数理解析研究所講究録
第 1620 巻 2009 年 23-26

An Invariance Property for Exchangeable Sequence: Application to Stock Price Data

Alok Goswami
Division of Theoretical Statistics & Mathematics
Indian Statistical Institute, Kolkata

Lecture delivered in Workshop on "Stochastic Numerics", RIMS, Kyoto University, Kyoto, July 7 - 9 : 2008

1 Some Probability Limit Results

Consider n objects arranged in a row and suppose m are selected at random (that is, with equal probability for each of $\binom{n}{m}$ selections). If we label selected objects by 0s and the rest of the objects by 1s, we get a random n-long binary sequence with m many 0s and $n-m$ many 1s. There will be $(m+1)$ runs of 1s (of which, some may be possibly of length zero) separated by 0s. Let $Y_1^n, \ldots, Y_{m+1}^n$ denote the lengths of these $m+1$ runs of 1s. Then $Y_1^n, \ldots, Y_{m+1}^n$ is a sequence of nonnegative interger valued random variables, which are cleraly not independent (they add up to $n-m$!). Further, for every vector $(l_1, \ldots, l_{m+1})$ of non-negative integers with $l_1 + \cdots + l_{m+1} = n-m$, we have

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

Clarly, the sequence $Y_1^n, \ldots, Y_{m+1}^n$ is exchangeable, that is, the joint distribution is invariant undr permutations of coordinates.

The question that we ask is: what happens as $n \to \infty$?

We show that when m also grows with n in an appropriate way, the random variables Y_j^n, $1 \leq j \leq m+1$ behave asymptotically like an i.i.d. sequenec of geometric random variables.

Theorem 1: Let $n \longrightarrow \infty$ and let $m \sim np$ for some $p \in (0,1)$. Then, for any $k \geq 1$,

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

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

Next, we consider a slightly different question. Consider the probability histogram (relative frequencies) generated by the random variables $Y_1^n, \ldots, Y_{m+1}^n$. This will give a (random) probability distribution on non-negative integers with the probability mass functions

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

What do these probability histograms look like for large n? In other words, do the empirical distributions of the Y_j^n, $1 \leq j \leq m+1$ converge to a limit, as $n \rightarrow \infty$? If $Y_1^n, \ldots, Y_{m+1}^n$ were IID Geometric(p), then of course, the histograms would resemble, for large n, Geometric(p) distribution. This is classical rsult. But here the $Y_1^n, \ldots, Y_{m+1}^n$ are only asymptotically i.i.d. with Geometric(p) distribution. It turns out, however, that, with probability one, the (random) probability histograms genrated by the $Y_1^n, \ldots, Y_{m+1}^n$ will, for large n, still resemble a Geometric(p) distribution.

Theorem 2: Let $n \longrightarrow \infty$ and let $m \sim np$ for some $p \in (0,1)$. Then,

$$P\left[\lim_{n \to \infty} \theta_n(l) = p(1-p)^l, \quad l = 0, 1, \ldots \right] = 1.$$

Denote the empirical distribution for the random variables in the nth row by P_n. Then, by the well-known Scheffe's Theorem, one gets the following result.

Corollary: The distributions P_n converge, with probability 1, to the Geometric(p) distribution, in total variation as well as in Kolmogorov distance. Further, the convergence $\theta_n(l) \to p(1-p)^l$ is uniform in l, with probability 1.

In the next section, we outline a connection of the above results with analysis of stock price data, which was the main motivation for these results. In particular, the above results provide an invariance theorem in probability. Further details on the stock price analysis and the detailed proofs of the results may be found in [1] and [2].

2 Connection with stock-price data: An Invariance Result

Much of what follows is basd on the principle of what is called hierarchical segmentation of the Stock Price Time Series. Given prices of a stock at equal intervals of time, consider the times of occurences of extreme values for the returns over succesive time intervals. This will generate a certain subset from among the set of all time points considered. Now suppose that the assumed model for stock prices implies that the returns over successive time intervals (of equal length) are i.i.d. or, more generally, exchangeable. Then it is clear that in picking the times of occurences of extreme values of such returns, all subsets (of a fixed size) from among the set of all time points are equally likely to show up.

To elaborate, let $\alpha, \beta \geq 0$ with $0 < \alpha + \beta < 1$. From a set of values $(x_1, x_2, \ldots, x_n)$, we want to choose those that form the lower 100α-percentile and those that form the upper 100β-percentile. It is clear that if k and l are integers satisfying

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

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 following theorem says that in case the data points are realizations of n exchangeable random variables, then this amounts to selecting $k + n - l + 1$ objects at random from a set of n objects. This gives the connecting link between analysis of stock price data and the limit results in the previous section.

Theorem 3: If $X_1, \ldots, X_n$ 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.

The importance of the above lies in the fact that under many of the standard theoretical models of stock prices (starting from the classical Black-Scholes' Geometric Brownian Motion model to the more recent Geometric Levy Process Model), the returns over successive intervals of time (of equal length) are i.i.d. Our results also cover the case when such returns are merely exchangeable, as is the case of Geometric Fractional Brownian Motion with Hurst index = 1. Our results would suggest that under any of these models, the histograms (empirical distributions) of the successive gaps between locations of extremes in the stock price returns should be close to an appropriate Geometric distribution, at least for large n. An outline of the findings with real-life stock price data is contained in the talk given by my collaborator Chii-Ruey Hwang and in his article in this volume. A number of interesting problems remain open and are being looked into.

3 References

[1] Chang Lo-Bin, Alok Goswami, Chii-Ruey Hwang (2008),*An Invariance Property for some Empirical Distributions with Applications to Finance*, manuscript.

[2] Chang Lo-Bin, Shu-Chun Chen, Fushing Hsieh, Chii-Ruey Hwang (2008),Max Palmer *An Empirical Invariance for the Stock Price*, in preparation.