On the waiting time till some patterns occur in i.i.d. sequences

Krzysztof Zajkowski

Institute of Mathematics, University of Białystok

Będlewo, July 9–13 2013

Patterns in independent sequences

Let (ξ_n) be a sequence of i.i.d. random letters from a finite alphabet Ω.

Patterns in independent sequences

Let (ξ_n) be a sequence of i.i.d. random letters from a finite alphabet Ω.
By a pattern (word) A of the length l we mean a finite ordered sequence of letters $a_1 a_2 ... a_l$.

Patterns in independent sequences

Let (ξ_n) be a sequence of i.i.d. random letters from a finite alphabet Ω.
By a pattern (word) A of the length l we mean a finite ordered sequence of letters $a_1 a_2 ... a_l$.
We are interested in a random variable τ_A of the first time that one observes the word A as a run in the sequence (ξ_n), i.e.

$$\tau_A = \inf\{n : \ \xi_{n-l+1} = a_1, \ \xi_{n-l+2} = a_2, ..., \xi_n = a_l\}.$$

Patterns in independent sequences

Let (ξ_n) be a sequence of i.i.d. random letters from a finite alphabet Ω.
By a pattern (word) A of the length l we mean a finite ordered sequence of letters $a_1 a_2 ... a_l$.
We are interested in a random variable τ_A of the first time that one observes the word A as a run in the sequence (ξ_n), i.e.

$$\tau_A = \inf\{n : \ \xi_{n-l+1} = a_1, \ \xi_{n-l+2} = a_2, ..., \xi_n = a_l\}.$$

What is the expected value of τ_A?

Gambling team technique; S.-Y. R. Li (1980)

A casino generates the sequence of letters (ξ_n).
Consider a flow of gamblers (a gambling team) visiting the casino.

Gambling team technique; S.-Y. R. Li (1980)

A casino generates the sequence of letters (ξ_n).
Consider a flow of gamblers (a gambling team) visiting the casino.
The nth gambler arrives right before ξ_n will be observed and places \$1 bet that $\xi_n = a_1$. If ξ_n is not a_1 the gambler loses his dollar.
If $\xi_n = a_1$ the casino pays fair odds $\frac{1}{Pr(\xi_n = a_1)}$.

Gambling team technique; S.-Y. R. Li (1980)

A casino generates the sequence of letters (ξ_n).
Consider a flow of gamblers (a gambling team) visiting the casino.
The nth gambler arrives right before ξ_n will be observed and places \$1 bet that $\xi_n = a_1$. If ξ_n is not a_1 the gambler loses his dollar.
If $\xi_n = a_1$ the casino pays fair odds $\frac{1}{Pr(\xi_n = a_1)}$. Next the gambler bets his entire capital on $\xi_{n+1} = a_2$. If it is not a_2, he goes home with nothing, otherwise he increases his capital by factor $\frac{1}{Pr(\xi_{n+1} = a_2)}$.

Gambling team technique; S.-Y. R. Li (1980)

A casino generates the sequence of letters (ξ_n).
Consider a flow of gamblers (a gambling team) visiting the casino.
The nth gambler arrives right before ξ_n will be observed and places \$1 bet that $\xi_n = a_1$. If ξ_n is not a_1 the gambler loses his dollar.
If $\xi_n = a_1$ the casino pays fair odds $\frac{1}{Pr(\xi_n = a_1)}$. Next the gambler bets his entire capital on $\xi_{n+1} = a_2$. If it is not a_2, he goes home with nothing, otherwise he increases his capital by factor $\frac{1}{Pr(\xi_{n+1} = a_2)}$.
Then he continuous in the same fashion until the entire word A is exhausted. If the gambler is lucky he leaves the game with total winnings of

$$[Pr(\xi_n = a_1)Pr(\xi_{n+1} = a_2)...Pr(\xi_{n+l-1} = a_l)]^{-1}$$

dollars, otherwise he loses his initial bet \$1.

Let X_n denote the total net gain of the casino. One can observe that $(X_n, \sigma(\xi_1, ..., \xi_n))$ is a martingale.

Let X_n denote the total net gain of the casino. One can observe that $(X_n, \sigma(\xi_1, ..., \xi_n))$ is a martingale.
The stopped martingale X_{τ_A} is given by

$$X_{\tau_A} = \tau_A - A * A,$$

where $A * A$ is the total winning of gamblers by time τ_A and can be described as follows.

Let X_n denote the total net gain of the casino. One can observe that $(X_n, \sigma(\xi_1, ..., \xi_n))$ is a martingale.
The stopped martingale X_{τ_A} is given by

$$X_{\tau_A} = \tau_A - A * A,$$

where $A * A$ is the total winning of gamblers by time τ_A and can be described as follows.
Let $A_{(k)}$ and $A^{(k)}$ denote subpatterns formed by first and last k letters of A, respectively; i.e.
$A_{(k)} = a_1 a_2 ... a_k$ and $A^{(k)} = a_{l-k+1} a_{l-k+2} ... a_l$.

