ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΕΓΙΣΤΟ ΜΗΚΟΣ ΡΟΗΣ ΕΠΙΤΥΧΙΩΝ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΜΑΘΗΜΑΤΙΚΑ ΤΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΤΩΝ ΑΠΟΦΑΣΕΩΝ»

Αλμπάνης Πυθαγόρας
Α.Μ. 233

Επιβλέπουσα Καθηγήτρια
Ευφροσύνη Σ. Μακρή

Πάτρα, Οκτώβριος 2010

Η εργασία αυτή έγινε στα πλαίσια του Διατμηματικού Προγράμματος Μεταπτυχιακών Σπουδών «Μαθηματικά των Υπολογιστών και των Αποφάσεων» υπό την επίβλεψη της Επίκουρου Καθηγήτριας Ευφροσύνης Μακρή.

Θα ήθελα να ευχαριστήσω την κυρία Μακρή για την καθοδήγησή της και την πολύτιμη βοήθειά της στην εκπόνηση αυτής της διπλωματικής εργασίας.

Περιεχόμενα

Εισαγωγή

Θεωρούμε μια ακολουθία αποτελεσμάτων n δυαδικών πειραμάτων διατεταγμένων σε γραμμή. Το αποτέλεσμα κάθε πειράματος είναι επιτυχία (S ή 1) ή αποτυχία (F ή 0). Ροή επιτυχιών είναι μια ακολουθία από συνεχόμενες επιτυχίες των οποίων προηγείται και έπεται μια αποτυχία ή τίποτε (αν η ροή είναι στην αρχή ή στο τέλος της ακολουθίας). Μήκος μιας ροής επιτυχιών είναι ο αριθμός των επιτυχιών που περιλαμβάνονται στη ροή.

Η μελέτη των τυχαίων μεταβλητών που σχετίζονται με ροές ήταν ιδιαίτερα αποτελεσματική σε πολλά επιστημονικά πεδία. Από το 1940 αποτέλεσαν τη βάση, στη δημιουργία ελέγχων υποθέσεων ($Wald\ and\ Wolfowitz$, 1940) και στον ποιοτικό έλεγχο ($Mosteler$, 1941, $Wolfowitz$, 1943). Στις μέρες μας, πέρα από τη Στατιστική, βρίσκει εφαρμογές και σε άλλες επιστημονικές περιοχές όπως στη Βιολογία (ακολουθίες DNA), στην Οικολογία και στην Αξιοπιστία μηχανικών συστημάτων.

Σκοπός της εργασίας αυτής είναι να παρουσιάσει μια επισκόπηση αποτελεσμάτων που αφορούν στη μελέτη της κατανομής της τυχαίας μεταβλητής L_n, η οποία παριστάνει το μέγιστο μήκος ροής επιτυχιών σε n δυαδικά περάματα. Η μελέτη γίνεται για την τυχαία μεταβλητή ορισμένη σε ακολουθίες ανεξάρτητων, ανταλλάξιμων και Μαρκοβιανά εξαρτημένων δυαδικών μεταβλητών. Αναπτύσσονται, επίσης, οι μέθοδοι που έχουν χρησιμοποιηθεί για τον προσδιορισμό της κατανομής της μελετούμενης τυχαίας μεταβλητής.

Συγκεκριμένα, στο πρώτο κεφάλαιο θα ασχοληθούμε με ανεξάρτητες (όχι κατ΄ ανάγκη ισόνομες) δοκιμές $Bernoulli$ και χρησιμοποιώντας συνδυαστική ανάλυση δίνουμε τη συνάρτηση πιθανότητας καθώς και τη συνάρτηση κατανομής της τυχαίας μεταβλητής. Σημαντική είναι η συνεισφορά της μελέτης των τυχαίων μεταβλητών $G_{n,k}$ και $N_{n,k}$ που παριστάνουν τον αριθμό των ροών επιτυχιών μήκους τουλάχιστον k και μήκους k, αντίστοιχα, σε n δυαδικά πειράματα. Αντίστοιχες μαθηματικές εκφράσεις, για τη συνάρτηση κατανομής της τυχαίας μεταβλητής L_n, αποδεικνύονται στο δεύτερο κεφάλαιο στο οποίο ασχολούμαστε με την τυχαία μεταβλητή ορισμένη σε ακολουθία ανταλλάξιμων δυαδικών δοκιμών καθώς και σε ακολουθία που προκύπτει από τη μέθοδο δειγματοληψίας $Pólya - Eggenberger$, ως ειδική περίπτωση της ανταλλαξιμότητας.

Στο τρίτο κεφάλαιο, ύστερα από μια σύντομη εισαγωγή στη θεωρία των Μαρκοβιανών αλυσίδων, θα προσδιορίσουμε την κατανομή της L_n με τη μέθοδο εμφύτευσης μιας τυχαίας μεταβλητής σε Μαρκοβιανή αλυσίδα στην περίπτωση των ανεξάρτητων (όχι κατ΄ ανάγκη ισόνομων) δοκιμών *Bernoulli*. Επίσης, θα δοθεί μια μαθηματική έκφραση, καθώς και φράγματα για τη συνάρτηση κατανομής του μέγιστου μήκους ροής επιτυχιών στην περίπτωση των μη ομογενών Μαρκοβιανά εξαρτημένων δυαδικών μεταβλητών.

Τέλος, θα αναφερθούμε στην εφαρμογή της κατανομής της τυχαίας μεταβλητής L_n στην αξιοπιστία των συνεχόμενων k-από-τα-n συστημάτων αποτυχίας. Ένα τέτοιο σύστημα αποτυγχάνει αν και μόνο, αν μεταξύ των n συνιστωσών του εμφανισθούν τουλάχιστον k συνεχόμενες αποτυχημένες συνιστώσες. Αφού παρουσιάσουμε εκφράσεις για τον υπολογισμό της αξιοπιστίας αυτών των συστημάτων, θα παραθέσουμε κάποια παραδείγματα στα οποία θα γίνεται περισσότερο εμφανής η σύνδεση αυτής με την τυχαία μεταβλητή L_n.

Στη συνέχεια δίνουμε ένα παράδειγμα για την κατανόηση των τυχαίων μεταβλητών που θα χρησιμοποιηθούν. Έστω ότι έχουμε μια ακολουθία αποτελούμενη από $n = 10$ δοκιμές *Bernoulli*,

$$1100111101$$

Τότε, έχουμε:

- Μια ροή επιτυχιών μήκους 2
- Μια ροή αποτυχιών μήκους 2
- Μια ροή επιτυχιών μήκους 4
- Μια ροή αποτυχιών μήκους 1
- Μια ροή επιτυχιών μήκους 1

Επομένως, $L_{10} = 4$, $G_{10,2} = 2$, $N_{10,2} = 3$ ενώ $G_{10,5} = N_{10,5} = 0$.

Κεφάλαιο 1

Μελέτη της μεταβλητής L_n σε ανεξάρτητες δοκιμές *Bernoulli*

1.1 Δεσμευμένη συνάρτηση κατανομής της μεταβλητής L_n δοθέντος του αριθμού των επιτυχιών στην ακολουθία

Στην παράγραφο αυτή θα δώσουμε μια έκφραση για την δεσμευμένη συνάρτηση κατανομής της μέγιστης ροής επιτυχιών δοθέντος του αριθμού των επιτυχιών σε n ανεξάρτητες και ισόνομες δοκιμές *Bernoulli*.Έτσι λοιπόν, αν συμβολίσουμε με S_n το συνολικό αριθμό των επιτυχιών σε μια ακολουθία από n δοκιμές *Bernoulli* τότε σύμφωνα με τους *Philippou* και *Makri* (1986) έχουμε το παρακάτω λήμμα που μας δίνει την πιθανότητα $P(L_n \leq k, S_n = r)$.

Λήμμα 1.1 *Αν L_n και S_n είναι τυχαίες μεταβλητές που παριστάνουν αντίστοιχα το μήκος της μέγιστης ροής επιτυχιών και το συνολικό αριθμό επιτυχιών σε n ανεξάρτητες δοκιμές Bernoulli με πιθανότητα επιτυχίας p $(0 < p < 1)$, τότε,*

$$P(L_n \leq k, S_n = r) = p^r q^{n-r} \sum_{i=0}^{k} \sum_{x_1,\ldots,x_{k+1}} \binom{x_1 + \ldots + x_{k+1}}{x_1, \ldots, x_{k+1}}, \qquad (1)$$

$0 \leq k \leq r \leq n$, όπου το εσωτερικό άθροισμα ορίζεται πάνω σε όλους τους μη αρνητικούς ακεραίους $x_1, x_2, \ldots, x_{k+1}$ οι οποίοι ικανοποιούν τους περιορισμούς $x_1 + 2x_2 + \ldots + (k+1)x_{k+1} = n - i$ και $x_1 + \ldots + x_{k+1} = n - r$.

Απόδειξη Οι *Philippou* και *Makri* (1986) κατάφεραν να αναπαραστήσουν τα n πειράματα μιας ακολουθίας, κατηγοριοποιώντας τα, ως εξής:
Χρησιμοποίησαν τα σύμβολα e_i και τους μη αρνητικούς ακεραίους x_i και έθεσαν ως:
$\rightarrow e_1 : F$, και θα έχουμε x_1 από αυτά

$\rightarrow e_2 : SF$, και θα έχουμε x_2 από αυτά

$\vdots$

$\rightarrow e_k : \underbrace{SS\cdots S}_{k-1}F$, και θα έχουμε x_k από αυτά

$\rightarrow e_{k+1} : \underbrace{SS\cdots S}_{k}F$, και θα έχουμε x_{k+1} από αυτά

$\rightarrow i$: είναι ο αριθμός των συνεχόμενων επιτυχιών που βρίσκονται στο τέλος της ακολουθίας με δυνατές τιμές $i = 0, 1, \ldots, k$.
Ένα τυπικό στοιχείο του γεγονότος $(L_n \leq k, S_n = r)$ είναι μια διάταξη της μορφής:

$$a_1 a_2 \ldots a_{x_1+\ldots+x_{k+1}} \underbrace{SS\ldots S}_{i},\ 0 \leq i \leq k$$

η οποία περιλαμβάνει x_1 από τα e_1, x_2 από τα e_2,... έτσι όπως έχουν ορισθεί.

Για σταθερά $x_1, x_2, \ldots, x_{k+1}$ το πλήθος των παραπάνω ακολουθιών είναι:

$$\binom{x_1 + \ldots + x_{k+1}}{x_1, \ldots, x_{k+1}} \tag{2}$$

και η πιθανότητα καθεμιάς εξ' αυτών είναι

$$\begin{aligned} & P(a_1 a_2 \ldots a_{x_1+\ldots+x_{k+1}} \underbrace{SS\ldots S}_{i}) \\ = \quad & P(\{e_1\})^{x_1} P(\{e_2\})^{x_2} \ldots P(\{e_{k+1}\})^{x_{k+1}} P(\underbrace{SS\ldots S}_{i}). \end{aligned}$$

Όμως, επειδή οι δοκιμές είναι ανεξάρτητες με την ίδια πιθανότητα επιτυχίας $(P(S) = p)$, έχουμε ότι:

$$\begin{aligned} P(a_1 a_2 \ldots a_{x_1+\ldots+x_{k+1}} \underbrace{SS\ldots S}_{i}) &= q^{x_1} p^{x_2} q^{x_2} p^{2x_3} \ldots p^{kx_{k+1}} q^{x_{k+1}} p^i \\ &= q^{x_1+x_2+\ldots+x_{k+1}} p^{x_2+2x_3+\ldots+kx_{k+1}+i} \\ &= q^{n-r} p^{x_2+2x_3+\ldots+kx_{k+1}+i} \\ &= q^{n-r} p^r, \end{aligned}$$

έτσι ώστε να ικανοποιούνται οι σχέσεις $x_1 + 2x_2 + \ldots + (k+1)x_{k+1} = n - i$ και $x_1 + \ldots + x_{k+1} = n - r$.

Έτσι λοιπόν αν αθροίσουμε πάνω σε όλες τις δυνατές τιμές των ακεραίων $x_1, x_2, \ldots x_{k+1}$ και σ' αυτές του δείκτη i, μπορούμε να γράψουμε για τη ζητούμενη πιθανότητα:

$$P(L_n \leq k, S_n = r) = p^r q^{n-r} \sum_{i=0}^{k} \sum_{x_1,\ldots,x_{k+1}} \begin{pmatrix} x_1 + \ldots + x_{k+1} \\ x_1, \ldots, x_{k+1} \end{pmatrix}. \qquad (3)$$

$\diamond$

Με τη χρήση του τύπου της πιθανότητας $P(L_n \leq k, S_n = r)$, μπορούμε να αποδείξουμε τη δεσμευμένη συνάρτηση κατανομής της L_n δοθέντος του αριθμού των επιτυχιών.

Θεώρημα 1.1 *Έστω S_n και L_n όπως στο Λήμμα 1.1. Τότε,*

$$P(L_n \leq k | S_n = r) = \begin{pmatrix} n \\ r \end{pmatrix}^{-1} \sum_{i=0}^{k} \sum_{x_1,\ldots,x_{k+1}} \begin{pmatrix} x_1 + \ldots + x_{k+1} \\ x_1, \ldots, x_{k+1} \end{pmatrix}, \qquad (4)$$

$0 \leq k \leq r \leq n$, *όπου το εσωτερικό άθροισμα ορίζεται πάνω σε όλους τους μη αρνητικούς ακεραίους* $x_1, \ldots, x_{k+1}$, *οι οποίοι ικανοποιούν τους περιορισμούς* $x_1 + 2x_2 + \ldots + (k+1)x_{k+1} = n - i$, $x_1 + \ldots + x_{k+1} = n - r$.

Απόδειξη Η πιθανότητα, σε μια ακολουθία n ανεξάρτητων και ισόνομων δοκιμών *Bernoulli*, να έχουμε r επιτυχίες ισούται με:

$$P(S_n = r) = \begin{pmatrix} n \\ r \end{pmatrix} p^r q^{n-r},\ 0 \leq r \leq n.$$

Γνωρίζουμε ότι η δεσμευμένη πιθανότητα ορίζεται ως:

$$P(L_n \leq k | S_n = r) = \tfrac{P(L_n \leq k, S_n = r)}{P(S_n = r)},\ 0 \leq k \leq n.$$

Κάνοντας χρήση του παραπάνω ορισμού και του Λήμματος 1.1 προκύπτει ότι:

$$\begin{aligned} P(L_n \leq k | S_n = r) &= \frac{p^r q^{n-r} \sum_{i=0}^{k} \sum_{x_1,\ldots,x_{k+1}} \begin{pmatrix} x_1 + \ldots + x_{k+1} \\ x_1, \ldots, x_{k+1} \end{pmatrix}}{\begin{pmatrix} n \\ r \end{pmatrix} p^r q^{n-r}} \\ &= \begin{pmatrix} n \\ r \end{pmatrix}^{-1} \sum_{i=0}^{k} \sum_{x_1,\ldots,x_{k+1}} \begin{pmatrix} x_1 + \ldots + x_{k+1} \\ x_1, \ldots, x_{k+1} \end{pmatrix}. \end{aligned}$$

$\diamond$

Πράδειγμα 1.1 Για να γίνει περισσότερο κατανοητή η μέθοδος των *Philippou* και *Makri* (1986), θεωρούμε μια ακολουθία αποτελούμενη από $n = 23$ δοκιμές *Bernoulli* υποθέτοντας ότι η μέγιστη ροή επιτυχιών σ' αυτή δεν ξεπερνά την τιμή $k = 3$. Έστω ότι η ακολουθία είναι:

$$\underbrace{SF}_{e_2}\underbrace{SSF}_{e_3}\underbrace{F}_{e_1}\underbrace{F}_{e_1}\underbrace{SSF}_{e_3}\underbrace{F}_{e_1}\underbrace{SF}_{e_2}\underbrace{SSSF}_{e_4}\underbrace{F}_{e_1}\underbrace{SF}_{e_2}\underbrace{F}_{e_1}\underbrace{SS}_{i}$$

Εφαρμόζοντας την προαναφερθείσα μεθοδολογία έχουμε $x_1 = 5, x_2 = 3, x_3 = 2, x_4 = 1, i = 2$ και μπορούμε πιο εύκολα να κατανοήσουμε και τους περιορισμούς των x_i. Πιο συγκεκριμένα ισχύουν:

- $x_1 + \ldots + x_{k+1} = n - r$: κάθε ένα x_i έχει μία μόνο αποτυχία και έτσι αθροίζοντάς τα θα έχουμε το σύνολο των αποτυχιών $(= n - r)$
- $x_1 + 2x_2 + \ldots + (k+1)x_{k+1} = n - i$: ο δείκτης j στο κάθε x_j, $j = 1, \ldots, k+1$, δηλώνει τον αριθμό των δοκιμών που περιλαμβάνει το εκάστοτε e_j. Έτσι, αν τα αθροίσουμε θα πάρουμε το συνολικό αριθμό δοκιμών εκτός από τις επιτυχίες που βρίσκονται στο τέλος της ακολουθίας $(= n - i)$.

Παράδειγμα 1.2 Έστω ότι έχουμε μια ακολουθία από $n = 10$ ανεξάρτητες και ισόνομες δοκιμές *Bernoulli* και έστω ότι θέλουμε η μεταβλητή L_n να μην ξεπερνάει την τιμή $k = 4$. Θα υπολογίσουμε τη δεσμευμένη συνάρτηση κατανομής $P(L_{10} \leq 4 | S_{10} = r)$. Για να πραγματοποιηθεί αυτό θα πρέπει αρχικά να υπολογίσουμε τις επιτρεπτές τιμές των ακεραίων x_i, $i = 1, \ldots, 5$, βρίσκοντας τις κοινές λύσεις των διοφαντικών εξισώσεων:

$$x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 10 - i, \quad 0 \leq i \leq 4.$$

$$x_1 + x_2 + x_3 + x_4 + x_5 = 10 - r, \quad 0 \leq 4 \leq r \leq 10.$$

Πιο συγκεκριμένα, οι λύσεις της πρώτης είναι:

- <u>$i = 0$:</u>
  $x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 10$

0 5 0 0 0	1 1 1 1 0	0 0 0 0 2	1 0 0 1 1	0 3 0 1 0
1 3 1 0 0	0 1 1 0 1	1 2 0 0 1	0 1 0 2 0	1 0 3 0 0
0 2 2 0 0	2 2 0 1 0	0 0 2 1 0	2 4 0 0 0	2 0 1 0 1
4 3 0 0 0	2 1 2 0 0	5 1 1 0 0	2 0 0 2 0	5 0 0 0 1
3 1 0 0 1	6 0 0 1 0	3 2 1 0 0	6 2 0 0 0	3 0 1 1 0
7 0 1 0 0	4 1 0 1 0	8 1 0 0 0	4 0 2 0 0	10 0 0 0 0

- $\underline{i=1:}$
  $x_1+2x_2+3x_3+4x_4+5x_5=9$

0 0 0 1 1	1 0 1 0 1	4 1 1 0 0	0 1 1 1 0	1 0 0 2 1
4 0 0 0 1	0 2 0 0 1	2 1 0 0 1	5 2 0 0 0	0 3 1 0 0
2 2 1 1 0	5 0 0 1 0	0 0 3 0 0	2 0 1 1 0	6 0 1 0 0
1 1 2 0 0	3 1 0 1 0	7 1 0 0 0	1 2 0 1 0	3 3 0 0 0
9 0 0 0 0	1 4 0 0 0	3 0 2 0 0		

- $\underline{i=2:}$
  $x_1+2x_2+3x_3+4x_4+5x_5=8$

0 1 2 0 0	2 0 2 0 0	0 2 0 1 0	3 1 1 0 0	0 4 0 0 0	3 0 0 1 1
0 0 1 0 1	4 2 0 0 0	0 0 0 2 0	4 0 0 1 0	1 1 0 0 1	5 0 1 0 0
1 2 1 0 0	6 1 0 0 0	1 0 1 1 0	8 0 0 0 0	2 1 0 1 0	2 3 0 0 0

- $\underline{i=3:}$
  $x_1+2x_2+3x_3+4x_4+5x_5=7$

0 1 0 0 1	2 1 1 0 0	0 2 1 0 0	2 0 0 0 1	0 0 1 1 0	3 0 0 0 1
1 1 0 1 0	4 0 1 0 0	1 3 0 0 0	5 1 0 0 0	1 0 2 0 0	7 0 0 0 0

- $\underline{i=4:}$
  $x_1+2x_2+3x_3+4x_4+5x_5=6$

0 1 0 1 0	2 2 0 0 0	0 3 0 0 0	2 0 0 1 0	0 0 2 0 0
3 0 1 0 0	1 1 1 0 0	4 1 0 0 0	1 0 0 0 1	6 0 0 0 0

Δουλεύοντας με όμοιο τρόπο για τον δεύτερο περιορισμό, έχουμε την δυνατότητα να βρούμε τις κοινές λύσεις βάζοντας κάθε φορά τις επιτρεπτές τιμές στην μεταβλητή r.
Προφανώς, στις περιπτώσεις που έχουμε $r=10$ και $r=9$ επιτυχίες δεν υπάρχουν κοινές λύσεις και κατ΄ επέκταση θα ισχύει ότι $P(L_{10}\leq 4, S_{10}=10)=0$ καθώς και $P(L_{10}\leq 4, S_{10}=9)=0$, αντίστοιχα. Το αποτέλεσμα αυτό είναι αναμενόμενο αφού δεν μπορούμε να βάλουμε τα πειράματα σε κατάλληλη σειρά, ώστε το μέγιστο μήκος ροής επιτυχιών να μην ξεπερνά την τιμή 4.

