)^2(Z_2^n - p_1)(Z_3^n - p_1)] = \sum_{i,j,k \in \{0,1\}} (i - p_1)^2(j - p_1)(k - p_1)p_{ijk},$$

if one uses the above formulas for the p_{ijk} and simplifies, one gets

$$E[(Z_1^n - p_1)^2(Z_2^n - p_1)(Z_3^n - p_1)] = p_1^3 - 3p_1^4 - 2p_1p_{11} + 5p_1^2p_{11} + p_{111} - 2p_1p_{111}.$$

We are now going to use the explicit formula for p_1, p_{11} and p_{111} , namely,

$$p_1 = \binom{n-l-1}{m-1} / \binom{n}{m},$$

$$p_{11} = \binom{n-2l-2}{m-2} / \binom{n}{m}$$

and

$$p_{111} = \binom{n-3l-3}{m-3} / \binom{n}{m}.$$

Writing θ_n for m/n , we get, after an easy simplification, that

$$p_1 = \theta_n(1-\theta_n)^l(1-a_n),$$

$$p_{11} = \theta_n^2(1-\theta_n)^{2l}(1-b_n)$$

and

$$p_{111} = \theta_n^3(1-\theta_n)^{3l}(1-c_n),$$

where

$$a_n = \left(\frac{l(l-1)}{2(1-\theta_n)} - \frac{l(l+1)}{2} \right) \frac{1}{n} + \mathbf{O}(n^{-2}),$$

$$b_n = \left(\frac{2l(2l-1)}{2(1-\theta_n)} - \frac{(2l+1)(2l+2)}{2} + \frac{1}{\theta_n} \right) \frac{1}{n} + \mathbf{O}(n^{-2}),$$

and

$$c_n = \left(\frac{3l(3l-1)}{2(1-\theta_n)} - \frac{(3l+2)(3l+3)}{2} + \frac{3}{\theta_n} \right) \frac{1}{n} + \mathbf{O}(n^{-2}).$$

Using these in the simplified formula for $E[(Z_1^n - p_1)^2(Z_2^n - p_1)(Z_3^n - p_1)]$, one can easily find that the “leading terms” nicely cancel out and one is left only with terms that are no bigger than $\mathbf{O}(n^{-1})$. It should be noted in this connection that, since $\theta_n \rightarrow p \in (0, 1)$, both θ_n and $1 - \theta_n$ may be assumed to be bounded away from 0 and are, of course, bounded above by 1. The upshot of all these is that

$$E[(Z_1^n - p_1)^2(Z_2^n - p_1)(Z_3^n - p_1)] = \mathbf{O}(n^{-1}).$$

Now

$$\sum_{n=1}^{\infty} \frac{S_3}{(m-1)^4} = \sum_{n=1}^{\infty} \frac{m(m-1)}{2(m+1)^3} E[(Z_1^n - p_1)^2(Z_2^n - p_1)(Z_3^n - p_1)],$$

so that the convergence of this series follows from the above and the fact that

$$\frac{m(m-1)}{2(m+1)^3} = \mathbf{O}(n^{-1}). \quad (\text{Recall that } m \sim np.)$$

The argument for convergence of $\sum_{n=1}^{\infty} \frac{S_4}{(m+1)^4}$ is similar but a little more delicate. Since $S_4 = \binom{m+1}{4} E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)]$, showing merely that $E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)] = \mathbf{O}(n^{-1})$ will no longer do.

We will actually show that

$$E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)] = \mathbf{O}(n^{-2}),$$

which will, of course, yield the required convergence.

As before, in the expression

$$\begin{aligned} & E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)] \\ &= \sum_{i,j,k,h \in \{0,1\}} (i - p_1)(j - p_1)(k - p_1)(h - p_1)p_{ijkh}, \end{aligned}$$

we use the formula for the p_{ijkh} and then simplify to get

$$E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)] = -3p_1^4 + 6p_1^2p_{11} - 4p_1p_{111} + p_{1111}.$$

We already know

$$p_1 = \theta_n(1 - \theta_n)^l(1 - a_n), \quad p_{11} = \theta_n^2(1 - \theta_n)^{2l}(1 - b_n)$$

and

$$p_{111} = \theta_n^3(1 - \theta_n)^{3l}(1 - c_n),$$

where a_n, b_n and c_n are as obtained before. Similar simplification for p_{1111} gives

$$p_{1111} = \theta_n^4(1 - \theta_n)^{4l}(1 - d_n),$$

where

$$d_n = \left(\frac{4l(4l-1)}{2(1-\theta_n)} - \frac{(4l+3)(4l+4)}{2} + \frac{6}{\theta_n} \right) \frac{1}{n} + \mathbf{O}(n^{-2}).$$

Using these in the formula for $E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)]$, one can easily see that

$$\begin{aligned} & E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)] \\ &= \theta_n^4(1 - \theta_n)^{4l} [-3(-4a_n) + 6(-2a_n - b_n) - 4(-a_n - c_n) + (-d_n) + \mathbf{O}(n^{-2})]. \end{aligned}$$