Let X_n denote the total net gain of the casino. One can observe that $(X_n, \sigma(\xi_1, ..., \xi_n))$ is a martingale.
The stopped martingale X_{τ_A} is given by

$$X_{\tau_A} = \tau_A - A * A,$$

where $A * A$ is the total winning of gamblers by time τ_A and can be described as follows.
Let $A_{(k)}$ and $A^{(k)}$ denote subpatterns formed by first and last k letters of A, respectively; i.e.
$A_{(k)} = a_1 a_2 ... a_k$ and $A^{(k)} = a_{l-k+1} a_{l-k+2} ... a_l$.
We take a notation: $[A_{(k)} = A^{(k)}] = 1$ if $A_{(k)} = A^{(k)}$ and $[A_{(k)} = A^{(k)}] = 0$ if not.

Let X_n denote the total net gain of the casino. One can observe that $(X_n, \sigma(\xi_1, ..., \xi_n))$ is a martingale.
The stopped martingale X_{τ_A} is given by

$$X_{\tau_A} = \tau_A - A * A,$$

where $A * A$ is the total winning of gamblers by time τ_A and can be described as follows.
Let $A_{(k)}$ and $A^{(k)}$ denote subpatterns formed by first and last k letters of A, respectively; i.e.
$A_{(k)} = a_1 a_2 ... a_k$ and $A^{(k)} = a_{l-k+1} a_{l-k+2} ... a_l$.
We take a notation: $[A_{(k)} = A^{(k)}] = 1$ if $A_{(k)} = A^{(k)}$ and $[A_{(k)} = A^{(k)}] = 0$ if not. Then

$$A * A = \sum_{k=1}^{l} \frac{[A_{(k)} = A^{(k)}]}{Pr(A_{(k)})} = \sum_{k=1}^{l} \frac{[A_{(k)} = A^{(k)}]}{Pr(\xi_1 = a_1) ... Pr(\xi_k = a_k)}.$$

Employing the optional stopping theorem we get

$$0 = EX_{\tau_A} = E\tau_A - A * A$$

Employing the optional stopping theorem we get

$$0 = EX_{\tau_A} = E\tau_A - A * A$$

and, as a consequence,

$$E\tau_A = A * A = \sum_{k=1}^{l} \frac{[A_{(k)} = A^{(k)}]}{Pr(A_{(k)})}.$$

Employing the optional stopping theorem we get

$$0 = EX_{\tau_A} = E\tau_A - A * A$$

and, as a consequence,

$$E\tau_A = A * A = \sum_{k=1}^{l} \frac{[A_{(k)} = A^{(k)}]}{Pr(A_{(k)})}.$$

Example

Let $\Omega = \{H, T\}$, $Pr(\xi = H) = p$ *and* $Pr(\xi = T) = 1 - p = q$.

Employing the optional stopping theorem we get

$$0 = EX_{\tau_A} = E\tau_A - A * A$$

and, as a consequence,

$$E\tau_A = A * A = \sum_{k=1}^{l} \frac{[A_{(k)} = A^{(k)}]}{Pr(A_{(k)})}.$$

Example

Let $\Omega = \{H, T\}$, $Pr(\xi = H) = p$ *and* $Pr(\xi = T) = 1 - p = q$. *Consider two patterns:*

$$A = \underbrace{HH...H}_{l} \quad \text{and} \quad B = T\,\underbrace{HH...H}_{l-1}.$$

Employing the optional stopping theorem we get

$$0 = EX_{\tau_A} = E\tau_A - A * A$$

and, as a consequence,

$$E\tau_A = A * A = \sum_{k=1}^{l} \frac{[A_{(k)} = A^{(k)}]}{Pr(A_{(k)})}.$$

Example

Let $\Omega = \{H, T\}$, $Pr(\xi = H) = p$ *and* $Pr(\xi = T) = 1 - p = q$.
Consider two patterns:

$$A = \underbrace{HH...H}_{l} \quad \text{and} \quad B = T\,\underbrace{HH...H}_{l-1}.$$

Then $E\tau_A = \frac{1}{p} + \frac{1}{p^2} + ... + \frac{1}{p^l}$ *and* $E\tau_B = \frac{1}{qp^{l-1}}$. *For* $p = q = \frac{1}{2}$,
$E\tau_A = 2^{l+1} - 2$ *and* $E\tau_B = 2^l$.