Όμως, για

- $\underline{r=8:}$
  $x_1+x_2+x_3+x_4+x_5=2$

Κοινές λύσεις				
για $i=0$	για $i=1$	για $i=2$	για $i=3$	για $i=4$
0 0 0 0 2	0 0 0 1 1	0 0 1 0 1	0 1 0 0 1	0 1 0 1 0
		0 0 0 2 0	0 0 1 1 0	0 0 2 0 0
				1 0 0 0 1

Έτσι,

$$\begin{aligned}
& P(L_{10} \leq 4, S_{10} = 8) \\
= \; & p^8q^2 \left[\begin{pmatrix} 0+0+0+0+2 \\ 0,0,0,0,2 \end{pmatrix} \right] + p^8q^2 \begin{pmatrix} 0+0+0+1+1 \\ 0,0,0,1,1 \end{pmatrix} \\
& + p^8q^2 \left[\begin{pmatrix} 0+0+1+0+1 \\ 0,0,1,0,1 \end{pmatrix} + \begin{pmatrix} 0+0+0+2+0 \\ 0,0,0,2,0 \end{pmatrix} \right] \\
& + p^8q^2 \left[\begin{pmatrix} 0+1+0+0+1 \\ 0,1,0,0,1 \end{pmatrix} + \begin{pmatrix} 0+0+1+1+0 \\ 0,0,1,1,0 \end{pmatrix} \right] \\
& + p^8q^2 \left[\begin{pmatrix} 0+1+0+1+0 \\ 0,1,0,1,0 \end{pmatrix} + \begin{pmatrix} 0+0+2+0+0 \\ 0,0,2,0,0 \end{pmatrix} + \begin{pmatrix} 1+0+0+0+1 \\ 1,0,0,0,1 \end{pmatrix} \right] \\
= \; & p^8q^2[1+2+2+1+2+2+2+1+2] \\
= \; & p^8q^2 \cdot 15 = 0.0146484375, \quad \text{για } p = q = 0.5
\end{aligned}$$

- $\underline{r = 7:}$
  $x_1 + x_2 + x_3 + x_4 + x_5 = 3$

  και θα έχουμε:

Κοινές λύσεις				
για $i=0$	για $i=1$	για $i=2$	για $i=3$	για $i=4$
0 1 1 0 1	0 1 1 1 0	0 1 2 0 0	0 2 1 0 0	0 3 0 0 0
0 0 2 1 0	0 2 0 0 1	0 2 0 1 0	1 1 0 1 0	1 1 1 0 0
1 0 0 1 1	0 0 3 0 0	1 1 0 0 1	1 0 2 0 0	2 0 0 1 0
0 1 0 2 0	1 0 1 0 1	1 0 1 1 0	2 0 0 0 1	
	1 0 0 2 0			

Έτσι,

$$
\begin{aligned}
&P(L_{10} \leq 4, S_{10} = 7) \\
= \; & p^7q^3\left[\binom{0+1+1+0+1}{0,1,1,0,1} + \binom{0+0+2+1+0}{0,0,2,1,0} + \binom{1+0+0+1+1}{1,0,0,1,1}\right. \\
&\left. + \binom{0+1+0+2+0}{0,1,0,2,0}\right] + p^7q^3\left[\binom{0+1+1+1+0}{0,1,1,1,0} + \binom{0+2+0+0+1}{0,2,0,0,1}\right. \\
&\left. + \binom{0+0+3+0+0}{0,0,3,0,0} + \binom{1+0+1+0+1}{1,0,1,0,1} + \binom{1+0+0+2+0}{1,0,0,2,0}\right] \\
&+p^7q^3\left[\binom{0+1+2+0+0}{0,1,2,0,0} + \binom{0+2+0+1+0}{0,2,0,1,0} + \binom{1+1+0+0+1}{1,1,0,0,1}\right. \\
&\left. + \binom{1+0+1+1+0}{1,0,1,1,0}\right] + p^7q^3\left[\binom{0+2+1+0+0}{0,2,1,0,0} + \binom{1+1+0+1+0}{1,1,0,1,0}\right. \\
&\left. + \binom{1+0+2+0+0}{1,0,2,0,0} + \binom{2+0+0+0+1}{2,0,0,0,1}\right] \\
&+p^7q^3\left[\binom{0+3+0+0+0}{0,3,0,0,0} + \binom{1+1+1+0+0}{1,1,1,0,0} + \binom{2+0+0+1+0}{2,0,0,1,0}\right] \\
= \; & 0.5^{10} \cdot 80 = 0.078125, \quad \text{για } p = q = 0.5
\end{aligned}
$$

Τέλος, θα υπολογίσουμε τις αντίστοιχες δεσμευμένες συναρτήσεις κατανομής, χρησιμοποιώντας τον τύπο του Θεωρήματος 1.1. Έτσι,

$$
\begin{aligned}
P(L_{10} \leq 4 | S_{10} = 8) &= \binom{10}{8}^{-1} \sum_{i=0}^{4} \sum_{x_1,\ldots,x_5} \binom{x_1+x_2+x_3+x_4+x_5}{x_1,x_2,x_3,x_4,x_5} \\
&= \frac{15}{45} = 0.333333
\end{aligned}
$$

και

$$
\begin{aligned}
P(L_{10} \leq 4 | S_{10} = 7) &= \binom{10}{7}^{-1} \sum_{i=0}^{4} \sum_{x_1,\ldots,x_5} \binom{x_1+x_2+x_3+x_4+x_5}{x_1,x_2,x_3,x_4,x_5} \\
&= \frac{80}{120} = 0.66667
\end{aligned}
$$

1.2. Αθροιστική συνάρτηση κατανομής της μεταβλητής L_n

Για την εύρεση, με ανάλογη μεθοδολογία, της αθροιστικής συνάρτησης κατανομής της τυχαίας μεταβλητής L_n, θα βασιστούμε στην τ.μ. $N_{n,k}$ η οποία μας δίνει τον αριθμό των ροών επιτυχιών μήκους k σε μια ακολουθία n δοκιμών *Bernoulli*. Η χρήση της $N_{n,k}$ είναι σημαντική διότι μπορούμε να αποκτήσουμε έμμεσα το ζητούμενο, υπολογίζοντας μια τιμή της συνάρτησης πιθανότητάς της.

Πιο συγκεκριμένα, αν έχουμε μια ακολουθία από n πειράματα στα οποία η μέγιστη ροή επιτυχιών δεν ξεπερνά σε μήκος την τιμή $k-1$ ($L_n \leq k-1$) τότε δεν δύναται να υπάρξει ροή επιτυχιών μήκους k ($N_{n,k} = 0$).

Επομένως ισχύει η ισότητα,

$$P(L_n \leq k-1) = P(N_{n,k} = 0).$$

Οι *Philippou* και *Makri* (1986) καθώς και ο *Hirano* (1986) ανεξάρτητα, έδωσαν ακριβή τύπο για την κατανομή της μεταβλητής $N_{n,k}$. Εφαρμόζοντας ανάλογη μεθοδολογία, όπως παρουσιάστηκε στην προηγούμενη παράγραφο, έχουμε το παρακάτω θεώρημα.

Θεώρημα 1.2 *Αν συμβολίσουμε με $N_{n,k}$ τον αριθμό των ροών επιτυχιών μήκους k ($k \geq 1$) σε n ($n \geq 1$) ανεξάρτητες δοκιμές με κοινή πιθανότητα επιτυχίας p ($0 < p < 1$), τότε,*

$$P(N_{n,k} = x) = \sum_{i=0}^{k-1} \sum_{x_1,\ldots,x_k} \left(\begin{array}{c} x_1 + \ldots + x_k + x \\ x_1, \ldots, x_k, x \end{array} \right) p^n \left(\frac{q}{p}\right)^{x_1+\ldots+x_k},$$

$$x = 0, \ldots, \lfloor n/k \rfloor,$$

όπου το εσωτερικό άθροισμα ορίζεται πάνω σε όλους τους μη αρνητικούς ακεραίους $x_1, \ldots, x_k$ με $x_1 + 2x_2 + \ldots + kx_k = n - kx - i$ και $q = 1 - p$.

Απόδειξη Ακολουθώντας τους *Philippou* και *Muwafi* (1982), παρατηρούμε ότι ένα τυπικό στοιχείο του γεγονότος ($N_{n,k} = x$) είναι μια διάταξη της μορφής:

$$a_1 a_2 \ldots a_{x_1+\ldots+x_k+x} \underbrace{SS\ldots S}_{i}, \ (0 \leq i \leq k-1)$$

όπου x_1 από τα a είναι $e_1 = F$, x_2 από τα a είναι $e_2 = SF, \ldots, x_k$ είναι $e_k = \underbrace{SS\ldots S}_{k-1}F$, x είναι $\widetilde{e}_k = \underbrace{SS\ldots S}_{k}$ και ικανοποιείται ο περιορισμός $x_1 + 2x_2 + \ldots + kx_k + kx + i = n$. Αν i και $x_1, \ldots, x_k$ είναι σταθερά τότε ο αριθμός αυτών των διατάξεων είναι:

$$\left(\begin{array}{c} x_1 + \ldots + x_k + x \\ x_1, \ldots, x_k, x \end{array} \right).$$

Επιπλέον, λαμβάνοντας υπόψη το γεγονός ότι οι δοκιμές είναι ανεξάρτητες μεταξύ τους και ότι έχουν την ίδια πιθανότητα επιτυχίας $P(S) = p$, μπορούμε να βρούμε την πιθανότητα της κάθε διάταξης ως εξής:

$$\begin{aligned}
& P(a_1 a_2 \ldots a_{x_1+\ldots+x_{k+1}+x} \underbrace{SS\ldots S}_{i}) \\
= \;& (P\{e_1\})^{x_1}(P\{e_2\})^{x_2}\cdots(P\{e_k\})^{x_k}(P\{\widetilde{e}_k\})^{x} P\{\underbrace{SS\ldots S}_{i}\} \\
= \;& q^{x_1+x_2+\ldots+x_k} p^{x_2+2x_3+\ldots+(k-1)x_k+kx+i} \\
= \;& q^{x_1+x_2+\ldots+x_k} p^{n-(x_1+\ldots+x_k)} \\
= \;& p^n \left(\frac{q}{p}\right)^{x_1+x_2+\ldots+x_k}.
\end{aligned}$$

Οι μη αρνητικοί ακέραιοι $x_1, x_2, \ldots, x_k$ ικανοποιούν τη σχέση $x_1 + 2x_2 + \ldots + kx_k = n - i - kx$ και $0 \le i \le k-1$.
Έτσι, αθροίζοντας πάνω στις δυνατές τιμές των ακεραίων $x_1, \ldots, x_k$ και του i προκύπτει ότι:

$$P(N_{n,k} = x) = \sum_{i=0}^{k-1} \sum_{x_1,\ldots,x_k} \left(\begin{array}{c} x_1 + \ldots + x_k + x \\ x_1, \ldots, x_k, x \end{array} \right) p^n \left(\frac{q}{p}\right)^{x_1+\ldots+x_k}, \qquad x = 0, \ldots, \lfloor n/k \rfloor,$$

όπου το εσωτερικό άθροισμα λαμβάνεται πάνω στους μη αρνητικούς ακεραίους για τους οποίους ισχύει $x_1 + 2x_2 + \ldots + kx_k = n - i - kx$.

$\diamond$

Παρατηρήσεις

1. Η κατανομή της μεταβλητής $N_{n,k}$ θα συμβολίζεται $B_k(n,p)$ και θα ονομάζεται διωνυμική κατανομή τάξεως k με διάνυσμα παραμέτρων (n,p).

2. Στην περίπτωση όπου $k = 1$ παίρνουμε την γνωστή διωνυμική κατανομή $B_1(n,p) = B(n,p)$.

Πράγματι,
για $k = 1$ έχουμε ότι $i = 0$ και ότι το $\{N_{n,1} = x\}$ παριστάνει το γεγονός να

έχουμε x επιτυχίες σε n δοκιμές *Bernoulli*. Επομένως,

$$\begin{aligned} P(N_{n,1}=x) &= \sum_{i=0}^{0}\sum_{x_1}\binom{x_1+x}{x_1,x}p^{n-x_1}\left(\frac{q}{p}\right)^{x_1} \\ \text{όπου } x_1=n-x &= \sum_{x_1}\binom{n-x+x}{x_1,x}p^{n-x_1}q^{x_1} \\ &= \binom{n}{x}p^x q^{n-x}. \end{aligned}$$

Στη συνέχεια ο *Godbole* (1990), στηριζόμενος στα αποτελέσματα των *Philippou* και *Makri* (1986), έδωσε έναν εναλλακτικό τύπο για την κατανομή της μεταβλητής $N_{n,k}$ ο οποίος:

(α) Δίνεται μέσω διωνυμικών συντελεστών.

(β) Δεν περιλαμβάνει άθροισμα πάνω σε σύνολο ακεραίων που είναι λύσεις διοφαντικών εξισώσεων.

(γ) Είναι περισσότερο επιδεκτική στον υπολογισμό οριακών κατανομών.

Πριν προχωρήσουμε στο θεώρημα και την απόδειξη του *Godbole* (1990), θα παραθέσουμε ένα λήμμα που μας δίνει το άνω και το κάτω φράγμα του αριθμού των αποτυχιών της εκάστοτε ακολουθίας, υπό τον περιορισμό ότι $N_{n,k}=x$.

Λήμμα 1.2 *Αν έχουμε μια ακολουθία με n δοκιμές Bernoulli και Y είναι ο αριθμός των αποτυχιών, τότε ισχύει η σχέση $\lfloor (n-kx)/k \rfloor \le Y \le n-kx$. Πιο συγκεκριμένα έχουμε $P(N_{n,k}=x, Y=y)>0$ αν και μόνο αν $\lfloor (n-kx)/k \rfloor \le y \le n-kx$.*

Στο θεώρημα που ακολουθεί δίνεται ο τύπος του *Godbole* (1990).

Θεώρημα 1.3 *Θεωρούμε n δοκιμές Bernoulli με πιθανότητα επιτυχίας p, και έστω ότι η $N_{n,k}$ παριστάνει τον αριθμό των ροών επιτυχιών μήκους k. Τότε,*

$$\begin{aligned} P(N_{n,k}=x) = & \sum_{\lfloor (n-kx)/k \rfloor \le y \le n-kx} q^y p^{n-y}\binom{y+x}{x} \\ & \times \sum_{0\le j\le \lfloor (n-kx-y)/k \rfloor} (-1)^j \binom{y+1}{j}\binom{n-kx-jk}{y}, \\ & \qquad x=0,\ldots,\lfloor n/k \rfloor. \end{aligned}$$

Για να μπορέσουμε να αποδείξουμε το θεώρημα πρέπει αρχικά να δούμε τα δύο παρακάτω λήμματα.

Λήμμα 1.3 *Έστω S_n παριστάνει τον αριθμό των επιτυχιών, τα $N_{n,k}$, L_n όπως έχουν ορισθεί παραπάνω και έστω X_j είναι το αποτέλεσμα της j δοκιμής. Τότε,*

$$
\begin{aligned}
& P(N_{n,k}=x) \\
= \ & p^{kx} \sum_{\lfloor (n-kx)/k \rfloor \le y \le n-kx} \binom{y+x}{x} \sum_{0\le i \le k-1} p^i P(L_{m_i} \le k-1, X_{m_i}=F, S_{m_i}=m_i-y),
\end{aligned}
$$

όπου $m_i = n-kx-i$ και $x=0,1,\ldots,\lfloor n/k \rfloor$.

Απόδειξη Σύμφωνα με το Λήμμα 1.2 και το Λήμμα 1.3 έχουμε:

$$
\begin{aligned}
P(N_{n,k}=x) &= \sum_y P(N_{n,k}=x, S_n=n-y) \\
&= \sum_y \sum_i \textstyle\sum_3 \begin{pmatrix} x_1+\ldots+x_k+x \\ x_1,\ldots,x_k,x \end{pmatrix} p^n \left(\dfrac{q}{p}\right)^{x_1+\ldots+x_k},
\end{aligned}
$$

όπου το άθροισμα $\sum_3$ ορίζεται πάνω σε όλους τους μη αρνητικούς ακεραίους x_j οι οποίοι ικανοποιούν τις (α) $\sum_{1\le j\le k} x_j = y$ και (β) $\sum_{1\le j \le k} jx_j = n-kx-i (= m_i)$. Γι αυτό,

$$
\begin{aligned}
P(N_{n,k}=x) &= \sum_y \sum_i \textstyle\sum_3 \begin{pmatrix} y+x \\ x_1,\ldots,x_k,x \end{pmatrix} p^{n-y} q^y \\
&= p^{kx} \sum_y \begin{pmatrix} y+x \\ x \end{pmatrix} \sum_i p^i p^{n-kx-i-y} q^y \textstyle\sum_3 \begin{pmatrix} x_1+\ldots+x_k \\ x_1,\ldots,x_k \end{pmatrix}.
\end{aligned}
$$

Με μια πιο προσεκτική ματιά στην απόδειξη για την κατανομή της $N_{n,k}$ από τους *Philippou* και *Makri* (1986), διαπιστώνουμε ότι η ποσότητα

$$
p^{m_i-y} q^y \textstyle\sum_3 \begin{pmatrix} x_1+\ldots+x_k \\ x_1,\ldots,x_k \end{pmatrix}
$$

ισούται με την πιθανότητα να έχουμε α)$L_{m_i} \le k-1$ β) $S_{m_i} = m_i - y$ και έτσι δεν επιτρέπεται καμία ροή $SS\ldots S$ στο τέλος της ακολουθίας, γ) $X_{m_i}=F$.

◇

Λήμμα 1.4 *Για S_n, L_n, και X_j ορισμένες όπως στο Λήμμα 1.3 και θεωρώντας ως $A_{m,k,y}$ το γεγονός $\{L_m \le k-1, X_m = F, S_m = m-y\}$, τότε έχουμε για* *κάθε* m,

$$P(A_{m,k,y}) = q^y p^{m-y} \sum_{0 \le j \le \lfloor (m-y)/k \rfloor} (-1)^j \binom{y}{j} \binom{m-jk-1}{y-1}.$$

Απόδειξη Θεωρούμε ότι κάθε στοιχείο του παραπάνω γεγονότος είναι μια ακολουθία $|SS\ldots SSF|SS\ldots SSF|\ldots|SS\ldots SSF|$, η οποία περιλαμβάνει α_j επιτυχίες $(0 \le a_j \le k-1)$ πριν από την j αποτυχία $(1 \le j \le y)$. Έτσι,

$$\begin{aligned} P(A_{m,k,y}) &= P(S_m = m-y, X_m = F) - P(S_m = m-y, X_m = F, L_m \ge k) \\ &= \binom{m-1}{y-1} q^y p^{m-y} - P(\bigcup_{1 \le j \le y} A_j), \end{aligned}$$

όπου $A_j = \{S_m = m-y, X_m = F, a_j \ge k\}$ και a_j όπως ορίστηκαν παραπάνω. Αυτό ουσιαστικά σημαίνει ότι για το γεγονός A_j θέλουμε να ισχύει:

- $S_m = m-y$: ο αριθμός των επιτυχιών σε m δοκιμές *Bernoulli* να είναι $m-y$.
- $X_m = F$: το αποτέλεσμα της δοκιμής m να είναι αποτυχία.
- $a_j \ge k$: ο αριθμός των επιτυχιών πριν από κάθε αποτυχία να είναι μεγαλύτερος ή ίσος του k.

Θεωρούμε ότι έχουμε y κάλπες, μία για κάθε F. Δηλαδή ο αριθμός των καλπών ισούται με τον αριθμό των αποτυχιών στην ακολουθία. Έτσι, καλούμαστε να τοποθετήσουμε τις επιτυχίες με όλους τους δυνατούς τρόπους. Για να μπορέσουμε να ικανοποιήσουμε τον περιορισμό $\alpha_j \ge k$, θα πρέπει να διανείμουμε τις $m-y$ όμοιες σφαίρες (δηλ. τις επιτυχίες), έτσι ώστε οι j διακεκριμένες κάλπες $(i_1, i_2, \ldots, i_j)$, να περιέχουν τουλάχιστον k σφαίρες (επιτυχίες) η καθεμία.