The interesting thing that happens now is that if one collects all the coefficients of n^{-1} from the quantity within the square brackets above, they all cancel out, thereby leaving only the $\mathbf{O}(n^{-2})$ terms. This shows that

$$E[(Z_1^n - p_1)(Z_2^n - p_1)(Z_3^n - p_1)(Z_4^n - p_1)] = \mathbf{O}(n^{-2}),$$

as was claimed. This completes the proof. \square

Proof of Theorem 2.3. The proof rests on the following simple yet crucial observation. Let Z_1, Z_2, \dots be a sequence of i.i.d. random variables having geometric distribution with parameter p . For integers $0 \leq m \leq n$, consider the conditional distribution of (Z_1, \dots, Z_{m+1}) given that $\sum_{i=1}^{m+1} Z_i = n - m$. Simple computations show that this conditional distribution is the same as the joint distribution of $(Y_1^n, \dots, Y_{m+1}^n)$. Once this is observed, the stated result then follows easily from a general result of Chen ([4], Theorem 6), using the additional fact that $\sum_{i=1}^{m+1} Z_i$ is a sufficient statistic for the mean $\mu = (1 - p)/p$ of the geometric distribution and that the variance $\sigma^2 = \mu(1 + \mu)$ is a continuous function of μ . \square

Proof of Theorem 2.4. Consider the space \mathbf{R}^∞ of all real sequences $\mathbf{x} = (x_l, l = 0, 1, \dots)$, equipped with the product topology. It is well known (see [2], p. 19) that weak convergence for probabilities on \mathbf{R}^∞ is determined by weak convergence of the finite-dimensional distributions. It thus follows easily from Theorem 2.3 that the sequence of \mathbf{R}^∞ -valued random variables $\mathbf{X}_n = \{\sqrt{m+1}(F_n(l) - F(l)), l = 0, 1, \dots\}$ converge weakly to a zero-mean discrete-time Gaussian process $\mathbf{X} = (X_0, X_1, \dots)$ with covariance kernel $\Sigma(l, k)$, as $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow p$. The transplant of the weak convergence result from \mathbf{R}^∞ to $D[0, \infty)$ (equipped with the Skorohod topology) is easily achieved via the continuous mapping theorem (see [2], p. 30), once one observes that the mapping $\phi : \mathbf{R}^\infty \rightarrow D[0, \infty)$ defined by $\phi(\mathbf{x})(t) = x_{[t]}$, $t \in [0, \infty)$ is continuous. This last observation follows easily from the criteria for convergence in $D[0, \infty)$ (see [2, p.112]). \square

Remark. Using the standard quantile transformation (see [12, p.97]) and the continuous mapping theorem, one can obtain the limiting distribution of the Kolmogorov-Smirnov statistic $\sup_t \sqrt{m+1}|F_n(t) - F(t)|$ to be that of $\sup_{t \geq 0} |W^0(1 - p^t)|$, where W^0 denotes the Brownian bridge. Clearly, this is not the usual Kolmogorov distribution.

Acknowledgment

L.-B. Chang gratefully acknowledges the partial support of the Center of Mathematical Modeling & Scientific Computing and the National Center for Theoretical Science, Hsinchu, Taiwan, and the National Science Council under grant 100-2115-M-009-007-MY2. Part of the research was done during the A. Goswami's visit to the Institute of Mathematics, Academia Sinica. He wishes to thank the Institute for their kind invitation and their tremendous hospitality. C.-R. Hwang has received funding from the National Science Council under grant number 101-2115-M-001-012-MY2, 98-2115-M-001-007-MY3 and 99-2118-M-001-007-MY2. Fushing truly appreciated all the generous supports from the Institute of Mathematics, Academia Sinica.

References

1. D. W. K. Andrews, Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* **4**(1988), 458-467.
2. P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, 1968.
3. Lo-Bin Chang, Shu-Chun Chen, Fushing Hsieh, Chii-Ruey Hwang and Max Palmer, Empirical invariance in stock market and related problems, manuscript, 2009.
4. Wen-Chen Chen, Limit theorems for general size distributions, *J. Appl. Probab.* **18**(1981), 139-147.
5. J. Davidson, An L_1 -convergence theorem for heterogeneous mixingale arrays with trending moments, *Statist. Probab. Lett.*, **16**(1994), 301-306.
6. R. M. De Jong, A strong law of large numbers for triangular mixingale arrays, *Statist. Probab. Lett.* **27**(1996), 1-9.
7. P. Diaconis and D. Freedman, Finite exchangeable sequences, *Ann. Probab.* **8** (1980), 745-764.
8. E. Eberlein, Application of generalized hyperbolic Lévy motions to finance, *Lévy Processes: Theory and Applications* (Edited by O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick), Birkhäuser, (2001), 319-336.

9. Stuart Geman and Lo-Bin Chang, Stock prices and the peculiar statistics of rare events, preprint, 2009.
10. Fushing Hsieh, Shu-Chun Chen and Chii-Ruey Hwang, Discovering stock dynamics through multidimensional volatility-phase, accepted by *Quantitative Finance* (2009).
11. Y. Hu, B. Oksendal and A. Sulem, Optimal portfolio in a fractional Black & Scholes Market, *Mathematical Physics and Stochastic Analysis: Essays in Honor of Ludwig Streit*, World Scientific, 267-279, 2000.
12. David Pollard, *Convergence of Stochastic Processes*, Springer Series in Statistics, xiv, Springer-Verlag, 1984.
13. H. Teicher, Almost certain convergence in double arrays, *Z. Wahrsch. Verw. Gebiete* **69**(1985), 331-345.