The generating function of τ_A

Modifying the preceding method one can obtain a formula for the generating function of τ_A. Let $0 < \alpha < 1$,

$$\begin{aligned} X_{\tau_A} &= 1 + \alpha + \alpha^2 + ... + \alpha^{\tau_A - 1} - \alpha^{\tau_A}(A * A)(\alpha) \\ &= \alpha^{\tau_A}\Big(\frac{1}{\alpha - 1} - (A * A)(\alpha)\Big) + \frac{1}{1 - \alpha}, \end{aligned}$$

The generating function of τ_A

Modifying the preceding method one can obtain a formula for the generating function of τ_A. Let $0 < \alpha < 1$,

$$\begin{aligned} X_{\tau_A} &= 1 + \alpha + \alpha^2 + ... + \alpha^{\tau_A - 1} - \alpha^{\tau_A}(A * A)(\alpha) \\ &= \alpha^{\tau_A}\Big(\frac{1}{\alpha - 1} - (A * A)(\alpha)\Big) + \frac{1}{1 - \alpha}, \end{aligned}$$

where

$$(A * A)(\alpha) = \sum_{k=1}^{l} \frac{[A_{(k)} = A^{(k)}]}{Pr(A_{(k)})\alpha^k}.$$

Observe that $(A * A)(1) = A * A$.

The generating function of τ_A

The optional stopping theorem implies

$$0 = E(X_{\tau_A}) = E(\alpha^{\tau_A})\Big(\frac{1}{\alpha - 1} - (A * A)(\alpha)\Big) + \frac{1}{1 - \alpha}.$$

The generating function of τ_A

The optional stopping theorem implies

$$0 = E(X_{\tau_A}) = E(\alpha^{\tau_A})\Big(\frac{1}{\alpha - 1} - (A * A)(\alpha)\Big) + \frac{1}{1-\alpha}.$$

Solving this relation for $E(\alpha^{\tau_A})$ we obtain

$$E(\alpha^{\tau_A}) = \frac{1}{1 + (1-\alpha)(A * A)(\alpha)}.$$

Competing patterns

Consider a set of m patterns (words) A_i $(1 \leqslant i \leqslant m)$ of lengths l_i, respectively.

Competing patterns

Consider a set of m patterns (words) A_i $(1 \leqslant i \leqslant m)$ of lengths l_i, respectively.
We assume that the set of patterns is reduced that is none of the patterns contains any other as a subpattern.

Competing patterns

Consider a set of m patterns (words) A_i ($1 \leqslant i \leqslant m$) of lengths l_i, respectively.
We assume that the set of patterns is reduced that is none of the patterns contains any other as a subpattern.
Let $\tau_i = \tau_{A_i}$ denote the stopping time until A_i occurs and τ be the stopping time till some of considered patterns is observed, i.e.

$$\tau = \min\{\tau_i : \ 1 \leqslant i \leqslant m\}.$$

Correlation functions of patterns

Let A and B be two patterns of the lengths l and m, respectively. We define a correlation function $(B * A)(\alpha)$ as

$$(B * A)(\alpha) = \sum_{k=1}^{\min\{l,m\}} \frac{[A_{(k)} = B^{(k)}]}{Pr(A_{(k)})\alpha^k}.$$

Example

Let $\Omega = \{H, T\}$ *and* (ξ_n) *be a sequence of i.i.d. letters in* Ω *with the distribution*

$$Pr(\xi_n = H) = p \quad \text{and} \quad Pr(\xi_n = T) = q = 1 - p.$$

Example

Let $\Omega = \{H, T\}$ *and* (ξ_n) *be a sequence of i.i.d. letters in* Ω *with the distribution*

$$Pr(\xi_n = H) = p \quad \text{and} \quad Pr(\xi_n = T) = q = 1 - p.$$

Consider two patterns $A = THH$ *and* $B = THTH$*. Then the correlation functions have the following forms:*

$$(B * A)(\alpha) = \frac{1}{pq\alpha^2}, \quad (A * B)(\alpha) = 0,$$

$$(A * A)(\alpha) = \frac{1}{p^2q\alpha^3}, \quad (B * B)(\alpha) = \frac{1}{pq\alpha^2} + \frac{1}{p^2q^2\alpha^4}.$$

Probability-generating functions

Let

$$g_\tau(\alpha) := E(\alpha^\tau) = \sum_{n=0}^{\infty} Pr(\tau = n)\alpha^n$$

the probability-generating function of random variable τ and

$$g_\tau^{A_i}(\alpha) = E(\alpha^\tau \mathbf{1}_{\{\tau=\tau_i\}}) = \sum_{n=0}^{\infty} Pr(\tau = \tau_i = n)\alpha^n, \quad 1 \leqslant i \leqslant m.$$

Since $Pr(\tau = n) = \sum_{i=1}^{m} Pr(\tau = \tau_i = n)$
we have that $g_\tau(\alpha) = \sum_{i=1}^{m} g_\tau^{A_i}(\alpha)$.