Εφόσον ικανοποιήσαμε τον περιορισμό βάζοντας τις λιγότερες σφαίρες που μπορεί να έχει κάθε μία κάλπη (k σε αριθμό), οι εναπομείνασες $m-y-jk$ πρέπει να τοποθετηθούν στις y κάλπες με οποιονδήποτε τρόπο. Αυτό επιτυγχάνεται με

$$\binom{m-y-jk+y-1}{y-1} = \binom{m-jk-1}{y-1} \text{ διαφορετικούς τρόπους.}$$

Σημειώνουμε ότι ο αριθμός αυτός είναι ίδιος για κάθε σύνολο με πληθικότητα j και χρησιμοποιώντας την αρχή «εγκλεισμού-αποκλεισμού»

$$P(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P(A_j) - \sum_{1 \le i < j \le n} (A_i \cap A_j) + \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_{n-1})$$

προκύπτει ότι,

$$\begin{aligned} P(A_{m,k,y}) &= q^y p^{m-y} \left[\binom{m-1}{y-1} - \sum_{1\leq j\leq y} (-1)^{j+1} \binom{y}{j} \binom{m-jk-1}{y-1} \right] \\ &= q^y p^{m-y} \sum_{0\leq j\leq y} (-1)^j \binom{y}{j} \binom{m-jk-1}{y-1}. \end{aligned}$$

Όπου $\binom{y}{j}$ είναι όλοι οι δυνατοί τρόποι για να τοποθετήσουμε τις αποτυχίες και $\binom{m-jk-1}{y-1}$ είναι οι τρόποι για να τοποθετήσουμε τις επιτυχίες που απέμειναν. Έτσι καταλήξαμε στο αποτέλεσμα που θέλαμε.

$\diamond$

Απόδειξη Θεωρήματος 1.3
Από τα Λήμματα 1.3, 1.4 και θέτοντας όπου $m_i = n - kx - i$ έχουμε:

$$\begin{aligned} &P(N_{n,k}) \\ &= p^{kx} \sum_y \binom{y+x}{x} \sum_{0\leq i\leq k-1} p^i q^y p^{m_i-y} \sum_{0\leq j\leq \lfloor (m_i-y)/k \rfloor} (-1)^j \binom{y}{j} \binom{m_i-jk-i}{y-1} \\ &= \sum_y q^y p^{n-y} \binom{y+x}{x} \sum_{0\leq j\leq \lfloor n-kx-y)/k \rfloor} (-1)^j \binom{y}{j} \sum_{0\leq i\leq k-1} \binom{m_i-jk-i}{y-1} \\ &= \sum_y q^y p^{n-y} \binom{y+x}{x} \sum_j (-1)^j \binom{y}{j} \left[\binom{n-kx-jk}{y} - \binom{n-kx-jk-k}{y} \right]. \end{aligned}$$

Ξρησιμοποιώντας την ιδιότητα του τριγώνου του *Pascal* έχουμε

$$P(N_{n,k} = x) = \sum_y q^y p^{n-y} \binom{y+x}{x} \sum_j (-1)^j \binom{y+1}{j} \binom{n-kx-jk}{y}.$$

$\diamond$

Επειδή όμως ο σκοπός είναι η εύρεση της αθροιστικής συνάρτησης κατανομής της L_n, έχουμε το πόρισμα που έχει δειχθεί και από τους *Burr* και *Cane* (1961).

Πόρισμα *Έστω L_n είναι το μέγιστο μήκος ροής επιτυχιών σε n δοκιμές Bernoulli. Τότε,*

$$P(L_n \leq k-1) = \sum_{\lfloor n/k \rfloor \leq y\leq n} q^y p^{n-y} \sum_{0\leq j\leq \lfloor n-y)/k \rfloor} (-1)^j \binom{y+1}{j} \binom{n-jk}{y}.$$

Απόδειξη Θέτοντας $x = 0$ στο Θεώρημα 1.3 παρατηρούμε ότι $(N_{n,k} = 0) = (L_n \leq k-1)$. Αυτό συμβαίνει γιατί αν το μήκος της μέγιστης ροής επιτυχιών δεν ξεπερνά το $k-1$ τότε δεν υπάρχει καμία ροή επιτυχιών με μήκος ίσο με k και αντιστρόφως. Έτσι προκύπτει το ζητούμενο.

$\diamond$

Παράδειγμα 1.3 Έστω ότι έχουμε μια ακολουθία από $n = 10$ ανεξάρτητες και ισόνομες δοκιμές *Bernoulli* με $k = 4$. Χρησιμοποιώντας τον τύπο του *Godbole* (1990) θα υπολογίσουμε την κατανομή $P(L_{10} \leq 3)$.

Αρχικά θα πρέπει να βρούμε τις δυνατές τιμές που μπορεί να πάρει η μεταβλητή y. Για $n = 10$ και $k = 4$ θα έχουμε ότι:

$$\left\lfloor \frac{n}{k} \right\rfloor \leq y \leq n \Longrightarrow \left\lfloor \frac{10}{4} \right\rfloor \leq y \leq 10 \Longrightarrow \lfloor 2.5 \rfloor \leq y \leq 10 \Longrightarrow 2 \leq y \leq 10.$$

Έτσι, θα υπολογίσουμε την πιθανότητα

$$P(L_{10} \leq 3) = \sum_{y=2}^{10} q^y p^{10-y} \sum_{0 \leq j \leq \lfloor 10-y)/4 \rfloor} (-1)^j \binom{y+1}{j} \binom{10-4j}{y}.$$

- <u>Για y=2</u> έχουμε: $0 \leq j \leq \lfloor 8/4 \rfloor \longrightarrow 0 \leq j \leq 2$

Έτσι

$$q^2p^8 \left[\binom{3}{0}\binom{10}{2} - \binom{3}{1}\binom{6}{2} + \binom{3}{2}\binom{2}{2} \right] = q^2p^8[45-45+3] = 3q^2p^8$$

- <u>Για y=3</u> έχουμε: $0 \leq j \leq \lfloor 7/4 \rfloor \longrightarrow 0 \leq j \leq 1$

Έτσι

$$q^3p^7 \left[\binom{4}{0}\binom{10}{3} - \binom{4}{1}\binom{6}{3} \right] = q^3p^7[120-80] = 40q^3p^7$$

- <u>Για y=4</u> έχουμε: $0 \leq j \leq \lfloor 6/4 \rfloor \longrightarrow 0 \leq j \leq 1$

Έτσι

$$q^4p^6 \left[\binom{5}{0}\binom{10}{4} - \binom{5}{1}\binom{6}{4} \right] = q^4p^6[210-75] = 135q^4p^6$$

- Για y=5 έχουμε: $0 \leq j \leq \lfloor 5/4 \rfloor \longrightarrow 0 \leq j \leq 1$

Έτσι

$$q^5p^5\left[\binom{6}{0}\binom{10}{5}-\binom{6}{1}\binom{6}{5}\right]=q^5p^5[252-36]=216q^5p^5$$

- Για y=6 έχουμε: $0 \leq j \leq \lfloor 4/4 \rfloor \longrightarrow 0 \leq j \leq 1$

Έτσι

$$q^6p^4\left[\binom{7}{0}\binom{10}{6}-\binom{7}{1}\binom{6}{6}\right]=q^6p^4[210-7]=203q^6p^4$$

- Για y=7 έχουμε: $j=0$

Έτσι

$$q^7p^3\binom{8}{0}\binom{10}{7}=120q^7p^3$$

- Για y=8 έχουμε: $j=0$

Έτσι

$$q^8p^2\binom{9}{0}\binom{10}{8}=45q^8p^2$$

- Για y=9 έχουμε: $j=0$

Έτσι

$$q^9p\binom{10}{0}\binom{10}{9}=10q^9p$$

- Για y=10 έχουμε: $j=0$

Έτσι

$$q^{10}\left(\begin{array}{c}11\\0\end{array}\right)\left(\begin{array}{c}10\\10\end{array}\right)=q^{10}$$

Επομένως για μια συγκεκριμένη τιμή $p = q = 0.5$ έχουμε:

$P(L_{10} \leq 3) = (0.5)^{10}[3{+}40{+}135{+}216{+}203{+}120{+}45{+}10{+}1] = 0.754882812.$

Βέβαια, θα καταλήξουμε στο ίδιο αποτέλεσμα αν χρησιμοποιήσουμε τον τύπο των *Philippou* και *Makri* (1986) για τη συνάρτηση πιθανότητας της τ.μ. $N_{k,n}$.

Παράδειγμα 1.4 Έστω ότι έχουμε μια ακολουθία από $n = 10$ ανεξάρτητες και ισόνομες δοκιμές *Bernoulli* και έστω $k = 4$. Υπολογίζοντας την πιθανότητα

$$P(N_{10,4}=0)=\sum_{i=0}^{3}\sum_{x_1,\ldots,x_4}\left(\begin{array}{c}x_1+\ldots+x_4\\x_1,\ldots,x_4\end{array}\right)p^{10}\left(\frac{q}{p}\right)^{x_1+\ldots+x_4},$$

υπολογίζουμε ουσιαστικά την $P(L_{10} \leq 3)$. Αρχικά θα πρέπει να επιλύσουμε τη διοφαντική εξίσωση

$$x_1+2x_2+3x_3+4x_4=10-i,\quad 0\leq i\leq 3$$

που αποτελεί τον περιορισμό για τους μη αρνητικούς ακεραίους x_j , $j = 0, 1, 2, 3, 4$ Πιο συγκεκριμένα έχουμε,

- $\underline{i = 0:}$
  $x_1 + 2x_2 + 3x_3 + 4x_4 = 10$

0 0 2 1	0 1 0 2	0 2 2 0	0 3 0 1
0 5 0 0	1 0 3 0	1 1 1 1	1 3 1 0
2 0 0 2	2 1 2 0	2 2 0 1	2 4 0 0
3 0 1 1	3 2 1 0	4 0 2 0	4 1 0 1
5 1 1 0	6 0 0 2	6 2 0 0	7 0 1 0
8 1 0 0	10 0 0 0	4 3 0 0	

- $\underline{i = 1:}$
  $x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 9$

0 0 3 0	0 1 1 1	0 3 1 0
1 0 0 2	1 1 2 0	1 2 0 1
1 4 0 0	2 0 1 1	2 2 1 0
3 0 2 0	3 1 0 1	3 3 0 0
4 1 1 0	5 0 0 1	5 2 0 0
6 0 1 0	7 1 0 0	9 0 0 0

- $\underline{i = 2:}$
  $x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 8$

0 0 0 2	0 1 2 0	0 2 0 1
0 4 0 0	1 0 1 1	1 2 1 0
2 3 0 0	2 0 2 0	2 1 0 1
3 1 1 0	4 2 0 0	4 0 0 1
5 0 1 0	6 1 0 0	8 0 0 0

- $\underline{i = 3:}$
  $x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 7$

0 0 1 0	0 2 1 0	1 1 0 1
1 3 0 0	1 0 2 0	2 1 1 0
3 2 0 0	3 0 0 1	4 0 1 0
5 1 0 0	7 0 0 0	

Έτσι για $p = q = 0.5$,

$$\begin{aligned}
P(N_{10,4} = 0) &= p^{10} \sum_{x_1+2x_2+3x_3+4x_4=10} \binom{x_1+x_2+x_3+x_4+0}{x_1, x_2, x_3, x_4, 0} \left(\frac{q}{p}\right)^{x_1+x_2+x_3+x_4} \\
&\quad + p^{10} \sum_{x_1+2x_2+3x_3+4x_4=9} \binom{x_1+x_2+x_3+x_4+0}{x_1, x_2, x_3, x_4, 0} \left(\frac{q}{p}\right)^{x_1+x_2+x_3+x_4} \\
&\quad + p^{10} \sum_{x_1+2x_2+3x_3+4x_4=8} \binom{x_1+x_2+x_3+x_4+0}{x_1, x_2, x_3, x_4, 0} \left(\frac{q}{p}\right)^{x_1+x_2+x_3+x_4} \\
&\quad + p^{10} \sum_{x_1+2x_2+3x_3+4x_4=7} \binom{x_1+x_2+x_3+x_4+0}{x_1, x_2, x_3, x_4, 0} \left(\frac{q}{p}\right)^{x_1+x_2+x_3+x_4} \\
&= p^{10} \cdot 401 + p^{10} \cdot 208 + p^{10} \cdot 108 + p^{10} \cdot 56 \\
&= (0.5)^{10} \cdot 773 = 0.7548828
\end{aligned}$$

Στη συνέχεια ο *Muselli* (1996) ακολουθώντας μια διαφορετική μέθοδο από αυτήν του *Godbole* (1990), παρουσίασε έναν ακόμα πιο απλό τύπο για τον υπολογισμό της πιθανότητας $P(L_n \leq k-1)$. Χρησιμοποίησε, εκτός από την τυχαία μεταβλητή S_n, την $G_{n,k}$ η οποία παριστάνει τον αριθμό των ροών επιτυχιών μήκους τουλάχιστον k σε μια ακολουθία n δοκιμών *Bernoulli*, όπου η καθεμία έχει πιθανότητα επιτυχίας p, $0 \leq p \leq 1$. Έτσι, στηριζόμενος στην ισότητα $P(L_n \geq k) = 1 - P(L_n \leq k-1) = P(G_{n,k} \geq 1) = \sum_{x=1}^{\lfloor \frac{n+1}{k+1} \rfloor} P(G_{n,k} = x)$, έφτασε στο επιθυμητό αποτέλεσμα. Στο επόμενο θεώρημα παρουσιάζεται μια πρώτη έκφραση για τη συνάρτηση πιθανότητας της τυχαίας μεταβλητής $G_{n,k}$.

Θεώρημα 1.4 *Αν με $G_{n,k}$ συμβολίσουμε τον αριθμό των ροών επιτυχιών μήκους k ή και παραπάνω, σε n δοκιμές Bernoulli, θα έχουμε:*

$$P(G_{n,k} = x) = \sum_{m=x}^{\lfloor \frac{n+1}{k+1} \rfloor} (-1)^{m-x} \binom{m}{x} \sum_{y=m-1}^{n-mk} \binom{y+1}{m} \binom{n-mk}{y} p^{n-y} q^y, \quad (5)$$

όπου k, n και x είναι θετικοί ακέραιοι.

Απόδειξη Θεωρούμε το γεγονός

A_j=(Μια ακολουθία αποτελούμενη από k συνεχόμενες επιτυχίες και η οποία ξεκινάει από το X_j),

όπου X_j είναι το αποτέλεσμα της $j-$οστής δοκιμής. Αν επίσης ορίσουμε ένα υποσύνολο του $\{1, \ldots, n\}$, το $J_x = (j_1, j_2, \ldots, j_x)$ (όπου j_i είναι διατεταγμένα κατά αύξουσα σειρά, δηλαδή $j_1 \leq \ldots \leq j_x$) το οποίο περιλαμβάνει ακριβώς x διαφορετικούς δείκτες και αν συμβολίσουμε με $\overline{A}_j$ το συμπλήρωμα του A_j, τότε μπορούμε να γράψουμε ότι:

$$P(G_{n,k} = x) = P\left(\bigcup_{j_1, j_2, \ldots, j_x} \left(A_{j_1} \cap A_{j_2} \cap \ldots \cap A_{j_x} \cap \bigcap_{j \notin J_x} \overline{A}_j \right) \right)$$

και χρησιμοποιώντας το θεώρημα «εγκλεισμού-αποκλεισμού» θα έχουμε:

$$P(G_{n,k} = x) = \sum_{m \geq x} (-1)^{m-x} \binom{m}{x} r_m,$$

όπου,

$$r_m = \sum_{j_1, \ldots, j_m} P(A_{j_1} \cap \ldots \cap A_{j_m}) = \sum_{y=m-1}^{n-mk} \sum_{j_1, \ldots, j_m} P(A_{j_1} \cap \ldots \cap A_{j_m}, S_n = n-y). \quad (6)$$

Αν έχουμε m ροές επιτυχιών με μήκος τουλάχιστον k οι οποίες έχουν ως αρχή τις θέσεις $j_1, \dots, j_m$ και αν συμβολίσουμε με y το συνολικό αριθμό των αποτυχιών, τότε συμπεραίνουμε ότι χρειαζόμαστε το λιγότερο $m-1$ αποτυχίες οι οποίες θα χωρίζουν τις ροές επιτυχιών. Από την άλλη πλευρά, αφού έχουμε m ροές επιτυχιών μήκους τουλάχιστον k, τότε αυτό σημαίνει ότι έχουμε τουλάχιστον mk επιτυχίες στην ακολουθία και αυτό συνεπάγεται ότι ο μέγιστος αριθμός αποτυχιών είναι ίσος με $n - mk$. Έτσι, προκύπτουν τα όρια για τον αριθμό των αποτυχιών στο άθροισμα.
Από τον ορισμό του γεγονότος A_j προκύπτει ότι, αν κάποια ακολουθία ξεκινάει από την θέση j τότε αυτόματα θα έχουμε ότι η δοκιμή $X_{j-1} = F$. Αυτό θα έχει σαν αποτέλεσμα η πιθανότητα $P(A_{j_1} \cap \ldots \cap A_{j_m} \cap S_n = n - y)$ να ειναι διαφορετική από το μηδέν όταν οι ροές επιτυχιών του συνόλου A_j είναι διακεκριμένες, δηλαδή όταν ισχύει:

$$j_1 + k + 1 \leq j_2, \dots, j_{m-1} + k + 1 \leq j_m,\ j_m + k - 1 \leq n.$$

Για $j_1 \geq 1$, συνδυάζοντας τις ανισότητες θα έχουμε ότι $r_m = 0$ για

$$1 + (k+1)(m-1) + k - 1 > n \Rightarrow m > \tfrac{n+1}{k+1} \Rightarrow m > \lfloor \tfrac{n+1}{k+1} \rfloor$$

και έτσι παίρνουμε ένα άνω φράγμα για την ποσότητα m. Όπου $\lfloor x \rfloor$ ο μεγαλύτερος ακέραιος που είναι μικρότερος από τον x.

Σε αντίθετη περίπτωση, έχουμε ότι

$$\sum_{j_1,\dots,j_m} P(A_{j_1} \cap \ldots \cap A_{j_m} \cap S_n = n - y) = N_{m,y} p^{n-y} q^y, \tag{7}$$

όπου η ποσότητα $N_{m,y}$ μας δίνει τον αριθμό των διαφορετικών τρόπων με τους οποίους μπορούμε να τοποθετήσουμε τα πειράματά μας έτσι ώστε να έχουμε ακριβώς $n - y$ επιτυχίες και κάθε φορά να υπάρχουν m ροές επιτυχιών μήκους τουλάχιστον k στην ακολουθία.

Για να μπορέσουμε να δώσουμε μια έκφραση για την ποσότητα $N_{m,y}$, λαμβάνουμε υπ' όψιν μας ότι:

- Εφόσον έχουμε y αποτυχίες, τότε μπορούμε να τοποθετήσουμε τις m ροές επιτυχιών μήκους k, έτσι ώστε η καθεμία να διαχωρίζεται από την άλλη το λιγότερο από μία αποτυχία με

  $\left(\begin{array}{c} y+1 \\ m \end{array} \right)$ διαφορετικούς τρόπους.

- Στη συνέχεια παρατηρούμε ότι μας απομένουν να τοποθετήσουμε $n-y-mk$ επιτυχίες και επειδή δεν υπάρχει πλέον κανένας περιορισμός, αυτό επιτυγχάνεται με

$$\binom{n-y-mk+y+1-1}{y+1-1} = \binom{n-mk}{y} \text{ διαφορετικούς τρόπους.}$$

Έτσι, προκύπτει ότι

$$N_{m,y} = \binom{y+1}{m}\binom{n-mk}{y} \tag{8}$$

και επειδή κάθε μία από τις ακολουθίες περιλαμβάνει y αποτυχίες και $n-y$ επιτυχίες, η πιθανότητα της καθεμίας είναι $p^{n-y}q^{y}$.

Αντικαθιστώντας τις σχέσεις (8) και (7) στη σχέση (6) προκύπτει η ισότητα

$$r_m = \sum_{y=m-1}^{n-mk} \binom{y+1}{m}\binom{n-mk}{y} p^{n-y}q^{y}$$

και κατ΄ επέκταση, προκύπτει ο παρακάτω τύπος για τη συνάρτηση πιθανότητας της τυχαίας μεταβλητής $G_{n,k}$,

$$P(G_{n,k}=x) = \sum_{m=x}^{\lfloor \frac{n+1}{k+1} \rfloor} (-1)^{m-x} \binom{m}{x} \sum_{y=m-1}^{n-mk} \binom{y+1}{m}\binom{n-mk}{y} p^{n-y}q^{y}.$$

$\diamond$

Παρατήρηση Αν κάνουμε μια εναλλαγή στη σειρά των αθροισμάτων στον τύπο (5) προκύπτει η ισοδύναμη σχέση:

$$P(G_{n,k}=x) = \sum_{y=x-1}^{n-kx} p^{n-y}q^{y} \sum_{m=x}^{\min(y+1, \lfloor \frac{n-y}{k} \rfloor)} (-1)^{m-x} \binom{m}{x}\binom{y+1}{m}\binom{n-mk}{y},$$

που είναι όμοια με την μαθηματική έκφραση του *Godbole* (1990), για τη συνάρτηση πιθανότητας της τυχαίας μεταβλητής $N_{n,k}$ ($P(N_{n,k}=x)$). Πιο συγκεκριμένα, αν θέσουμε στον τύπο (5) όπου $j=m-x$, προκύπτει ο τύπος του *Godbole* (1990).

