# A martingale approach to the study of occurrence of sequence patterns in repeated experiments

MATH 833-Fall 2012 Presenter: Dae Han Kang

## 1. Introduction

This paper applies the concept of stopping times of martingales to problems regarding the occurrence of sequence patterns in repeated experiments. For every given set of finite sequences, we compute the expected time till one of them occurs. Moreover we compute the probability for each sequence to be the first occurrence. Given a collection of n finite sequences of outcomes, Theorem 3.1. gives a system of n+1 linear equations on these n+1 quantities(the expected waiting time and one probability for each competing sequence).

## 2. Basic setup and definitions

Consider a random sequence with countable state space Σ. Let $Z_1, Z_2, \cdots$ be a sequence of i.i.d. random variables having the same distribution as a fixed random variable Z. For given sequences $A=(a_1, a_2, \cdots, a_m)$ and $B$ over Σ, define stopping times $N_B$ and $N_{AB}$ as follows.

$N_B$=min{ $k$ |$B$ is a connected subsequence of $(Z_1, \cdots, Z_k)$ }, that is, $N_B$ is the waiting time until $B$ appears as a run in the process $Z_1, Z_2, \cdots$.
Similarly define
$N_{AB}$=min{ $k$ |$B$ is a connected subsequence of $(a_1, \cdots a_m, Z_1, \cdots, Z_k)$ }.

**Definition 2.2** For every pair $(i,j)$ of integers, write

$$\delta_{ij} = P(Z=b_j)^{-1} \quad \text{if } 1 \le i \le m, 1 \le j \le n \text{ and } a_i = b_j$$
$$= 0 \quad \text{otherwise.}$$

Then define $A*B = \delta_{11}\delta_{22}\cdots\delta_{mm} + \delta_{21}\delta_{32}\cdots\delta_{m,m-1} + \cdots + \delta_{m1}$ .(2.3)

## 3. Main Results

**Theorem 3.1.** Let $Z, Z_1, Z_2, \cdots$ be discrete i.i.d random variables and $A_1, \cdots, A_n$ be finite sequences of possible values of $Z$ not containing one another. Let $A$ be another such sequence not containing any $A_i$. Given the starting sequence $A$, let $p_i$ be the probability that $A_i$ precedes the remaining n-1 sequences in a realization of the process $Z_1, Z_2, \cdots$. Then for every $i$,

$$\sum_{j=1}^{n} p_j A_j{}^* A_i = EN + A^* A_i \qquad (3.1)$$

where $N$ is the stopping time when any $A_j$ occurs. In particular when $A$ is void, we have, for every $i$,

$$\sum_{j=1}^{n} p_j A_j{}^* A_i = EN \qquad (3.2)$$

## Outline of the proof

This theorem is based on the following lemma.

**Lemma 2.4.** Given a starting sequence $A$, the expected waiting time for a sequence $B$ is $EN_{AB} = B^* B - A^* B$, provided that $B$ is not a connected subsequence of $A$. In particular the expected waiting tie of the sequence $B$ (without a starting sequence) is $B^* B$.

**Sketch of the proof.**

1) $EN_{AB}$ is dominated by a geometric random variable so that it is finite.

2) Choose a suitable martingale to compute $EN_{AB}$

$$X_k = (a_1, \cdots, a_m, Z_1, \cdots, Z_k)^* B - k \quad (k = 0, 1, 2, \cdots)$$

3) Consider the stopped process $X_{k \wedge N_{AB}}$

We see that $X_{k \wedge N_{AB}}$ is also a martingale and $X_{N_{AB}} = B^* B - N_{AB}$.

4) Show that $X_{k \wedge N_{AB}}$ is uniformly integrable.

We have $|X_{k \wedge N_{AB}}| \le |(a_1, \cdots, a_m, Z_1, \cdots, Z_k)^* B| + k \le B^* B + N_{AB}$

5) Conclusion

Since $X_{k \wedge N_{AB}}$ is uniformly integrable, $EX_{N_{AB}} = EX_0 = A^* B$.

Therefore $EN_{AB} = B^* B - A^* B$ by 3).

References

[1] Shuo-Yen Robert Li(1980). A Martingale Approach to the Study of Occurrence of Sequence Patterns in Reapeated Experiments. Ann.Probab., Volume 8,No.6,1171-1176.