The system of equations; H. Gerber, S-Y.R. Li (1981)

One can obtain the following system of linear equations

$$\begin{cases} g_\tau(\alpha) - \sum_{j=1}^{m} g_\tau^{A_j}(\alpha) & = \quad 0 \\ g_\tau(\alpha) + (1-\alpha)\sum_{j=1}^{m}(A_j * A_i)(\alpha) g_\tau^{A_j}(\alpha) & = \quad 1 \quad (1 \leqslant i \leqslant m), \end{cases}$$

The system of equations; H. Gerber, S-Y.R. Li (1981)

One can obtain the following system of linear equations

$$\begin{cases} g_\tau(\alpha) - \sum_{j=1}^m g_\tau^{A_j}(\alpha) & = & 0 \\ g_\tau(\alpha) + (1-\alpha)\sum_{j=1}^m (A_j * A_i)(\alpha) g_\tau^{A_j}(\alpha) & = & 1 \quad (1 \leqslant i \leqslant m), \end{cases}$$

which is equivalent to this one

$$\sum_{j=1}^m [1 - (1-\alpha)(A_j * A_i)(\alpha)] g_\tau^{A_j}(\alpha) = 1 \quad (1 \leqslant i \leqslant m). \qquad (1)$$

Let $\mathcal{A}$ denotes a matrix formed by correlations functions $A_j * A_i$, i.e.

$$\mathcal{A}(\alpha) = \Big[(A_j * A_i)(\alpha)\Big]_{1 \leqslant i,j \leqslant m}$$

and $\mathcal{A}^j(\alpha)$ is the matrix arisen by replacing the j-th column of $\mathcal{A}(\alpha)$ by the column vector $[1]_{1 \leqslant i \leqslant m}$.

Let $\mathcal{A}$ denotes a matrix formed by correlations functions $A_j * A_i$, i.e.

$$\mathcal{A}(\alpha) = \Big[(A_j * A_i)(\alpha)\Big]_{1 \leqslant i,j \leqslant m}$$

and $\mathcal{A}^j(\alpha)$ is the matrix arisen by replacing the j-th column of $\mathcal{A}(\alpha)$ by the column vector $[1]_{1 \leqslant i \leqslant m}$.

Theorem

The solution of the system of linear equations (1) has the following form

$$g_{\tau}^{A_i}(\alpha) = \frac{\det \mathcal{A}^i(\alpha)}{\sum_{j=1}^{m} \det \mathcal{A}^j(\alpha) + (1-\alpha) \det \mathcal{A}(\alpha)} \quad (1 \leqslant i \leqslant m).$$

Corollary

The probability $Pr(\tau = \tau_i)$ *that the pattern* A_i *precedes all the remaining* $m-1$ *patterns is equal to* $g_\tau^{A_i}(1)$*, that is*

$$Pr(\tau = \tau_i) = \frac{\det \mathcal{A}^i(1)}{\sum_{j=1}^m \det \mathcal{A}^j(1)}.$$

Conway's formula

Let us emphasize that the above corollary is the generalization of the Conway's formula. For two patterns we get

$$\frac{Pr(\tau = \tau_1)}{Pr(\tau = \tau_2)} = \frac{\det \mathcal{A}^1}{\det \mathcal{A}^2} = \det \begin{bmatrix} 1 & (A_2 * A_1) \\ 1 & (A_2 * A_2) \end{bmatrix} : \det \begin{bmatrix} (A_1 * A_1) & 1 \\ (A_1 * A_2) & 1 \end{bmatrix}$$

$$= \frac{(A_2 * A_2) - (A_2 * A_1)}{(A_1 * A_1) - (A_1 * A_2)}.$$

The expected waiting time of τ

Corollary

The expected waiting time till one of patterns is observed is given by

$$E\tau = \frac{\det \mathcal{A}(1)}{\sum_{j=1}^{m} \det \mathcal{A}^j(1)}.$$

The expected waiting time of τ

Corollary

The expected waiting time till one of patterns is observed is given by

$$E\tau = \frac{\det \mathcal{A}(1)}{\sum_{j=1}^{m} \det \mathcal{A}^{j}(1)}.$$

For one pattern A its expected waiting time is equal to $A * A$.

- H. Gerber, S-Y.R. Li, *The occurrence of sequence of patterns in repeated experiments and hitting times in a Markov chain*, Stochastic Processes and Their Applications, 11, 101-108 (1981).
- S-Y.R. Li, *A martingale approach to the study of occurrence of sequence patterns in repeated experiments*, The Annals of Probability, Vol. 8. (1980), 1171-1176.
- U. Ostaszewska, K. Zajkowski, *On the waiting time till some patterns occur in i.i.d. sequences*, arXiv:1302.4859.
- K. Zajkowski, *Penney's game between many players*, arXiv:1212.3973.