Χρησιμοποιώντας το αποτέλεσμα του Θεωρήματος 1.4 ο *Muselli* (1996) απέδειξε έναν ακόμα πιο απλό τύπο, ο οποίος περιλαμβάνει ένα μοναδικό άθροισμα, για τη συνάρτηση πιθανότητας της τυχαίας μεταβλητής $G_{n,k}$ και έτσι ο υπολογισμός της γίνεται ακόμη ευκολότερος.

Θεώρημα 1.5 *Αν $G_{n,k}$ είναι ο αριθμός των ροών επιτυχιών μήκους k ή και περισσότερο σε n δοκιμές Bernoulli, τότε έχουμε:*

$$P(G_{n,k}=x)=\sum_{m=x}^{\lfloor\frac{n+1}{k+1}\rfloor}(-1)^{m-x}\binom{m}{x}p^{mk}q^{m-1}\left(\binom{n-mk}{m-1}+q\binom{n-mk}{m}\right),$$

όπου k, n και x είναι θετικοί ακέραιοι.

Απόδειξη Κάνοντας μια αντικατάσταση στον τύπο (5) θέτοντας ως $j=n-mk-y$, θα έχουμε:

$$\begin{aligned}
&P(G_{n,k}=x)\\
=&\sum_{m=x}^{\lfloor\frac{n+1}{k+1}\rfloor}\binom{m}{x}(-1)^{m-x}\sum_{j=0}^{n-mk-m+1}\binom{n-mk-j+1}{m}\\
&\times\binom{n-mk}{j}p^{mk+j}q^{n-mk-j}\\
=&\sum_{m=x}^{\lfloor\frac{n+1}{k+1}\rfloor}(-1)^{m-x}\binom{m}{x}p^{mk}q^{n-mk}\sum_{j=0}^{n-mk-m+1}\binom{n-mk-j+1}{m}\\
&\times\binom{n-mk}{j}(p/q)^{j}. \qquad (9)
\end{aligned}$$

Χρησιμοποιώντας την ιδιότητα του τριγώνου *Pascal*, θα έχουμε ότι:

$$\begin{aligned}
&\sum_{j=0}^{n-mk-m+1}\binom{n-mk-j+1}{m}\binom{n-mk}{j}(p/q)^{j}\\
=&\sum_{j=0}^{n-mk-m+1}\left(\binom{n-mk-j}{m-1}+\binom{n-mk-j}{m}\right)\binom{n-mk}{j}(p/q)^{j}\\
=&\sum_{j=0}^{n-mk-m+1}\left(\binom{n-mk-j}{n-mk-m+1-j}+\binom{n-mk-j}{n-mk-m-j}\right)\binom{n-mk}{j}(p/q)^{j}\\
=&\binom{n-mk}{m-1}(1/q)^{n-mk-m+1}+\binom{n-mk}{m}(1/q)^{n-mk-m}, \qquad (10)
\end{aligned}$$

αν εφαρμόσουμε τη σχέση ($Feller$, 1968, $p.63$)

$$\sum_{v\geq 0}\binom{h}{v}\binom{h-v}{r-v}t^v=\binom{h}{r}(1+t)^r.$$

Τέλος αντικαθιστώντας τη σχέση (10) στην (9) προκύπτει το ζητούμενο.

$\diamond$

Εκμεταλευόμενοι τη σχέση που συνδέει τις τυχαίες μεταβλητές, $G_{n,k}$ και L_n μπορούμε να πάρουμε έναν αντίστοιχα απλό τύπο για τη συνάρτηση κατανομής της δεύτερης. Πιο συγκεκριμένα έχουμε το παρακάτω πόρισμα.

Πόρισμα *Αν L_n παριστάνει το μήκος της μεγαλύτερης ροής επιτυχιών σε n δοκιμές Bernoulli, έχουμε*

$$P(L_n\leq k-1)=\sum_{m=0}^{\lfloor\frac{n+1}{k+1}\rfloor}(-1)^m p^{mk}q^{m-1}\left(\binom{n-mk}{m-1}+q\binom{n-mk}{m}\right),$$

όπου k και n είναι θετικοί ακέραιοι.

Απόδειξη Γνωρίζοντας ότι ισχύει η σχέση:

$$P(L_n\geq k)=P(G_{n,k}\geq 1)=\sum_{x=1}^{\lfloor\frac{n+1}{k+1}\rfloor}P(G_{n,k}=x)$$

και λαμβάνοντας υπόψη το Θεώρημα 1.5 έχουμε,

$$\begin{aligned}P(L_n\leq k-1) &= 1-P(L_n\geq k)=1-\sum_{x=1}^{\lfloor\frac{n+1}{k+1}\rfloor}P(G_{n,k}=x)\\ &= 1-\sum_{x=1}^{\lfloor\frac{n+1}{k+1}\rfloor}\sum_{m=x}^{\lfloor\frac{n+1}{k+1}\rfloor}(-1)^{m-x}\binom{m}{x}p^{mk}q^{m-1}\\ &\quad\times\left(\binom{n-mk}{m-1}+q\binom{n-mk}{m}\right).\end{aligned}$$

$\diamond$

Παράδειγμα 1.5 Έστω $n=10$ δοκιμές $Bernoulli$ με πιθανότητα επιτυχίας p, $(0\leq p\leq 1)$. Αν $k=4$, τότε σύμφωνα με τον τύπο του $Muselli$ (1996),

$$P(L_{10}\leq 3)=\sum_{m=0}^{2}(-1)^m p^{4m}q^{m-1}\left(\binom{10-4m}{m-1}+q\binom{10-4m}{m}\right).$$

Για $m = 0$

$$(-1)^0 p^0 q^{-1}\left(\binom{10}{-1} + q\binom{10}{0}\right) = q^{-1}(0+q) = 1$$

Για $m = 1$

$$(-1)^1 p^4 q^0\left(\binom{10-4}{0} + q\binom{6}{1}\right) = -p^4[1+6q] = -p^4 - 6p^4 q$$

Για $m = 2$

$$(-1)^2 p^8 q^1\left(\binom{2}{1} + q\binom{2}{2}\right) = -p^8 q^1[2+q] = 2p^8 q + p^8 q^2$$

Για $p = q = 0.5$,

$$\begin{aligned} P(L_{10} \leq 3) &= 1 - p^4 - 6p^5 + 2p^9 + p^{10} \\ &= 0.0009765625 + 0.00390625 - 0.1875 - 0.0625 + 1 \\ &= 0.7548828125 \end{aligned}$$

1.3 Συνάρτηση πιθανότητας της τυχαίας μεταβλητής L_n

Αργότερα οι *Makri* και *Psillakis* (2009) μέσω της συνάρτησης πιθανότητας της τυχαίας μεταβλητής $G_{n,k}$, έδωσαν την ακριβή κατανομή της μεταβλητής L_n. Έτσι, στην περίπτωση που τα πειράματα είναι ανεξάρτητα (όχι αναγκαστικά ι-σόνομα), με $P(X_i = 1) = p_i$, $i = 1, 2, \ldots, n$, απέδειξαν ότι η κατανομή της $G_{n,k}$, ικανοποιεί έναν αναδρομικό τύπο. Πιο συγκεκριμένα έχουμε το παρακάτω θεώρημα.

Θεώρημα 1.6 *Η συνάρτηση πιθανότητας* $g(x; k, 1, n) = P(G_{n,k} = x), x \in S_{G_{n,k}}$ *ικανοποιεί τον αναδρομικό τύπο:*

$$g(x; k, 1, n) = \sum_{i=0}^{n-1} \beta_i g(x - I_A(i); k, i+2, n) + \beta\delta_{x,1}, \quad n \geq k \tag{11}$$

όπου $\beta_i = q_{i+1}\left(\prod_{j=1}^{i} p_j\right)$, $i = 0, 1, \ldots, n-1$, $\beta = \prod_{j=1}^{n} p_j$, $A = \{k, k+1, \ldots, n-1\}$, $S_{G_{n,k}} = \{0, 1, \ldots, \lfloor\frac{n+1}{k+1}\rfloor\}$, $I_A(i) = 1$ *αν* $i \in A$

με αρχικές συνθήκες

$$g(x;k,1,n) = 0, \quad \text{αν } x < 0 \text{ ή } x > \lfloor \tfrac{n+1}{k+1} \rfloor,$$

$$g(x;k,1,n) = \begin{cases} 1, \text{αν } x = 0 \\ 0, \text{αν } x > 0 \end{cases}, \text{για } 0 \leq n < k.$$

Απόδειξη Αρχικά είναι φανερό ότι ισχύει το θεώρημα για $x < 0$, $x > \lfloor \frac{n+1}{k+1} \rfloor$ και $0 \leq n < k$. Στη συνέχεια θα ορίσουμε τα γεγονότα

- $A_i = (X_1 = X_2 = \ldots = X_i = 1, X_{i+1} = 0)$, $i = 0, \ldots, n-1$.
- $A_n = (X_1 = X_2 = \ldots = X_n = 1)$.

Τότε για $n \geq k$, $x = 0, 1, \ldots, \lfloor \frac{n+1}{k+1} \rfloor$ έχουμε:

$$\begin{aligned} g(x;k,1,n) &= P(\bigcup_{i=0}^{n}[(G_{n,k} = x) \cap A_i]) \\ &= \sum_{i=0}^{k-1} P(G_{n,k} = x|A_i)P(A_i) + \sum_{i=k}^{n-1} P(G_{n,k} = x|A_i)P(A_i) \\ &\quad + P(G_{n,k} = x|A_n)P(A_n). \end{aligned}$$

Για κάθε ένα άθροισμα έχουμε ότι:

- <u>Για $i = 0, 1, \ldots, k-1$</u>
  Από τον ορισμό του συνόλου A_i, προκύπτει ότι:

$$\begin{aligned} A_0 &= (X_1 = 0) \\ A_1 &= (X_1 = 1, X_2 = 0) \\ A_2 &= (X_1 = X_2 = 1, X_3 = 0) \\ &\vdots \\ A_{k-1} &= (X_1 = X_2 = \ldots = X_{k-1} = 1, X_k = 0). \end{aligned}$$

  Έτσι, η πιθανότητα να έχουμε x ροές επιτυχιών μήκους τουλάχιστον k μεταξύ των δοκιμών 1 και n δοθέντος του A_i, ισούται με την πιθανότητα να έχω τον ίδιο αριθμό ροών μεταξύ των δοκιμών $1+i+1$, n. Δηλαδή,

$$P(G_{n,k} = x|A_i) = g(x;k,i+2,n)$$

  Αυτό γίνεται περισσότερο κατανοητό, αν παρατηρήσουμε τον τρόπο που έχουν ορισθεί τα $A_0, A_1, A_2, \ldots, A_{k-1}$, οπου στην χειρότερη περίπτωση δεν θα έχουμε καμία ροή επιτυχιών και στην καλύτερη υπάρχει ροή επιτυχιών με μήκος το πολύ $k-1$.

- <u>Για $i = k, k+1, \ldots, n-1$</u>
  Όμοια με παραπάνω, θα έχουμε ότι:

  $$P(G_{n,k} = x | A_i) = g(x-1; k, i+2, n)$$

  και αυτό συμβαίνει λόγω του ορισμού των A_i και των τιμών του i, $i = k, k+1, \ldots, n-1$, που σημαίνει ότι θα έχει ήδη δημιουργηθεί μια ροή μήκους τουλάχιστον k στα πρώτα i πειράματα.

- Επίσης, έχουμε ότι:

  $$P(G_{n,k} = x | A_n) \;\; = \;\; \delta_{x,1}$$

όπου, $P(A_i) = p_1 p_2 \cdots p_i q_{i+1}, \quad i = 0, 1, \ldots, n-1$ και $P(A_n) = p_1 p_2 \cdots p_n$. Έτσι προκύπτει το ζητούμενο.

◇

Επανερχόμαστε στην εύρεση της συνάρτησης πιθανότητας της μεταβλητής L_n. Ισχύει ότι:

$$P(L_n = k) = \begin{cases} P(G_{n,1} = 0) & \text{αν } k = 0 \\ \\ P(G_{n,k+1} = 0) - P(G_{n,k} = 0) & \text{αν } k = 1, 2, \ldots, n-1 \\ \\ 1 - P(G_{n,k} = 0) & \text{αν } k = n \end{cases}$$

- <u>Όταν $k = 0$</u>, δηλαδή όταν το μέγιστο μήκος ροής επιτυχιών είναι 0, τότε δεν εμφανίζεται καμία επιτυχία στην ακολουθία και αυτό έχει σαν αποτέλεσμα να ισχύει $G_{n,1} = 0$.

- <u>Όταν $k = 1, 2, \ldots, n-1$</u>

$$\begin{aligned} P(L_n = k) &= P(L_n \leq k) - P(L_n \leq k-1) \\ &= P(G_{n,k+1} = 0) - P(G_{n,k} = 0). \end{aligned}$$

- <u>Όταν $k = n$</u>, δηλαδή όταν το μέγιστο μήκος ροής επιτυχιών ειναι n, τότε όλα τα πειράματα είναι επιτυχίες και με βάση το παραπάνω σκεπτικό προκύπτει το ζητούμενο.

Χρησιμοποιώντας μεθόδους συνδυαστικής ανάλυσης οι *Makri*, *Philippou* και *Psillakis* (2007α) απέδειξαν μια μαθηματική έκφραση για τον άμεσο υπολογισμό της συνάρτησης πιθανότητας της τυχαίας μεταβλητής L_n.
Για την απόδειξη του αντίστοιχου θεωρήματος της ακριβούς κατανομής της L_n, θα χρησιμοποιηθεί το παρακάτω λήμμα.

Λήμμα 1.5 (*Freud*, 1956 *Riordan*, 1964, *p.* 104)
Ο αριθμός των διαφορετικών διανομών των n όμοιων σφαιριδίων σε r διακεκριμένες κάλπες, έτσι ώστε καθεμιά να περιέχει το πολύ k σφαιρίδια δίνεται από τον τύπο:

$$C(n,r,k) = \sum_{j=0}^{r}(-1)^j \binom{r}{j} \binom{n-(k+1)j+r-1}{r-1},$$

αν $n > 0$; 1, αν $n = 0$; και 0 διαφορετικά.

Απόδειξη Θεωρούμε την γεννήτρια συνάρτηση:

$$g(t) = (1+t+t^2+\ldots+t^k)^r = \underbrace{(t^0+t+t^2+\ldots+t^k)\cdots(t^0+t+t^2+\ldots+t^k)}_{r \text{ φορές}}$$

Επειδή αυτή η συνάρτηση είναι ένα πολυώνυμο ώς προς t, βαθμού rk, σημαίνει ότι είναι της μορφής:

$$g(t) = \alpha_0 t^0 + \alpha_1 t^1 + \alpha_2 t^2 + \ldots + \alpha_n t^n + \ldots + \alpha_{rk} t^{rk},$$

όπου α_i είναι συντελεστές του πολυωνύμου και είναι πραγματικοί αριθμοί. Κάθε ένας από τους παράγοντες α_i μας δίνει τους διαφορετικούς τρόπους με τους οποίους μπορούν να συνδυαστούν οι δυνάμεις του t, κάθε μία από διαφορετικό παράγοντα, έτσι ώστε να προκύπτει κάθε φορά το αντίστοιχο t^i

$$\begin{aligned}
g(t) &= (1+t+t^2+\ldots+t^k)^r \\
&= \underbrace{(t^0+t+t^2+\ldots+t^k)\cdots(t^0+t+t^2+\ldots+t^k)}_{r \text{ φορες}} \\
&= \left(\frac{1-t^{k+1}}{1-t}\right)^r \\
&= \frac{(1-t^{k+1})^r}{(1-t)^r}
\end{aligned}$$

$$
\begin{aligned}
&= (1-t^{k+1})^r(1-t)^{-r} \\
&= \sum_{i=0}^{r}\binom{r}{i}(-t^{k+1})^i\sum_{j=0}^{\infty}\binom{r+j-1}{j}t^j.
\end{aligned}
$$

Ο συντελεστής του t^n παριστάνει όλους τους δυνατούς τρόπους με τους οποίους μπορούμε να διανείμουμε n διακεκριμένες σφαίρες σε r κάλπες έτσι ώστε κάθε μία να έχει χωρητικότητα το πολύ k σφαίρες. Έτσι, για τον α_n έχουμε:

$$
\alpha_n = \sum_{i(k+1)+j=n}(-1)^n\binom{r}{i}\binom{r+j-1}{r-1}
$$

και επειδή $i(k+1)+j=n \Longrightarrow j=n-i(k+1)$ προκύπτει τελικά ότι:

$$
\alpha_n = \sum_{i=0}^{\lfloor \frac{n}{k} \rfloor}(-1)^i\binom{r}{i}\binom{n+r-i(k+1)-1}{r-1}.
$$

Χρησιμοποιώντας διωνυμικούς συντελεστές, έχουμε το παρακάτω θεώρημα που μας δίνει έναν ακριβή τύπο για τη συνάρτηση πιθανότητας της L_n.

$\diamond$

Θεώρημα 1.7 *Έστω L_n είναι το μήκος της μεγαλύτερης ροής επιτυχιών σε n δοκιμές Bernoulli με πιθανότητα επιτυχίας p. Τότε:*

$$
\begin{aligned}
P(L_n=0) &= q^n \\
P(L_n=k) &= p^n\sum_{y=\lfloor n/(k+1)\rfloor}^{n-k}\left(\frac{q}{p}\right)^y\sum_{i=1}^{y+1}\binom{y+1}{i} \\
&\quad \times C(n-y-ik, y+1-i, k-1),\ k=1,\dots,n. \qquad (12)
\end{aligned}
$$

Απόδειξη Προφανώς το (1) ισχύει. Για $k=1,\dots,n$, θεωρούμε την τυχαία μεταβλητή Y_n ως το συνολικό αριθμό των αποτυχιών σε μια ακολουθία από n δοκιμές *Bernoulli*. Αν πάρουμε το γεγονός $\{L_n=k, Y_n=y\}$, τότε ένα τυπικό στοιχείο αυτού θα είναι μια ακολουθία αποτελούμενη από y αποτυχίες και $n-y$ επιτυχίες, υπό τον περιορισμό ότι το μέγιστο μήκος ροής επιτυχιών της να ισούται με k.
Αν θεωρήσουμε ότι:

- Κάθε επιτυχία έχει πιθανότητα p (υπάρχουν $n-y$ επιτυχίες)

- Κάθε αποτυχία έχει πιθανότητα q (υπάρχουν y αποτυχίες)
- $N(n, y, k)$ είναι ο αριθμός των στοιχείων του γεγονότος $\{L_n = k, Y_n = y\}$

έχουμε ότι:

$$P(L_n = k, Y_n = y) = N(n, y, k)p^{n-y}q^y.$$

Επειδή, η κάθε διάταξη περιέχει y αποτυχίες θα δημιουργούνται $y+1$ κελιά στα οποία θα διανέμονται με κατάλληλο τρόπο οι σφαίρες - επιτυχίες. Τα κελιά αυτά τα συμβολίζουμε ως $U_1, U_2, \ldots, U_{y+1}$ και μάλιστα το κελί U_1 βρίσκεται πριν την πρώτη αποτυχία και το U_{y+1} βρίσκεται μετά την $y-$οστή αποτυχία. Αν έχουμε, για παράδειγμα, $y = 5$ αποτυχίες, τότε θα έχουμε τα ακόλουθα 6 κελιά:

$$\underbrace{\ldots}_{U_1} F \underbrace{\ldots}_{U_2} F \underbrace{\ldots}_{U_3} F \underbrace{\ldots}_{U_4} F \underbrace{\ldots}_{U_5} F \underbrace{\ldots}_{U_6}$$

Έτσι, η περίπτωσή μας ανάγεται σε πρόβλημα διανομής $n - y$ όμοιων σφαιρών σε $y + 1$ διακεκριμένα κελιά κάτω από τους περιορισμούς του προβλήματος και έτσι θα μπορέσουμε να ορίσουμε τον αριθμό $N(n, y, k)$.

Έστω ότι i κελιά από τα $U_1, U_2, \ldots, U_{y+1}$, $i = 1, \ldots, \min\{y + 1, \lfloor (n - y)/k \rfloor\}$ περιέχουν ροές επιτυχιών μήκους k. Αν αυτά θέλουμε να τα επιλέξουμε από τα συνολικά $y + 1$, μπορούμε να το πραγματοποιήσουμε με $\binom{y+1}{i}$ διαφορετικούς τρόπους. Ακόμη, αν συμβολίσουμε με $(j_1, j_2, \ldots, j_i)$ τις i επιλογές από τις $y + 1$, τότε κάθε φορά που κάνουμε μία επιλογή από αυτές αυτόματα σημαίνει ότι μια διάταξη από k συνεχόμενες επιτυχίες λαμβάνει χώρα σε κάθε ένα από τα i κελιά $U_{j_1}, U_{j_2}, \ldots, U_{j_i}$ με μοναδικό τρόπο. Εφόσον τοποθετηθούν αυτά, μας απομένουν $n - y - ik$ επιτυχίες οι οποίες πρέπει να διανεμηθούν σε $y + 1 - i$ κελιά έτσι ώστε κανένα από αυτά να μην περιέχει k συνεχόμενες επιτυχίες ή παραπάνω, δηλαδή το πολύ $k - 1$ επιτυχίες.
Σύμφωνα με το Λήμμα 1.5 ο αριθμός των τρόπων που μπορεί να πραγματοποιηθεί η συγκεκριμένη διανομή των επιτυχιών ισούται με

$$C(\underbrace{n - y - ik}_{\text{οι επιτυχίες}}, \underbrace{y + 1 - i}_{\text{τα κελιά}}, \underbrace{k - 1}_{\text{χωρητικότητα}}).$$

Έτσι προκύπτει ότι:

$$N(n, y, k) = \sum_{i=1}^{y+1} \binom{y+1}{i} C(n - y - ik, y + 1 - i, k - 1).$$

Αθροίζοντας πάνω στα y, $y = \lfloor n/(k+1) \rfloor, \ldots, n-k$, προκύπτει το ζητούμενο. Δηλαδή:

$$\begin{aligned} P(L_n = k) &= \sum_{y=\lfloor n/(k+1)\rfloor}^{n-k} P(L_n = k, Y_n = y) \\ &= \sum_{y=\lfloor n/(k+1)\rfloor}^{n-k} N(n,y,k) p^{n-y} q^y \\ &= \sum_{y=\lfloor n/(k+1)\rfloor}^{n-k} \sum_{i=1}^{y+1} \binom{y+1}{i} C(n-y-ik, y+1-i, k-1) p^n p^{-y} q^y \\ &= p^n \sum_{y=\lfloor n/(k+1)\rfloor}^{n-k} \left(\frac{q}{p}\right)^y \sum_{i=1}^{y+1} \binom{y+1}{i} C(n-y-ik, y+1-i, k-1). \end{aligned}$$

$\diamond$

Παράδειγμα 1.6 Δίνεται μια ακολουθία από $n = 10$ δοκιμές *Bernoulli* διατεταγμένες σε σειρά με πιθανότητα επιτυχίας $p = 1/2$. Χρησιμοποιώντας το θεώρημα των *Makri*, *Philippou* και *Psillakis* (2007α), θα υπολογίσουμε την πιθανότητα το μέγιστο μήκος ροής επιτυχιών να είναι ίσο με $k = 4$.

Αρχικά με βάση τα δεδομένα μας έχουμε ότι η μεταβλητή Y θα παίρνει τιμές:

$$y = \left\lfloor \frac{n}{k+1} \right\rfloor, \ldots, n-k = \left\lfloor \frac{10}{5} \right\rfloor, \ldots, (10-4) = 2, \ldots, 6.$$

Στη συνέχεια χρησιμοποιώντας τον τύπο (12) του Θεωρήματος 1.7 έχουμε ότι:

$$P(L_{10} = 4) = \left(\frac{1}{2}\right)^{10} \sum_{y=2}^{6} \left(\frac{0.5}{0.5}\right)^y \sum_{i=1}^{y+1} \binom{y+1}{i} C(10-y-i4, y+1-i, 3). (13)$$

Έτσι θα έχουμε ότι:

- Για $y = 2$:

$$\binom{3}{1} C(4,2,3) + \binom{3}{2} C(0,1,3) + \binom{3}{3} C(-4,0,3)$$

$$\binom{3}{1} \cdot 0 + \binom{3}{2} \cdot 0 + \binom{3}{3} \cdot 0 = 0$$

- Για $y = 3$:

$$\begin{aligned}
& \binom{4}{1} C(3,3,3) + \binom{4}{2} C(-1,2,3) + \binom{4}{3} C(-5,1,3) \\
& + \binom{4}{4} C(-9,0,3) \\
= & \binom{4}{1} \cdot 5 + \binom{4}{2} C(-1,2,3) \cdot 0 + \binom{4}{3} \cdot 0 + \binom{4}{4} \cdot 0 \\
= & 20
\end{aligned}$$

- Για $y = 4$:

$$\begin{aligned}
& \binom{5}{1} C(2,4,3) + \binom{5}{2} C(-2,3,3) + \binom{5}{3} C(-6,2,3) \\
& + \binom{5}{4} C(-10,1,3) + \binom{5}{5} C(-14,0,3) \\
= & \binom{5}{1} \cdot 0 + \binom{5}{2} \cdot 0 + \binom{5}{3} \cdot 0 \\
& + \binom{5}{4} \cdot 0 + \binom{5}{5} \cdot 0 \\
= & 0
\end{aligned}$$

- Για $y = 5$:

$$\begin{aligned}
& \binom{6}{1} C(1,5,3) + \binom{6}{2} C(-3,4,3) + \binom{6}{3} C(-7,3,3) \\
& + \binom{6}{4} C(-11,2,3) + \binom{6}{5} C(-15,1,3) + \binom{6}{6} C(-19,0,3) \\
= & 0
\end{aligned}$$

- Για $y = 6$:

$$\binom{7}{1} C(0,6,3) + \binom{7}{2} C(-4,5,3) + \binom{7}{3} C(-8,4,3)$$

$$
\begin{aligned}
&+\binom{7}{4}C(-12,3,3)+\binom{7}{5}C(-16,2,3)+\binom{7}{6}C(-21,1,3)\\
&+\binom{7}{7}C(-24,0,3)\\
=\ &0
\end{aligned}
$$

Έτσι,

$$P(L_{10}=4)=\left(\frac{1}{2}\right)^{10}\cdot 20=0.01953125.$$

Κεφάλαιο 2

Μελέτη της μεταβλητής L_n σε ανταλλάξιμες δυαδικές ακολουθίες

2.1 Περιγραφή μοντέλου

Ορισμός: Οι τυχαίες μεταβλητές $\xi_1, \xi_2, \ldots, \xi_n$ θα χαρακτηρίζονται ανταλλάξιμες αν για κάθε n ισχύει η ισότητα:

$$P(\xi_1 \leq x_1, \ldots, \xi_n \leq x_n) = P(\xi_{\pi_1} \leq x_1, \ldots, \xi_{\pi_n} \leq x_n),$$

για κάθε μετάθεση $\pi = (\pi_1, \ldots, \pi_n)$ του $\{1, 2, \ldots, n\}$ και για κάθε $x_i \in R$, $i = 1, \ldots, n$.
Με άλλα λόγια θα πρέπει η από κοινού κατανομή των μεταβλητών $\xi_1, \ldots, \xi_n$ να παραμένει σταθερή υπό οποιαδήποτε ματάθεση.

Βασισμένος στην υδρολογία ο *Eryilmaz* (2005α) περιέγραψε ένα μοντέλο το οποίο βοηθάει περισσότερο στην κατανόηση της προέλευσης των ανταλλάξιμων μεταβλητών. Πιο συγκεκριμένα:

Έστω ότι βρισκόμαστε σε μια περιοχή στην οποία τα προηγούμενα n χρόνια έχουν πραγματοποιηθεί μετρήσεις για τις πλημμύρες (ξηρασίες) που έχουν σημειωθεί, δημιουργώντας έτσι μια ακολουθία από n ανεξάρτητες και ισόνομες μεταβλητές $X_1, X_2, \ldots, X_n$.

Θεωρούμε επίσης ότι οι $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$ παριστάνουν τις πλημμύρες (ξηρασίες) για τα επόμενα m χρόνια στην ίδια περιοχή. Στη συνέχεια διατάσσουμε τις μετρήσεις $X_1, X_2, \ldots, X_n$, παίρνοντας ως τυχαίο κατώφλι την r-οστή μικρότερη μέτρηση και την συμβολίζουμε με $X_{r:n}$.
Χρησιμοποιώντας το κατώφλι αυτό, για $i = 1, 2, \ldots, m$ ορίζουμε τις τυχαίες

μεταβλητές:

$$\xi_i = \begin{cases} 1, & \text{αν } X_{n+i} > X_{r:n} \\ \\ 0, & \text{διαφορετικά} \end{cases}$$

Στη συνέχεια γίνεται μια σχηματική περιγραφή του μοντέλου

$$\left.\begin{array}{l} X_1 \\ X_2 \\ \vdots \\ X_n \end{array}\right\} \longrightarrow^{X_{r:n}} \left\{\begin{array}{l} X_{n+1} > X_{r:n} \Longrightarrow \xi_1 = 1 \\ X_{n+2} < X_{r:n} \Longrightarrow \xi_2 = 0 \\ \vdots \\ X_{n+m} > X_{r:n} \Longrightarrow \xi_m = 1 \end{array}\right.$$

στην οποία παρατηρούμε ότι, από τις παρελθοντικές μετρήσεις παίρνουμε το κατώφλι ($X_{r:n}$) και ύστερα από σύγκριση αυτού με τις μελλοντικές, προκύπτει η δυαδική ακολουθία των στοιχείων $\xi_1, \xi_2, \ldots, \xi_m$ που δεν είναι κλασικές δοκιμές *Bernoulli*, αλλά ανταλλάξιμες. Ο λόγος είναι ότι επιλέξαμε με τυχαίο τρόπο το κατώφλι, ενώ σε αντίθετη περίπτωση θα ασχολούμασταν με ανεξάρτητες και ισόνομες δοκιμές *Bernoulli*.

Για να μπορέσουμε να μελετήσουμε την τ.μ. L_n θα πρέπει να ορίσουμε ως:

$$\begin{aligned} \lambda_k &= P(\xi_1 = \xi_2 = \ldots = \xi_k = 1), \\ \\ S_m &= \left|\{i \leq m : X_{n+i} > X_{r:n}\}\right|, \end{aligned}$$

όπου $|A|$ δηλώνει τον αριθμό των στοιχείων του A, με συνάρτηση πιθανότητας

$$P(S_m = k)\frac{\binom{m-k+r-1}{m-k}\binom{n-r+k}{k}}{\binom{m+n}{n}}, \quad k = 0, 1, \ldots, m.$$

2.2 Αθροιστική συνάρτηση κατανομής της L_n

Ένα πολύ σημαντικό θεώρημα που αφορά ανταλλάξιμες δυαδικές ακολουθίες, είναι αυτό των *George* και *Bowman* (1995). Οι δύο αυτοί ερευνητές, έδωσαν μαθηματικές εκφράσεις για την κατανομή του αθροίσματος των ανταλλάξιμων δυαδικών τυχαίων μεταβλητών καθώς και για την από κοινού κατανομή εκφρασμένες συναρτήσει των λ_i. Έτσι έχουμε το παρακάτω θεώρημα.

Θεώρημα 2.1 *Έστω $X_1, X_2, \ldots, X_n$ είναι ανταλλάξιμες δυαδικές τυχαίες μεταβλητές και έστω:*

$$\lambda_k = P(X_1 = X_2 = \ldots = X_k = 1),\ k = 1, 2, \ldots n \text{ και } \lambda_0 = 1.$$

Τότε έχουμε ότι

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \sum_{j=0}^{n-l}(-1)^j \binom{n-l}{j} \lambda_{l+j} \qquad (14)$$

και

$$P(S_n = l) \quad = \quad \binom{n}{l} \sum_{j=0}^{n-l}(-1)^j \binom{n-l}{j} \lambda_{l+j}, \qquad (15)$$

όπου $S_n = \sum_{i=1}^{n} X_i$ και $l = \sum_{i=1}^{n} x_i$.

Απόδειξη Έστω $X_1, \ldots, X_n$ είναι ανταλλάξιμες, θα δειχθεί ότι η πιθανότητα $P(X_1 = 1, \ldots, X_l = 1, X_{l+1} = 0, \ldots, X_n = 0)$ δίνεται από τον τύπο (14).
Από τον ορισμό της πιθανότητας λ_k και χρησιμοποιώντας το θεώρημα «εγκλεισμού-αποκλεισμού» έχουμε ότι:

$$\begin{aligned}
\lambda_l &= P(X_1 = 1, \ldots, X_l = 1) \\
&= \sum_{x_{l+1}, \ldots, x_n} P(X_1 = 1, \ldots, X_l = 1, X_{l+1} = x_{l+1}, \ldots, X_n = x_n) \\
&= P(X_1 = 1, \ldots, X_l = 1, X_{l+1} = 0, \ldots, X_n = 0) \\
&\quad + \binom{n-l}{1} \sum_{x_{l+2}, \ldots, x_n} P(X_1 = 1, \ldots, X_l = 1, X_{l+1} = 1, X_{l+2} = x_{l+2}, \ldots, X_n = x_n) \\
&\quad - \binom{n-l}{2} \sum_{x_{l+3}, \ldots, x_n} P(X_1 = 1, \ldots, X_{l+1} = 1, X_{l+2} = 1, X_{l+3} = x_{l+3}, \ldots, X_n = x_n) \\
&\quad + \ldots + (-1)^{n-l+1} P(X_1 = 1, \ldots, X_n = 1) \\
&= P(X_1 = 1, \ldots, X_l = 1, X_{l+1} = 0, \ldots, X_n = 0) + \sum_{j=1}^{n-l}(-1)^{j+1} \binom{n-l}{j} \lambda_{l+j}.
\end{aligned}$$

Έτσι,

$$P(X_1 = 1, \ldots, X_l = 1, X_{l+1} = 0, \ldots, X_n = 0) = \sum_{j=0}^{n-l}(-1)^j \binom{n-l}{j} \lambda_{l+j}.$$

$\diamond$

Ο τύπος (15) προκύπτει άμεσα από τον τύπο (14) λόγω του ότι οι μεταβλητές είναι ανταλλάξιμες και τα αποτελέσματά μας δεν επηρρεάζονται από τις ακριβείς θέσεις των επιτυχιών στην ακολουθία, αλλά μόνο από το συνολικό αριθμό αυτών.

Το Θεώρημα 2.1 αποτελεί το θεμέλιο για την εύρεση της κατανομής της τυχαίας μεταβλητής L_n για μια ακολουθία από ανταλλάξιμες δυαδικές δοκιμές. Αν πάρουμε τη δεσμευμένη συνάρτηση κατανομής $P(L_n < k|S_n = l)$, $0 \leq l \leq n$ τότε από τον ορισμό της δέσμευσης θα έχουμε:

$$\begin{aligned} P(L_n < k|S_n = l) &= \frac{P(L_n < k, S_n = l)}{P(S_n = l)} \\ &= \frac{N(n,k,l)P(l \text{ επιτυχίες και } n-l \text{ αποτυχίες})}{\begin{pmatrix} n \\ l \end{pmatrix} P(l \text{ επιτυχίες και } n-l \text{ αποτυχίες})} \\ &= \frac{N(n,k,l)}{\begin{pmatrix} n \\ l \end{pmatrix}}, \end{aligned} \tag{16}$$

όπου ο συντελεστής $N(n,k,l)$ παριστάνει τους διαφορετικούς τρόπους με τους οποίους μπορούμε να έχουμε τις l επιτυχίες και ταυτόχρονα το μέγιστο μήκος ροής επιτυχιών να είναι μικρότερο από την τιμή k. Παρατηρούμε ότι η ισότητα (16) ισχύει και στην περίπτωση των ανεξάρτητων και ισόνομων δοκιμών και μάλιστα έχουμε αποδείξει ότι :

$$P(L_n < k|S_n = l) = \begin{pmatrix} n \\ l \end{pmatrix}^{-1} \sum_{i=0}^{\lfloor l/k \rfloor} (-1)^i \begin{pmatrix} n-l+1 \\ i \end{pmatrix} \begin{pmatrix} n-ik \\ n-l \end{pmatrix}. \tag{17}$$

. Το αποτέλεσμα αυτό οφείλεται στο γεγονός ότι, όταν δεσμεύουμε με τον αριθμό των επιτυχιών απλοποιούνται οι ποσότητες λ_i.

Έτσι, χρησιμοποιώντας το θεώρημα ολικής πιθανότητας προκύπτει το ζητούμενο και πιο συγκεκριμένα:

$$P(L_n < k) = \sum_{l=0}^{n} P(L_n < k|S_n = l)P(S_n = l)$$

και με αντικατάσταση από τις σχέσεις (15) και (17), έχουμε ότι:

$$\begin{aligned} P(L_n < k) &= \sum_{l=0}^{n} \sum_{i=0}^{min(\lfloor l/k \rfloor, n-l+1)} \sum_{j=0}^{n-l} (-1)^i (-1)^j \begin{pmatrix} n-l+1 \\ i \end{pmatrix} \\ &\times \begin{pmatrix} n-ik \\ n-l \end{pmatrix} \begin{pmatrix} n-l \\ j \end{pmatrix} \lambda_{l+j}. \end{aligned}$$

Αντικαθιστώντας την ποσότητα

$$
\begin{aligned}
\lambda_{l+j} &= \frac{\begin{pmatrix} n-r+l+j \\ l+j \end{pmatrix}}{\begin{pmatrix} n+l+j \\ n \end{pmatrix}} \\
&= \frac{\begin{pmatrix} m-l+r-1 \\ m-l \end{pmatrix}\begin{pmatrix} n-r+l \\ l \end{pmatrix}}{\begin{pmatrix} m+n \\ n \end{pmatrix}}
\end{aligned}
$$

για τις μεταβλητές του μοντέλου που έχει περιγραφεί στην προηγούμενη παράγραφο, θα έχουμε ότι η συνάρτηση κατανομής θα ισούται:

$$
\begin{aligned}
P(L_m < k) &= \begin{pmatrix} m+n \\ n \end{pmatrix}^{-1} \sum_{l=0}^{m} \sum_{i=0}^{min(\lfloor l/k \rfloor, m-l+1)} \left[(-1)^i \begin{pmatrix} m \\ l \end{pmatrix}^{-1} \begin{pmatrix} m-l+1 \\ i \end{pmatrix} \right. \\
&\left. \times \begin{pmatrix} m-ik \\ m-l \end{pmatrix} \begin{pmatrix} m-l+r-1 \\ m-l \end{pmatrix} \begin{pmatrix} n-r+l \\ j \end{pmatrix} \right].
\end{aligned}
$$

Οι *Eryilmaz* και *Demir* (2007) χρησιμοποιώντας την τυχαία μεταβλητή $G_{n,k}$ καθώς επίσης και το γεγονός ότι η δεσμευμένη συνάρτηση κατανομής της τ.μ. L_n δοθέντος του αριθμού των επιτυχιών είναι ίδια τόσο στις ανεξάρτητες και ισόνομες όσο και στις ανταλλάξιμες δοκιμές, κατάφεραν να δώσουν έναν ακριβή τύπο για την κατανομή της τυχαίας μεταβλητής L_n.
Για να το επιτύχουν αυτό έκαναν χρήση μιας νέας τυχαίας μεταβλητής, η οποία συμβολίζεται με η_i.

Ορισμός: Η τυχαία μεταβλητή η_i μας δίνει το μήκος της ροής επιτυχιών μέχρι την $i-$οστή δοκιμή. Δηλαδή ισχύει ότι:

$$(\eta_i = k) = (\xi_i = \xi_{i-1} = \ldots = \xi_{i-(k-1)} = 1, \xi_{i-k} = 0),$$

όπου $\xi_1, \xi_2, \ldots$ είναι ανταλλάξιμες δυαδικές δοκιμές. Το βασικό χαρακτηριστικό αυτής είναι ότι φτάνει, για παράδειγμα, μέχρι την $i-$οστή δοκιμή, κοιτάει πίσω και μας δίνει το μήκος της ροής επιτυχιών έως τη συγκεκριμένη θέση της ακολουθίας. Αυτό που πρέπει να γνωρίζουμε είναι ότι οι μεταβλητές $(\eta_i)_{i\geq 1}$ έχουν τη Μαρκοβιανή ιδιότητα, μόνο στην περίπτωση όπου οι $\xi_1, \xi_2, \ldots$ είναι ανεξάρτητες και ισόνομες, ενώ δεν συμβαίνει το ίδιο όταν οι $\xi_1, \xi_2, \ldots$ είναι ανταλλάξιμες. Η ιδιαιτερότητα αυτή γίνεται περισσότερο εμφανής στο παρακάτω παράδειγμα.

Παράδειγμα 2.1(*Eryilmaz* και *Demir* (2007)) Έστω $\xi_1, \xi_2, \xi_3, \xi_4$ είναι ανταλλάξιμες δοκιμές και τα $\eta_i,\ i = 1, 2, 3, 4$ όπως έχουν ορισθεί παραπάνω. Θα υπολογίσουμε τη δεσμευμένη πιθανότητα $P(\eta_4 = 2|\eta_3 = 1, \eta_2 = 0, \eta_1 = 1)$. Με βάση τον ορισμό της δεσμευμένης πιθανότητας, έχουμε ότι:

$$P(\eta_4 = 2|\eta_3 = 1, \eta_2 = 0, \eta_1 = 1) = \frac{P\left(\eta_4 = 2, \eta_3 = 1, \eta_2 = 0, \eta_1 = 1\right)}{P\left(\eta_3 = 1, \eta_2 = 0, \eta_1 = 1\right)}. \quad (18)$$

Από τον ορισμό της τυχαίας μεταβλητής η_i συμπεραίνουμε ότι στη συγκεκριμένη ακολουθία ισχύει

- $\eta_4 = 2 \to \xi_4 = 1, \xi_3 = 1, \xi_2 = 0$
- $\eta_3 = 1 \to \xi_3 = 1, \xi_2 = 0$
- $\eta_2 = 0 \to \xi_2 = 0$
- $\eta_1 = 1 \to \xi_1 = 1$

Αντικαθιστώντας στον τύπο (18) έχουμε ότι :

$$P(\eta_4 = 2|\eta_3 = 1, \eta_2 = 0, \eta_1 = 1) = \frac{P\left(\xi_4 = 1, \xi_3 = 1, \xi_2 = 0, \xi_1 = 1\right)}{P\left(\xi_3 = 1, \xi_2 = 0, \xi_1 = 1\right)}.$$

Λόγω του ότι οι $\xi_1, \xi_2, \xi_3, \xi_4$ είναι ανταλλάξιμες μπορούμε να εφαρμόσουμε το θεώρημα των *George* και *Bowman* (1995) για να υπολογίσουμε τις απο κοινού πιθανότητες των $\xi_i,\ i = 1, 2, 3, 4$, που εμφανίζονται στον αριθμητή και στον παρονομαστή του παραπάνω πηλίκου. Πιο συγκεκριμένα:

$$\begin{aligned} P(\xi_4 = 1, \xi_3 = 1, \xi_2 = 0, \xi_1 = 1) &= \sum_{j=0}^{4-3} (-1)^j \begin{pmatrix} 4-3 \\ j \end{pmatrix} \lambda_{3+j} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \lambda_3 - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \lambda_4 \\ &= \lambda_3 - \lambda_4. \end{aligned}$$

Με τον ίδιο τρόπο έχουμε ότι:

$$\begin{aligned} P(\xi_3 = 1, \xi_2 = 0, \xi_1 = 1) &= \sum_{j=0}^{3-2} (-1)^j \begin{pmatrix} 3-2 \\ j \end{pmatrix} \lambda_{2+j} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \lambda_2 - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \lambda_3 \\ &= \lambda_2 - \lambda_3. \end{aligned}$$

Έτσι έχουμε για την δεσμευμένη πιθανότητα ότι:

$$P(\eta_4 = 2|\eta_3 = 1, \eta_2 = 0, \eta_1 = 1) = \frac{\lambda_3 - \lambda_4}{\lambda_2 - \lambda_3}.$$

Επίσης θα πρέπει να υπολογίσουμε την πιθανότητα, στην περίπτωση όπου η $i-$οστή δοκιμή εξαρτάται μόνο από την $(i-1)-$οστή δοκιμή, όταν δηλαδή η ακολουθία των η_i έχει την Μαρκοβιανή ιδιότητα.

$$P(\eta_4 = 2|\eta_3 = 1) = \frac{P(\eta_4 = 2, \eta_3 = 1)}{P(\eta_3 = 1)}. \tag{19}$$

Όμοια με παραπάνω ισχύει ότι:

- $\eta_4 = 2 \to \xi_4 = 1, \xi_3 = 1, \xi_2 = 0$
- $\eta_3 = 1 \to \xi_3 = 1, \xi_2 = 0$

Αντικαθιστώντας στον τύπο (19):

$$P(\eta_4 = 2|\eta_3 = 1) = \frac{P\left(\xi_4 = 1, \xi_3 = 1, \xi_2 = 0\right)}{P\left(\xi_3 = 1, \xi_2 = 0\right)},$$

και αν εφαρμόσουμε το θεώρημα των *George* και *Bowman* (1995) θα μπορέσουμε να φτάσουμε στο ζητούμενο. Έτσι,

$$\begin{aligned} P(\xi_4 = 1, \xi_3 = 1, \xi_2 = 0) &= \sum_{j=0}^{3-2}(-1)^j \binom{3-2}{j} \lambda_{2+j} \\ &= \binom{1}{0}\lambda_2 - \binom{1}{1}\lambda_3 \\ &= \lambda_2 - \lambda_3. \end{aligned}$$

και

$$\begin{aligned} P(\xi_3 = 1, \xi_2 = 0 &= \sum_{j=0}^{2-1}(-1)^j \binom{2-1}{j} \lambda_{1+j} \\ &= \binom{1}{0}\lambda_1 - \binom{1}{1}\lambda_2 \\ &= \lambda_1 - \lambda_2. \end{aligned}$$

Καταλήγουμε στο συμπέρασμα ότι ισχύει η αρχική υπόθεση, ότι οι ακολουθίες των η_i δεν έχουν τη Μαρκοβιανή ιδιότητα όταν οι ξ_i είναι ανταλλάξιμες και αυτό γιατί

$$\begin{aligned} P(\eta_4 = 2|\eta_3 = 1, \eta_2 = 0, \eta_1 = 1) &= \frac{\lambda_3 - \lambda_4}{\lambda_2 - \lambda_3} \\ &\neq \frac{\lambda_2 - \lambda_3}{\lambda_1 - \lambda_2} = P(\eta_4 = 2|\eta_3 = 1). \end{aligned}$$

Θεώρημα 2.2 *Αν η τυχαία μεταβλητή $G_{n,k}$ παριστάνει τον αριθμό των ροών επιτυχιών μεγέθους τουλάχιστον k σε n ανταλλάξιμες δοκιμές τότε η συνάρτηση πιθανότητας αυτής δίνεται από τον τύπο*

$$\begin{aligned} P(G_{n,k} = x) &= \sum_{l=kx}^{\alpha_1} \sum_{j=0}^{\alpha_2} \sum_{i=0}^{n-l} (-1)^j (-1)^i \binom{n-l-x+1}{j} \binom{n-l+1}{x} \\ &\times \binom{n-k(x+j)}{n-l} \binom{n-l}{i} \lambda_{l+i}, \quad x = 0, \ldots, \left\lfloor \frac{n+1}{k+1} \right\rfloor, \end{aligned}$$

όπου $\alpha_1 = \min(n, n-x+1)$, και $\alpha_2 = \min(\lfloor (l-kx)/k \rfloor, n-l-x+1)$.

Απόδειξη Παίρνουμε την πιθανότητα,

$$\begin{aligned} P(G_{n,k} = x) &= \sum_{\chi_1} P(G_{n,k} = x | S_n = l) P(S_n = l) \\ &= \sum_{l \in \chi_1} \sum_{j \in \chi_2} \sum_{i \in \chi_3} (-1)^j (-1)^i \binom{n-l-x+1}{j} \binom{n-l+1}{x} \\ &\times \binom{n-k(x+j)}{n-l} \binom{n-l}{i} \lambda_{l+i}, \end{aligned}$$

όπου :
$\chi_1 \equiv \{l : 0 \leq l \leq n, l - kx \geq 0, n - l + 1 \geq x\}$,
$\chi_2 \equiv \left\{j : j \leq n - l - x + 1, j \leq \lfloor \frac{l-kx}{k} \rfloor \right\}$,
$\chi_3 \equiv \{i : 0 \leq i \leq n - l\}$.

◇

Εφόσον γνωρίζουμε τη συνάρτηση πιθανότητας της τ.μ. $G_{n,k}$, μπορούμε να βρούμε τη συνάρτηση κατανομής της L_n, όπως φαίνεται στο παρακάτω λήμμα.

Λήμμα 2.1 *Αν η L_n παριστάνει το μέγιστο μήκος ροής επιτυχιών σε μια* *ακολουθία από n ανταλλάξιμες δυαδικές δοκιμές, ένας ακριβής τύπος για τη συνάρτηση κατανομής αυτής είναι:*

$$\begin{aligned} P(L_n < k) &= \sum_{l=0}^{n} \sum_{j=0}^{min(\lfloor l/k \rfloor, n-l+1)} \sum_{i=0}^{n-l} (-1)^j (-1)^i \begin{pmatrix} n-l+1 \\ j \end{pmatrix} \\ &\quad \times \begin{pmatrix} n-kj \\ n-l \end{pmatrix} \begin{pmatrix} n-l \\ i \end{pmatrix} \lambda_{l+i}, \quad k = 1, \ldots, n. \end{aligned}$$

Απόδειξη Εφόσον θέλουμε το μήκος της μέγιστης ροής επιτυχιών να μην ξεπερνά την τιμή k, τότε αυτό σημαίνει ότι δεν υπάρχει καμία ροή επιτυχιών που να έχει μήκος ίσο ή μεγαλύτερο του k. Έτσι, μπορούμε να γράψουμε ότι:

$$\begin{aligned} P(L_n < k) &= P(G_{n,k} = 0) \\ &= \sum_{l=0}^{n} \sum_{j=0}^{min(\lfloor l/k \rfloor, n-l+1)} \sum_{i=0}^{n-l} (-1)^j (-1)^i \begin{pmatrix} n-l+1 \\ j \end{pmatrix} \\ &\quad \times \begin{pmatrix} n-kj \\ n-l \end{pmatrix} \begin{pmatrix} n-l \\ i \end{pmatrix} \lambda_{l+i}. \end{aligned}$$

$\diamond$

Βάσει του Θεωρήματος 2.7 έχουμε τη δυνατότητα κάθε φορά να υπολογίζουμε την πιθανότητα $P(L_n \geq k)$ που, όπως θα δούμε και σε επόμενο κεφάλαιο, μας δίνει την αξιοπιστία ένος συνεχόμενου k-από-τα-n συστήματος, το οποίο λειτουργεί αν και μόνο αν στην ακολουθία υπάρχουν τουλάχιστον k συνεχόμενες συνιστώσες που λειτουργούν. Αυτό επιτυγχάνεται χρησιμοποιώντας την ισότητα $P(L_n \geq k) = 1 - P(L_n < k)$ και τον τύπο του Λήμματος 2.1. Ακόμη, στην περίπτωση όπου $2k \geq n$ ο υπολογισμός της συγκεκριμένης πιθανότητας γίνεται ευκολότερος, βάσει ενός απλού τύπου, όπως αυτός παρουσιάζεται στο παρακάτω λήμμα.

Λήμμα 2.2 *Για $2k \geq n$ έχουμε ότι:*

$$P(L_n \geq k) = (n-k+1)\lambda_k - (n-k)\lambda_{k+1}.$$

Απόδειξη Στην περίπτωση που έχουμε $2k \geq n$ και χρησιμοποιώντας την τ.μ. η_i ισχύει η ισότητα

$$P(L_n \geq k) = \sum_{i=k}^{n-1} P(\eta_i \geq k, \eta_{i+1} = 0) + P(\eta_n \geq k).$$

Για να γίνει πιο σαφής η ορθότητα αυτής, δεν έχουμε παρά να δούμε πώς προκύπτουν οι δύο όροι του δεύτερου μέλους.

- $P(\eta_i \geq k, \eta_{i+1} = 0)$: Από τον ορισμό της μεταβλητής η_i γνωρίζουμε ότι κινείται προς τα πίσω και φράσσεται πάντα από μια αποτυχία. Βάζοντας την ποσότητα $\eta_{i+1} = 0$ καταφέρνουμε να την φράξουμε και από μπροστά.

- $P(\eta_n \geq k)$: Στην περίπτωση αυτή δεν χρειάζεται να κάνουμε κάτι για να φράξουμε γιατί η ακολουθία τελειώνει.

Στη συνέχεια υπολογίζουμε καθεμία εξ' αυτών:

$$\begin{aligned}
& P(\eta_i \geq k, \eta_{i+1} = 0) \\
= \ & P(\eta_{i+1} = 0) - P(\eta_i < k, \eta_{i+1} = 0) \\
= \ & P(\eta_{i+1} = 0) - \sum_{j=0}^{k-1} P(\eta_i = j, \eta_{i+1} = 0) \\
= \ & (1 - \lambda_1) - \sum_{j=0}^{k-1} P(\xi_i = 1, \xi_{i-1} = 1, \ldots, \xi_{i-j+1} = 1, \xi_{i-j} = 0, \xi_{i+1} = 0) \\
= \ & (1 - \lambda_1) - \sum_{j=0}^{k-1} \sum_{k=0}^{2} (-1)^k \binom{2}{k} \lambda_{r+k} \ \ (\textit{George} \text{ και } \textit{Bowman}) \ (1995) \\
= \ & (1 - \lambda_1) - \sum_{j=0}^{k-1} (\lambda_j - 2\lambda_{j+1} + \lambda_{j+2}) = \lambda_k - \lambda_{k+1}.
\end{aligned}$$

Επίσης ισχύει ότι :

$$P(\eta_n \geq k) = \lambda_k.$$

Με μια απλή αντικατάσταση προκύπτει το ζητούμενο.

$\diamond$

Παράδειγμα 2.2 Έστω ότι έχουμε $n = 10$ ανταλλάξιμες δοκιμές και έστω ότι $k = 4$. Θέλουμε να βρούμε την πιθανότητα η τυχαία μεταβλητή L_n να παίρνει τιμές μικρότερες του k χρησιμοποιώντας τον τύπο των *Eryilmaz* και *Demir* (2007), δηλαδή:

$$\begin{aligned}
P(L_n < k) \ = \ & \sum_{l=0}^{n} \sum_{j=0}^{min(\lfloor l/k \rfloor, n-l+1)} \sum_{i=0}^{n-l} (-1)^j (-1)^i \binom{n-l+1}{j} \\
& \times \binom{n-kj}{n-l} \binom{n-l}{i} \lambda_{l+i}.
\end{aligned}$$

Εφόσον έχουμε $n = 10$, ο αριθμός των επιτυχιών θα παίρνει τιμές $l = 0, \ldots, 10$ και έτσι θα μπορέσουμε να βρούμε τη ζητούμενη πιθανότητα.

- *Για $l = 0$:* $j = 0$ και $i = 0, \ldots, 10$

$$\begin{aligned}
&\sum_{i=0}^{10}(-1)^i \binom{11}{j}\binom{10-4j}{10}\binom{10}{i}\lambda_i \\
= \ &\binom{11}{0}\binom{10}{10}\left[\binom{10}{0}\lambda_0 - \binom{10}{1}\lambda_1 + \binom{10}{2}\lambda_2 - \binom{10}{3}\lambda_3\right. \\
&+ \binom{10}{4}\lambda_4 - \binom{10}{5}\lambda_5 + \binom{10}{6}\lambda_6 - \binom{10}{7}\lambda_7 \\
&\left.+ \binom{10}{8}\lambda_8 - \binom{10}{9}\lambda_9 + \binom{10}{10}\lambda_{10}\right] \\
= \ &\lambda_0 - 10\lambda_1 + 45\lambda_2 - 120\lambda_3 + 210\lambda_4 - 252\lambda_5 + 210\lambda_6 - 120\lambda_7 + 45\lambda_8 \\
&-10\lambda_9 + \lambda_{10}
\end{aligned}$$

- *Για $l = 1$:* $j = 0$ και $i = 0, \ldots, 9$

$$\begin{aligned}
&\sum_{i=0}^{9}(-1)^i \binom{10}{j}\binom{10}{9}\binom{9}{i}\lambda_{i+1} \\
= \ &\binom{10}{0}\binom{10}{9}\left[\binom{9}{0}\lambda_1 - \binom{9}{1}\lambda_2 + \binom{9}{2}\lambda_3 - \binom{9}{3}\lambda_4\right. \\
&+ \binom{9}{4}\lambda_5 - \binom{9}{5}\lambda_6 + \binom{9}{6}\lambda_7 - \binom{9}{7}\lambda_8 \\
&\left.+ \binom{9}{8}\lambda_9 - \binom{9}{9}\lambda_{10}\right] \\
= \ &10\lambda_1 - 90\lambda_2 + 360\lambda_3 - 840\lambda_4 + 1260\lambda_5 - 1260\lambda_6 + 840\lambda_7 - 360\lambda_8 \\
&+90\lambda_9 - 10\lambda_{10}
\end{aligned}$$

- *Για $l = 2$:* $j = 0$ και $i = 0, \ldots, 8$ Ακολουθώντας τον ίδιο τρόπο υπολογισμού έχουμε:

$$\sum_{i=0}^{8}(-1)^i \binom{9}{j}\binom{10}{8}\binom{8}{i}\lambda_{2+i}$$

$$= \quad 45\lambda_2 - 360\lambda_3 + 1260\lambda_4 - 2520\lambda_5 + 3150\lambda_6 - 2520\lambda_7 + 1260\lambda_8$$

$$- \quad 360\lambda_9 + 45\lambda_{10}$$

- *<u>Για $l = 3$:</u>* $j = 0$ και $i = 0, \ldots, 7$

$$\sum_{i=0}^{7}(-1)^i \binom{8}{j}\binom{10}{7}\binom{7}{i}\lambda_{3+i}$$

$$= \quad 120\lambda_3 - 840\lambda_4 + 2520\lambda_5 - 4200\lambda_6 + 4200\lambda_7 - 2520\lambda_8$$

$$+ \quad 840\lambda_9 - 120\lambda_{10}$$

- *<u>Για $l = 4$:</u>* $j = 0, 1$ και $i = 0, \ldots, 6$

$$\sum_{i=0}^{6}(-1)^i \binom{7}{0}\binom{10}{6}\binom{6}{i}\lambda_{4+i} - \sum_{i=0}^{6}(-1)^i \binom{7}{1}\binom{6}{6}\binom{7}{i}\lambda_{4+i}$$

$$= \quad 203\lambda_4 - 1218\lambda_5 + 3045\lambda_6 - 4060\lambda_7 + 3045\lambda_8 - 1218\lambda_9 + 203\lambda_{10}$$

- *<u>Για $l = 5$:</u>* $j = 0, 1$ και $i = 0, \ldots, 5$

$$\sum_{i=0}^{5}(-1)^i \binom{6}{0}\binom{10}{5}\binom{5}{i}\lambda_{5+i} - \sum_{i=0}^{5}(-1)^i \binom{6}{1}\binom{6}{5}\binom{5}{i}\lambda_{5+i}$$

$$= \quad 216\lambda_5 - 1080\lambda_6 + 2160\lambda_7 - 2160\lambda_8 + 1080\lambda_9 - 216\lambda_{10}$$

- *<u>Για $l = 6$:</u>* $j = 0, 1$ και $i = 0, \ldots, 4$

$$\sum_{i=0}^{4}(-1)^i \binom{5}{0}\binom{10}{4}\binom{4}{i}\lambda_{6+i} - \sum_{i=0}^{4}(-1)^i \binom{5}{1}\binom{6}{4}\binom{4}{i}\lambda_{6+i}$$

$$= \quad 135\lambda_6 - 540\lambda_7 + 810\lambda_8 - 540\lambda_9 + 135\lambda_{10}$$

- *<u>Για $l = 7$:</u>* $j = 0, 1$ και $i = 0, \ldots, 3$

$$\sum_{i=0}^{3}(-1)^i \binom{4}{0}\binom{10}{3}\binom{3}{i}\lambda_{7+i} - \sum_{i=0}^{3}(-1)^i \binom{4}{1}\binom{6}{3}\binom{3}{i}\lambda_{7+i}$$

$$= \quad 40\lambda_7 - 120\lambda_8 + 120\lambda_9 - 40\lambda_{10}$$

- <u>*Για $l=8$:*</u> $j=0,1,2$ και $i=0,\dots,2$

$$\begin{aligned}
&\sum_{i=0}^{2}(-1)^i\binom{3}{0}\binom{10}{2}\binom{2}{i}\lambda_{8+i}-\sum_{i=0}^{2}(-1)^i\binom{3}{1}\binom{6}{2}\binom{2}{i}\lambda_{8+i}\\
&+\sum_{i=0}^{2}(-1)^i\binom{3}{2}\binom{2}{2}\binom{2}{i}\lambda_{8+i}\\
=\;& 3\lambda_8-6\lambda_9+3\lambda_{10}
\end{aligned}$$

- <u>*Για $l=9$:*</u> $j=0,1,2$ και $i=0,1$

$$\begin{aligned}
&\sum_{i=0}^{1}(-1)^i\binom{2}{0}\binom{10}{1}\binom{1}{i}\lambda_{9+i}-\sum_{i=0}^{1}(-1)^i\binom{2}{1}\binom{6}{1}\binom{1}{i}\lambda_{9+i}\\
&+\sum_{i=0}^{1}(-1)^i\binom{2}{2}\binom{2}{1}\binom{1}{i}\lambda_{9+i}\\
=\;& 0
\end{aligned}$$

- <u>*Για $l=10$:*</u> $j=0,1,2$ και $i=0$

$$\begin{aligned}
&\sum_{i=0}^{0}(-1)^i\binom{1}{0}\binom{10}{0}\binom{0}{0}\lambda_{10}-\sum_{i=0}^{0}(-1)^i\binom{1}{1}\binom{6}{0}\binom{0}{0}\lambda_{10}\\
&+\sum_{i=0}^{0}(-1)^i\binom{1}{2}\binom{2}{0}\binom{0}{0}\lambda_{10}\\
=\;& 0
\end{aligned}$$

Επομένως, προκύπτει ότι:

$$P(L_{10}<4)=\lambda_0-7\lambda_4+6\lambda_5+3\lambda_8-4\lambda_9+\lambda_{10}.$$

Μπορούμε να υπολογίσουμε τις τιμές των λ_i και να βρούμε μια τιμή για τη συνάρτηση κατανομής της L_n, για συγκεκριμένη ακολουθία ανταλλάξιμων τυχαίων μεταβλητών.

2.3. Pólya-Eggenberger μέθοδος

Περιγραφή

Έχουμε μια κάλπη η οποία αρχικά περιέχει w άσπρα και b μαύρα σφαιρίδια και κάθε φορά επιλέγεται ένα από αυτά με τυχαίο τρόπο. Στη συνέχεια το ξανατοποθετούμε μέσα στην κάλπη, προσθέτοντας s ακόμη σφαιρίδια του ίδιου χρώματος. Η διαδικασία αυτή επαναλαμβάνεται n φορές και έτσι δημιουργείται μια δυαδική ακολουθία ως εξής:

$$\xi_i = \begin{cases} 1, \text{αν η } i\text{-οστή επιλογή είναι άσπρο σφαιρίδιο} \\ \\ 0, \text{αν η } i\text{-οστή επιλογή είναι μαύρο σφαιρίδιο} \end{cases}$$

όπου τα ξ_i είναι ανταλλάξιμα με:

$$\begin{aligned} \lambda_n &= P(\xi_1 = \xi_2 = \ldots = \xi_n = 1) \\ &= \underbrace{\frac{w}{w+b}}_{1 \text{ επιλογή}} \underbrace{\frac{w+s}{w+b+s}}_{2 \text{ επιλογή}} \cdots \underbrace{\frac{w+(n-1)s}{w+b+(n-1)s}}_{n\text{-οστή επιλογή}} \\ &= \prod_{j=0}^{n-1} \frac{w+js}{w+b+js}. \end{aligned} \tag{20}$$

Ακόμη, μπορούμε να βρούμε την πιθανότητα μιας τέτοιας ακολουθίας που περιλαμβάνει y αποτυχίες και $n-y$ επιτυχίες.

$$\begin{aligned} p_n(y) &= P(y \text{ αποτυχίες και } n-y \text{ επιτυχίες}) \\ &= \underbrace{\frac{w}{w+b}\frac{w+s}{w+b+s}\cdots\frac{w+(n-y-1)s}{w+b+(n-y-1)s}}_{n-y \text{ επιτυχίες}} \\ &\times \underbrace{\frac{b}{w+b+(n-y)s}\frac{b+s}{w+b+(n-y-1)s}\cdots\frac{b+(y-1)s}{w+b+(n-1)s}}_{y \text{ αποτυχίες}}. \end{aligned}$$

Έτσι,

$$p_n(y) = \frac{\prod_{j=0}^{n-y-1}(w+js)\prod_{j=0}^{y-1}(b+js)}{\prod_{j=0}^{n-1}(w+b+js)}, \quad 0 \le y \le n.$$

Παρατηρούμε ότι η πιθανότητα $p_n(y)$ είναι ένα κλάσμα οι όροι του οποίου είναι γινόμενα. Αυτό σημαίνει ότι σε όποια σειρά και αν βάλω τους παράγοντες (δηλαδή τις αποτυχίες και τις επιτυχίες) το αποτέλεσμα θα είναι το ίδιο. Γι΄ αυτό το λόγο και στηριζόμενοι στον ορισμό των ανταλλάξιμων δοκιμών, ισχυριζόμαστε ότι τα ξ_i βρίσκονται κάτω από αυτήν τη μορφή εξάρτησης.

Παρατήρηση Στην περίπτωση όπου η παράμετρος s πάρει την τιμή 0, τότε από την ισότητα (20) έχουμε ότι:

$$\lambda_i \;=\; \lambda_1^i, i = 1, 2, \ldots, n \text{ και } \lambda_0 = 1$$

όπου,

$$\lambda_1 \;=\; \frac{w}{w+b} = p,$$

με p $(0 < p < 1)$, να είναι η κοινή πιθανότητα επιτυχίας για κάθε μία δοκιμή. Με άλλα λόγια, στη συγκεκριμένη περίπτωση $(s = 0)$, έχουμε ανεξάρτητες και ισόνομες δοκιμές *Bernoulli* και μάλιστα:

$$p_n(y) = p^{n-y}q^y.$$

Αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής L_n

Οι *Makri*, *Philippou* και *Psillakis* (2007α), με τη βοήθεια της συνδυαστικής ανάλυσης, έδωσαν έναν τύπο για την κατανομή της τυχαίας μεταβλητής $N_{n,k,s}$, η οποία μας δίνει τον αριθμό των ροών επιτυχιών (επιλογών λευκών σφαιριδίων) μεγέθους k σε n επιλογές σύμφωνα με τη δειγματοληπτική μέθοδο *Pólya – Eggenberger*. Με τη χρήση αυτής, μπορούμε να βρούμε την αθροιστική συνάρτηση κατανομής της L_n και πιο συγκεκριμένα παρατηρούμε ότι το γεγονός $\{L_n < k\}$ ταυτίζεται με το $\{N_{n,k,s} = 0\}$, διότι αν η μέγιστη ροή επιτυχιών έχει μήκος μικρότερο του k, τότε δεν θα υπάρχει καμία ροή επιτυχιών μήκους k και αντίστροφα.

Πριν περάσουμε στο θεώρημα για τη συνάρτηση πιθανότητας της μεταβλητής $N_{n,k,s}$ στο 0 και την απόδειξη αυτού, θα πρέπει να αναφέρουμε το παρακάτω λήμμα.

Λήμμα 2.3 *Έστω $C(\alpha; i, r-i; m-1, n-1)$ παριστάνει τον αριθμό των διανομών α μη διακεκριμένων σφαιριδίων σε r διακεκριμένα κελιά, εκ΄ των οποίων τα i έχουν χωρητικότητα $m-1$ και τα υπόλοιπα $r-i$ έχουν χωρητικότητα $n-1$. Τότε:*

$$C(\alpha; i, r-i; m-1, n-1)$$

$$= \sum_{j_1=0}^{\lfloor \alpha/m \rfloor} \sum_{j_2=0}^{\lfloor (\alpha - mj_1)/n \rfloor} (-1)^{j_1+j_2} \binom{i}{j_1} \binom{r-i}{j_2} \binom{\alpha - mj_1 - nj_2 + r - 1}{r-1},$$

όπου $\lfloor x \rfloor$ είναι ο μέγιστος ακέραιος που είναι μικρότερος από το x

Απόδειξη Ο παραπάνω τύπος προκύπτει παίρνοντας τη γεννήτρια συνάρτηση $(1-t^m)^i(1-t^n)^{r-1}(1-t)^{-r}$. Αναλύοντας τον κάθε ένα παράγοντα έχουμε ότι:

$$\begin{aligned} (1-t^m)^i &= \sum_{j_1=0}^{i} \binom{i}{j_1} (-t^m)^{j_1} \\ &= \sum_{j_3} \binom{i}{j_1} (-1)^{j_1} t^{mj_1}. \end{aligned} \tag{21}$$

Με παρόμοιο τρόπο έχουμε ότι:

$$\begin{aligned} (1-t^n)^{r-i} &= \sum_{j_2=0}^{r-i} \binom{r-i}{j_2} (-t^n)^{j_2} \\ &= \sum_{j_2} \binom{r-i}{j_2} (-1)^{j_2} t^{nj_2}. \end{aligned} \tag{22}$$

Και για τον τρίτο παράγοντα έχουμε:

$$(1-t)^{-r} = \sum_{j_3=0}^{\infty} \binom{r+j_3-1}{r-1} t^{j_3}.$$

Επειδη όμως ισχύει ότι $mj_1 + nj_2 + j_3 = \alpha$, τότε παρατηρούμε ότι ο συντελεστής του t^α είναι η ποσότητα:

$$\begin{aligned} &\sum_{j_1} \sum_{j_2, j_3} \binom{i}{j_1} (-1)^{j_1} \binom{r-i}{j_2} (-1)^{j_2} \binom{r+j_3-1}{r-1} \\ &= \sum_{j_1=0}^{\lfloor \alpha/m \rfloor} \sum_{j_2=0}^{\lfloor (\alpha - mj_1)/n \rfloor} (-1)^{j_1+j_2} \binom{i}{j_1} \binom{r-i}{j_2} \binom{\alpha - mj_1 - nj_2 + r - 1}{r-1}. \end{aligned}$$

◇

Πόρισμα *Αν $C(\alpha; i, r-i; m-1, n-1)$ είναι όπως ορίστηκε στο Λήμμα 2.3, τότε:*

$$\begin{aligned} C(\alpha; i, r-i; m-1, m-1) &= C(\alpha, r, m-1) \\ &= \sum_{j=0}^{\lfloor a/m \rfloor} (-1)^j \binom{r}{j} \binom{\alpha - mj + r - 1}{r-1}. \end{aligned}$$

Θεώρημα 2.3 *Έστω ότι έχουμε μια κάλπη που περιλαμβάνει w άσπρα και b μαύρα σφαιρίδια και έστω ότι επιλέγουμε τυχαία n από αυτά, κάθε φορά και από ένα σύμφωνα με την Pólya − Eggenberger μέθοδο. Αν $N_{n,k,s}$ και L_n παριστάνουν τον αριθμό των ροών επιτυχιών μήκους k και το μέγιστο μήκος ροής επιτυχιών αντίστοιχα, έχουμε ότι:*

$$\begin{aligned} P(L_n < k) &= P(N_{n,k,s} = 0) \\ &= \sum_{y=y_1}^{y_2} p_n(y) C(n-y, y+1, k-1), \end{aligned}$$

όπου $y_1 = \lfloor \frac{n}{k} \rfloor$ και $y_2 = n$.

Απόδειξη Παίρνουμε το γεγονός $\{N_{n,k,s} = 0, Y_n = y\}$ και παρατηρούμε ότι κάθε στοιχείο αυτού, είναι μια ακολουθία που περιλαμβάνει y αποτυχίες και $n - y$ επιτυχίες τοποθετημένες με τέτοιο τρόπο ώστε το μήκος των ροών επιτυχιών να μην ξεπερνάει την τιμή $k - 1$. Αφού έχουμε y αποτυχίες, θα έχουμε και $y + 1$ κελιά. Έτσι, ο αριθμός των διαφορετικών τρόπων που μπορούμε να τοποθετήσουμε τις $n - y$ επιτυχίες σε $y + 1$ κελιά έτσι ώστε σε κανένα να μην τοποθετηθούν περισσότερα από $k - 1$ είναι, σύμφωνα με το πόρισμα, $C(n - y, y + 1, k - 1)$. Επίσης η πιθανότητα καθεμιάς απο τις ακολουθίες είναι ίση με $p_n(y)$.

◇

Στη συνέχεια οι *Makri*, *Philippou* και *Psillakis* (2007b) με παρόμοιο σκεπτικό, αλλά με την χρήση πλέον της μεταβλητής $G_{n,k}$ έδωσαν μια έκφραση για την αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής L_n.

Λήμμα 2.4 *Ο αριθμός των διανομών των α μη διακεκριμένων σφαιριδίων σε r διακεκριμένα κελιά, όπου κάθε ένα από τα m, $0 \le m \le r$ να περιλαμβάνει το πολύ k σφαιρίδια, δίνεται από τον τύπο:*

$$H_m(a, r, k) = \sum_{j=0}^{\lfloor \alpha/(k+1) \rfloor} (-1)^j \binom{m}{j} \binom{\alpha - (k+1)j + r - 1}{a - (k+1)j}.$$

Απόδειξη Ο αριθμός $H_m(a, r, k)$ είναι ίσος με τον συντελεστή του t^α αν αναλύσουμε τη γεννήτρια συνάρτηση $g(t) = (1 - t^{k+1})^m (1 - t)^{-r}$.

◇

Χρησιμοποιώντας το Λήμμα 2.4, αποδείχθηκε μια μαθηματική έκφραση για τη συνάρτηση πθανότητας της μεταβλητής $G_{n,k}$, όπως φαίνεται στο παρακάτω θεώρημα.

Θεώρημα 2.4 *Αν $G_{n,k}$ είναι ο αριθμός των ροών επιτυχιών με μήκος τουλάχιστον k, τότε η συνάρτηση πιθανότητας δίνεται από τον τύπο:*

$$\begin{aligned} & P(G_{n,k} = x) \\ = & \sum_{y=0}^{n-kx} p_n(y) \binom{y+1}{x} H_{y+1-x}(n-y-kx, y+1, k-1), x = 0, 1, \ldots, \left\lfloor \frac{n+1}{k+1} \right\rfloor. \end{aligned}$$

Απόδειξη Αρχικά έχουμε y αποτυχίες που σημαίνει ότι δημιουργούνται $y+1$ κελιά. Επειδή όμως θέλουμε να βρούμε την πιθανότητα $P(G_{n,k} = x)$, ο αριθμός των διαφορετικών τρόπων βάσει των οποίων, τα x κελιά μπορούν να επιλεγούν από τα συνολικά $y+1$, είναι ίσος με $\binom{y+1}{x}$, $x = 0, 1, \ldots, \lfloor (n+1)/(k+1) \rfloor$. Εφόσον έχουμε τοποθετήσει k επιτυχίες στα x κελιά, μας απομένουν να τοποθετήσουμε $n-y-kx$ επιτυχίες. Αυτό θα πραγματοποιηθεί διανέμοντας τις επιτυχίες σε $y+1$ κελιά χωρίς όμως τα εναπομείναντα $y+1-x$ να περιέχουν πάνω από $k-1$. Ο αριθμός αυτός των διαφορετικών τρόπων διανομής σφαιριδίων δίνεται από το Λήμμα 3.4 με τον αριθμό $H_{y+1-x}(n-y-kx, y+1, k-1)$.

$\diamond$

Αν λάβουμε υπόψη τη γνωστή σχέση που συνδέει τις τυχαίες μεταβλητές $G_{n,k}$ και L_n, τότε μπορούμε να βρούμε έναν τύπο για την αθροιστική συνάρτηση κατανομής της δεύτερης.

Πόρισμα *Αν L_n είναι το μήκος της μεγαλύτερης ροής επιτυχιών σύμφωνα με την μέθοδο $Pólya - Eggenberger$, τότε εχουμε ότι:*

$$P(L_n < k) = \sum_{y=0}^{n} p_n(y) \sum_{i=0}^{\lfloor (n-y)/k \rfloor} (-1)^i \binom{y+1}{i} \binom{n-ki}{n-ki-y}.$$

Απόδειξη Επειδή θέλουμε η μέγιστη ροή επιτυχιών να μην ξεπερνά την τιμή k τότε δε θα υπάρχει καμία ροή επιτυχιών με μέγεθος τουλάχιστον k. Χρησιμοποιώντας το Θεώρημα 2.4 έχουμε για τις αντίστοιχες πιθανότητες των γεγονότων ότι:

$$P(L_n < k) \ = \ P(G_{n,k} = 0)$$

$$
\begin{aligned}
&= \sum_{y=0}^{n} p_n(y) \binom{y+1}{0} \sum_{j=0}^{\lfloor n-y \rfloor/k} (-1)^j \binom{y+1}{j} \binom{n-y-(k-1+1)j+y}{n-y-kj} \\
&= \sum_{y=0}^{n} p_n(y) \sum_{j=0}^{\lfloor n-y \rfloor/k} (-1)^j \binom{y+1}{j} \binom{n-kj}{n-y-kj}.
\end{aligned}
$$

$\diamond$

Κεφάλαιο 3

Μέθοδος Εμφύτευσης

3.1 Εισαγωγικές έννοιες

Έστω ότι έχουμε ένα χώρο καταστάσεων $\Omega = \{1, \ldots, n\}$, και $(Y_t) = \{Y_0, Y_1, \ldots, Y_t, \ldots\}$ είναι μια ακολουθία από τυχαίες μεταβλητές ορισμένες στο Ω. Θα ορίσουμε κάποιες έννοιες τις οποίες θα συναντήσουμε σε αυτό το κεφάλαιο.

- **Ορισμός** Η ακολουθία (Y_t) θα ονομάζεται πεπεραμένη Μαρκοβιανή αλυσίδα, αν για κάθε $\{Y_0 = i_0, Y_1 = i_1, \ldots, Y_{t-1} = i_{t-1}, Y_t = i_t\}$, $t = 1, 2, \ldots$ έχουμε:

  $$P(Y_t = i_t | Y_{t-1} = i_{t-1}, \ldots, Y_0 = i_0) = P(Y_t = i_t | Y_{t-1} = i_{t-1}),$$

  δηλαδή η ακολουθία είναι Μαρκοβιανή αλυσίδα, αν η κατάσταση της θέσης i_t εξαρτάται μόνο από αυτή της αμέσως προηγούμενης, i_{t-1}.

- Οι *πιθανότητες μετάβασης* του συστήματος τη χρονική στιγμή t ορίζονται ως:

  $$p_{i,j}(t) = P(Y_t = j | Y_{t-1} = i), \; i, j \in \Omega.$$

  Οι $p_{i,j}$ με $1 \leq i, j \leq n$ αναπαρίστανται με την χρήση ενός $n \times n$ πίνακα, τον λεγόμενο πίνακα μετάβασης, ως εξής:

  $$M = \begin{pmatrix} p_{11}(t) & p_{12}(t) & \cdots & p_{1n}(t) \\ p_{21}(t) & p_{22}(t) & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1}(t) & p_{n2}(t) & \cdots & p_{nn}(t) \end{pmatrix}_{n \times n}.$$

- **Ορισμός** Μια Μαρκοβιανή αλυσίδα $(Y_0, Y_1, \ldots)$ είναι ομογενής αν οι πιθανότητες μετάβασης είναι σταθερές στο χρόνο. Ισχύει δηλαδή ότι $P(Y_t = j | Y_{t-1} = i) = p_{i,j}$ για $i, j \in \Omega$ και για όλα τα $t = 1, 2, \ldots$

- **Ορισμός** Μια κατάσταση i λέγεται *απορροφητική* αν η πιθανότητα μετάβασης (ενός βήματος) p_{ii} ισούται με 1.

Ύστερα από μια μικρή εισαγωγή στη θεωρία των Μαρκοβιανών αλυσίδων, μπορούμε να περάσουμε στην περιγραφή της μεθόδου εμφύτευσης, μιας τυχαίας μεταβλητής, σε Μαρκοβιανή αλυσίδα.

Για οποιοδήποτε n ορίζουμε ως Γ_n=$\{0, 1, \ldots, n\}$ ένα σύνολο δεικτών και ως $\Omega = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ ένα πεπερασμένο σύνολο καταστάσεων.

Ορισμός Μια μη μηδενική ακέραια τυχαία μεταβλητή $X_{n,k}$ εμφυτεύεται σε Μαρκοβιανή αλυσίδα όταν:

1. Υπάρχει μία πεπερασμένη Μαρκοβιανή αλυσίδα $(Y_t; t \in \Gamma_n)$ ορισμένη στο χώρο καταστάσεων Ω.

2. Υπάρχει μια πεπεραμένη διαμέριση $\{C_x; x = 0, 1, \ldots, l\}$ του χώρου καταστάσεων Ω.

3. Για κάθε $x = 0, 1, \ldots, l$ ισχύει $P(X_{n,k} = x) = P(Y_n \in C_x)$.

Έχουμε επίσης ότι:

- Λ_t είναι ο $m \times m$ πίνακας μετάβασης της πεπερασμένης Μαρκοβιανής αλυσίδας $(\{Y_t; t \in \Gamma_n\}, \Omega)$.

- U_r είναι ένα $1 \times m$ μοναδιαίο διάνυσμα έχοντας μονάδα στο $r-$οστό στοιχείο και 0 οπουδήποτε αλλού.

- U_r' είναι ο ανάστροφος πίνακας του U_r, διάστασης $m \times 1$.

- Για κάθε μία διαμέριση C_x ορίζουμε ένα διάνυσμα διάστασης $1 \times m$ τέτοιο ώστε: $U(C_x) = \sum_{r:\alpha_r \in C_x} U_r$.

Πριν δώσουμε το θεώρημα το οποίο μας δίνει τη συνάρτηση πιθανότητας της τυχαίας μεταβλητής $X_{n,k}$, καλό θα ήταν να αναφέρουμε τις εξισώσεις $Chapman-Kolmogorov$ γιατί θα είναι χρήσιμες στην απόδειξη του θεωρήματος. Δίνουμε το ακόλουθο λήμμα.

Λήμμα 3.1 *Έστω μια ομογενής Μαρκοβιανή αλυσίδα* $\{Y_n; n \in N\}$ *η οποία έχει πιθανότητες μετάβασης n βημάτων:*

$$p_{ij}^{(n)} = P(Y_{n+m} = j|Y_m = i) = P(Y_n = j|Y_0 = i)$$

και χώρο καταστάσεων $S = \{0, 1, \ldots\}$. *Τότε ισχύει ότι:*

(α) $p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}$, $i, j \in S$ και $m, n \in N$.

(β) Αν $P^{(1)} = P = (p_{ij})$ και $P^{(n)} = (p_{ij}^{(n)})$, $i, j \in S$ τότε $P^{(n)} = P^n = P \cdot P \cdots P$, όπου

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0s} \\ p_{10} & p_{11} & \cdots & p_{1s} \\ \vdots & \vdots & \vdots & \cdots \\ p_{s0} & p_{s1} & \cdots & p_{ss} \end{pmatrix}.$$

(γ) Αν $\alpha_k = P(Y_0 = k)$, $k \in S$ τότε $P(Y_n = j) = \sum_{k \in S} \alpha_k p_{kj}^{(n)}$, $\forall j \in S$.

Το ακόλουθο θεώρημα έχει δοθεί από τους Fu και $Koutras$ (1994) για τη συνάρτηση πιθανότητας μιας τυχαίας μεταβλητής σύμφωνα με τη μέθοδο εμφύτευσης.

Θεώρημα 3.1 *Αν η τυχαία μεταβλητή* $X_{n,k}$ *μπορεί να εμφυτευθεί σε πεπερασμένη Μαρκοβιανή αλυσίδα τότε:*

$$P(X_{n,k} = x) = \pi_0 \left(\prod_{t=1}^{n} \Lambda_t \right) U'(C_x),$$

όπου $\pi_0 = (P(Y_0 = \alpha_1), P(Y_0 = \alpha_2), \ldots, P(Y_0 = \alpha_m))$ *είναι ο πίνακας πιθανοτήτων των αρχικών καταστάσεων της Μαρκοβιανής αλυσίδας, διάστασης* $(1 \times m)$.

Απόδειξη Λόγω του ότι η μεταβλητή είναι εμφυτευμένη σε Μαρκοβιανή αλυσίδα, σύμφωνα με τον ορισμό θα έχουμε ότι:

$$P(X_{n,k} = x) = P(Y_n \in C_x) = \sum_{\alpha_r \in C_x} P(Y_n = \alpha_r) \tag{23}$$

Εφαρμόζοντας την ισότητα των $Chapman - Kolmogorov$ έχουμε ότι:

$$P(Y_n=\alpha_r)=\sum_{\alpha_j\in\Omega}P(Y_n=\alpha_r|Y_0=\alpha_j)P(Y_0=\alpha_j)$$

$$\begin{aligned}
&= (P(Y_0=\alpha_1),\ldots,P(Y_0=\alpha_m))\\
&\times\begin{pmatrix} P(Y_n=\alpha_1|Y_0=\alpha_1) & \cdots & P(Y_n=\alpha_r|Y_0=\alpha_1) & \cdots\\ P(Y_n=\alpha_1|Y_0=\alpha_2) & \cdots & P(Y_n=\alpha_r|Y_0=\alpha_2) & \cdots\\ \vdots & \cdots & \vdots & \cdots\\ P(Y_n=\alpha_1|Y_0=\alpha_m) & \cdots & P(Y_n=\alpha_r|Y_0=\alpha_m) & \cdots \end{pmatrix}\begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix}\\
&= (\ \cdots,\ \ P(Y_n=\alpha_r|Y_0=\alpha_1)P(Y_0=\alpha_1)+\ \ \cdots+\ \ P(Y_n=\alpha_r|Y_0=\alpha_m)P(Y_0=\alpha_m),\ \ \cdots\ )\\
&\times\begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix}\\
&= P(Y_n=\alpha_r|Y_0=\alpha_1)P(Y_0=\alpha_1)+\ldots+P(Y_n=\alpha_r|Y_0=\alpha_m)P(Y_0=\alpha_m)\\
&= \pi_0\left(\prod_{t=1}^{n}\Lambda_t\right)U'_r
\end{aligned}$$

και έτσι η σχέση (23) μπορεί να γραφεί ως:

$$\begin{aligned}
P(X_{n,k}=x) &= \sum_{\alpha_r\in C_x}\pi_0\left(\prod_{t=1}^{n}\Lambda_t\right)U'_r\\
&= \pi_0\left(\prod_{t=1}^{n}\Lambda_t\right)\sum_{\alpha_r\in C_x}U'_r\\
&= \pi_0\left(\prod_{t=1}^{n}\Lambda_t\right)U'(C_x),\quad \text{για κάθε } x=0,\ldots,l.
\end{aligned}$$

◇

Παρατήρηση Εάν η Μαρκοβιανή αλυσίδα είναι ομογενής (ανεξάρτητες και ισόνομες δοκιμές *Bernoulli*) τότε Λ_t=Λ και για κάθε $t\in\Gamma_n$,

$$\prod_{t=1}^{n}\Lambda_t(X)=\Lambda^n$$

και κατ΄ επέκταση

$$P(X_{n,k}=x)=\pi_0\Lambda^nU'(C_x),\quad x=0,\ldots,l.$$

3.2 Συνάρτηση πιθανότητας της τυχαίας μεταβλητής L_n

Όπως έχουμε αναφέρει και σε προηγούμενες ενότητες, πολλές φορές υπολογίζουμε έμμεσα την κατανομή της τυχαίας μεταβλητής L_n, χρησιμοποιώντας άλλοτε τη μεταβλητή $N_{n,k}$ και άλλοτε την $G_{n,k}$. Αυτήν την τακτική θα ακολουθήσουμε και σ΄ αυτή την παράγραφο, βρίσκοντας αρχικά τη συνάρτηση πιθανότητας της πρώτης.

Πριν περάσουμε στη διαδικασία εύρεσης της κατανομής θα πρέπει να αναφέρουμε τα εξής:

- Αν Ω είναι ο χώρος καταστάσεων, τότε κάθε στοιχείο αυτού συμβολίζεται με το ζεύγος (x, i), όπου x παριστάνει τον αριθμό των ροών επιτυχιών και το i σχετίζεται με τον αριθμό των συνεχόμενων επιτυχιών στην ακολουθία τη δεδομένη χρονική στιγμή.

- Με βάση τον παραπάνω συμβολισμό οι πιθανότητες μετάβασης ορίζονται ως:

$$p_t(y, j; x, i) = P(Y_t = (y, j)|Y_{t-1} = (x, i)).$$

- Για $i = 0$ έχουμε ότι $Y_t = (x, 0)$, που παριστάνει το γεγονός ότι στην ακολουθία μέχρι την $t-$οστή δοκιμή έχουμε x εμφανίσεις από ροές επιτυχιών και ότι η τελευταία δοκιμή είναι αποτυχία ή ότι αρχίζουμε να μετράμε για την επόμενη ροή. Γενικά, αν το σύστημα βρίσκεται στην κατάσταση (x, i) μετά την $(t-1)$ δοκιμή και η t δοκιμή είναι αποτυχία τότε το σύστημα μπαίνει στην κατάσταση $(x, 0)$. Ο λόγος είναι ότι κάθε φορά που συμβαίνει μία αποτυχία τότε αρχίζουμε να μετράμε την επόμενη ροή από την αρχή.

- Με βάση το παραπάνω σκεπτικό ισχύει επίσης ότι:

$$p_t(x, 0; x, i) = q_t.$$

- Τέλος, θα θεωρούμε ότι $P(Y_0 = (0, 0)) = 1$ και ότι η τελευταία κατάσταση της Μαρκοβιανής αλυσίδας είναι απορροφητική γι΄ αυτό και η τελευταία γραμμή στον πίνακα μετάβασης είναι η $(0,\ 0,\ \ldots,\ 0,\ 1)$.

Ας εφαρμόσουμε τα παραπάνω για την εύρεση της συνάρτησης πιθανότητας της τ.μ. $N_{n,k}$:

α) Χώρος καταστάσεων

Θεωρούμε, τον χώρο καταστάσεων:

$$\Omega = \Omega(N_{n,k}) = \{(x,i) : x = 0,1,\ldots,l \text{ και } i = 0,1,\ldots,k-1\},$$

το l είναι η μεγαλύτερη τιμή που μπορεί να πάρει το x και ισούται με $\lfloor \frac{n}{k} \rfloor$, καθώς και την πεπερασμένη Μαρκοβιανή αλυσίδα $\{Y_t : t \in \Gamma_n\}$ ορισμένη πάνω στο Ω. Για κάθε μία από τις ακολουθίες μήκους t,

$$S\ F\ F\ S\ \ldots\ F\underbrace{S\ S\ldots\ S}_{m},$$

m είναι ο αριθμός των συνεχόμενων επιτυχιών μετρώντας προς τα πίσω που βρίσκονται στο τέλος της ακολουθίας ($m = 0$ αν η $t-$οστή δοκιμή είναι αποτυχία είτε είναι θετικό πολλαπλάσιο του k άρα, μπορώ να πάρω μία ακόμη ροή επιτυχιών). Έτσι, ορίζουμε τη Μαρκοβιανή αλυσίδα ως $Y_t = (x,i)$ όπου το x μας δίνει τον αριθμό των συνεχόμενων k επιτυχιών και $m = i\ mod(k)$.

β) Διαμερίσεις

Για δεδομένο x ορίζουμε τα υποσύνολα $C_x = \{(x,i) : i = 0,1,\ldots,k-1\}$ και $x = 0,1,\ldots,l = \lfloor n/k \rfloor$ που είναι στοιχεία των διαμερίσεων του αρχικού συνόλου καταστάσεων $\Omega(N_{n,k})$ και $P(N_{n,k} = x) = P(Y_n \in C_x)$.

γ) Πίνακας μετάβασης

Η δημιουργία του πίνακα μετάβασης βασίζεται στις πιθανότητες μετάβασης κατά την μετακίνησή μας από την δοκιμή $t-1$ στην t. Έτσι έχουμε ότι:

$$P_{(x,i),(y,j)} = P\left(Y_t = (y,j) | Y_{t-1} = (x,i)\right)$$

- $= q$, αν $y = x$ και $j = 0$, για $i = 0,1,\ldots,k-1$. Δηλαδή αν στη θέση t δεν αλλάζει ο αριθμός των ροών επιτυχιών και επίσης έχουμε $j = 0$ τότε έχουμε αποτυχία στο τέλος.
- $= p$, αν $y = x$ και $j = i+1$, για $i = 0,1,\ldots,k-2$. Αφού ο αριθμός των ροών επιτυχιών δεν αλλάζει ενώ ο αριθμός των επιτυχιών αυξάνεται κατά ένα τότε συναντάμε επιτυχία στο τέλος.
- $= p$, αν $y = x+1$ και $j = 0$, για $i = k-1$ και $x = 0,1,\ldots,l-1$.
- $= 1$, αν $y = x = l$ και $j = i = k-1$.

- 0, διαφορετικά.

Στηριζόμενοι στους παραπάνω κανόνες προκύπτει ο πίνακας μετάβασης για την μεταβλητή $N_{n,k}$, η γενική μορφή του οποίου είναι:

$$\Lambda_t(N) = \begin{bmatrix} A_t & B_t & & & & \\ & A_t & B_t & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & A_t & B_t \\ & & & & & A_t^\star \end{bmatrix}$$

όπου ο πίνακας B_t είναι διάστασης $k \times k$ και έχει επιτυχία στη θέση $(k,1)$ και 0 οπουδήποτε αλλού, ο A_t είναι πίνακας διάστασης $k \times k$ της μορφής:

$$A_t = \begin{bmatrix} q_t & p_t & 0 & 0 & . & . & . \\ q_t & 0 & p_t & 0 & . & . & . \\ q_t & 0 & 0 & p_t & . & . & . \\ . & . & . & . & . & . & \\ q_t & . & . & . & . & & \end{bmatrix}$$

και ο πίνακας $A_t^\star$ είναι της ίδιας μορφής με τον A_t, μόνο που στην τελευταία γραμμή έχει το διάνυσμα

$$\begin{pmatrix} 0, & . & . & . & 0, & 1 \end{pmatrix}.$$

Παράδειγμα 3.1 Σαν εφαρμογή όλων των παραπάνω, μπορούμε να χρησιμοποιήσουμε την τ.μ. $N_{5,2}$ και πιο συγκεκιμένα:

α) Χώρος καταστάσεων

$\Omega(N_{5,2}) = \{(0,0),\ (0,1),\ (1,0),\ (1,1),\ (2,0),\ (2,1)\}$

β) Διαμερίσεις

$C_0 = \{(0,0),\ (0,1)\},\ C_1 = \{(1,0),\ (1,1)\},\ C_2 = \{(2,0),\ (2,1)\}$

<u>Πίνακας μετάβασης</u>

$$\begin{array}{c} \\ (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \\ (2,0) \\ (2,1) \end{array}
\begin{array}{c} \begin{array}{cccccc} (0,0) & (0,1) & (1,0) & (1,1) & (2,0) & (2,1) \end{array} \\
\begin{bmatrix} q & p & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 & 0 \\ 0 & 0 & q & p & 0 & 0 \\ 0 & 0 & q & 0 & p & 0 \\ 0 & 0 & 0 & 0 & q & p \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Αν έχουμε μια συγκεκριμένη ακολουθία, όπως φαίνεται παρακάτω:

$$F\ S\ F\ F\ S\ S\ S\ F\ S\ S$$

και θέσουμε όπου k την τιμή 2, τότε έχουμε ότι:

$$Y=(x,i)\longrightarrow Y_1=(0,0)\ ,Y_2=(0,1)\ ,Y_3=(0,0)\ ,Y_4=(0,0)$$

$$Y_5=(0,1),\ Y_6=(1,0),\ Y_7=(1,1),\ Y_8=(1,0),\ Y_9=(1,1),\ Y_{10}=(2,0)$$

Αφού ικανοποιήσαμε όλες τις συνθήκες για να μπορέσει η τ.μ. $N_{n,k}$ να εμφυτευθεί σε Μαρκοβιανή αλυσίδα, έχουμε τη δυνατότητα να εφαρμόσουμε σε αυτή το Θεώρημα 4.1 της προηγούμενης παραγράφου, δηλαδή:

$$P(N_{n,k}=x)=\pi_0\left(\prod_{t=1}^{n}\Lambda_t(N)\right)U'(C_x),\quad x=0,\ldots,l.$$

Το αποτέλεσμα αυτής της ενέργειας είναι η εύρεση της συνάρτησης πιθανότητας της τ.μ. L_n.

Θεώρημα 3.2 *Για δεδομένη τιμή του n έχουμε ότι:*

$$\begin{aligned} P(L_n=k) &= \pi_{0,k+1}\left(\prod_{t=1}^{n}\Lambda_t(k+1)\right)U'(k+1) \\ &\quad -\pi_{0,k}\left(\prod_{t=1}^{n}\Lambda_t(k)\right)U'(k),\ k=0,\ldots,n, \end{aligned}$$

όπου $\pi_{0,j}=(1,0,\ldots,0)$, $U(j)=(1,1,\ldots,1,0)(j=k,\ k+1)$ είναι $1\times(j+1)$ διανύσματα και Λ_t, όπως έχει οριστεί παραπάνω.

Απόδειξη Μέσω της γνωστής σχέσης που συνδέει τις παραπάνω τυχαίες μεταβλητές έχουμε ότι :

$$P(L_n < k) = P(N_{n,k} = 0)$$

για κάθε $k = 0, 1, \ldots, n$. Επίσης ισχύει ότι:

$$\begin{aligned} P(L_n = k) &= P(L_n < k+1) - P(L_n < k) \\ &= P(N_{n,k+1} = 0) - P(N_{n,k} = 0) \end{aligned}$$

και έτσι προκύπτει το ζητούμενο.

$\diamond$

Παράδειγμα 3.2 Θα υπολογίσουμε τη συνάρτηση πιθανότητας της τυχαίας μεταβλητής L_n, με την μέθοδο εμφύτευσης, έχοντας $n = 5$ δοκιμές *Bernoulli* διατεταγμένες σε γραμμή καθώς και $k = 2$.
Όπως έχουμε αναφέρει, για να μπορέσουμε να βρούμε την κατανομή μιας τυχαίας μεταβλητής, η οποία εμφυτεύεται σε Μαρκοβιανή αλυσίδα, πρέπει να κατασκευάσουμε:

1. κατάλληλο χώρο καταστάσεων,

2. μία διαμέριση του χώρου καταστάσεων,

3. τον αντίστοιχο πίνακα μετάβασης.

Έχουμε μια ακολουθία από 5, X_1, X_2, X_3, X_4, X_5, δοκιμές *Bernoulli* με πιθανότητα επιτυχίας $p_t = P(X_t = 1)$ και πιθανότητα αποτυχίας $q_t = P(X_t = 0)$.
Για τις συγκεκριμένες τιμές των n και k, έχουμε ότι: $l = \lfloor 5/2 \rfloor = 2$, $x = 0, 1$ και $i = 0, 1$. Έτσι, ο χώρος καταστάσεων είναι : $\Omega = \{(0,0), (0,1), (1,0), (1,1), (2,0), (2,1)\}$ και η κατάλληλη διαμέριση του Ω είναι: $C_0 = \{(0,0), (0,1)\}, C_1 = \{(1,0), (1,1)\}, C_2 = \{(2,0), (2,1)\}$.

Για να επιτύχουμε το στόχο μας θα πρέπει να χρησιμοποιήσουμε τον τύπο του Θεωρήματος 4.2 για $n = 5$ και $k = 2$ και πιο συγκεκριμένα:

$$\begin{aligned} P(L_5 = 2) &= \pi_{0,3} \left(\prod_{t=1}^{5} \Lambda_t(3) \right) U'(3) \\ &\quad - \pi_{0,2} \left(\prod_{t=1}^{5} \Lambda_t(2) \right) U'(2), \ k = 0, \ldots, 5. \end{aligned}$$

Ουσιαστικά καλούμαστε να υπολογίσουμε τις πιθανότητες $P(N_{5,2} = 0)$ και $P(N_{5,3} = 0)$. Για την πρώτη έχουμε:

- $\pi_0 = (P(Y_0 = \alpha_0), P(Y_1 = \alpha_1), P(Y_1 = \alpha_2)) = (1, 0, 0)$.

- $U'(C_0) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

- Ο πίνακας μετάβασης:

$$\Lambda_t = \begin{bmatrix} q_t & p_t & 0 \\ q_t & 0 & p_t \\ 0 & 0 & 1 \end{bmatrix}.$$

Από το θεώρημα των Fu και $Koutras$ (1994) προκύπτει ότι:

$$P(N_{5,2} = 0) = \pi_0 \left(\prod_{t=1}^{5} \Lambda_t \right) U'(C_0)$$

Ας υπολογίσουμε πρώτα το $\prod_{t=1}^{5} \Lambda_t$.

$$\begin{aligned} \prod_{t=1}^{5} \Lambda_t &= \begin{bmatrix} q_t & p_t & 0 \\ q_t & 0 & p_t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_t & p_t & 0 \\ q_t & 0 & p_t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_t & p_t & 0 \\ q_t & 0 & p_t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_t & p_t & 0 \\ q_t & 0 & p_t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_t & p_t & 0 \\ q_t & 0 & p_t \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \alpha'_{11} & \alpha'_{12} & \alpha'_{13} \\ \alpha'_{21} & \alpha'_{22} & \alpha'_{23} \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

Στη συνέχεια,

$$\begin{aligned} & P(N_{5,2} = 0) = (1, 0, 0) \\ \times & \begin{bmatrix} \alpha'_{11} & \alpha'_{12} & \alpha'_{13} \\ \alpha'_{21} & \alpha'_{22} & \alpha'_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\ = & \begin{bmatrix} \alpha'_{11} & \alpha'_{12} & \alpha'_{13} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\ = & \alpha'_{11} + \alpha'_{12}, \end{aligned}$$

όπου:
$\alpha'_{11} = q_2q_4q_5 + q_2q_3p_4q_5 + q_1p_2q_3q_5$

$\alpha'_{12} = q_2q_3q_4p_5 + q_2p_3q_4p_5 + p_1q_2q_3q_4p_5 + q_1p_2q_3q_4p_5$
$\alpha'_{13} = q_2q_3q_4p_5 + q_2p_3p_4 + q_1p_2q_3p_4p_5 + q_1p_2p_3 + p_1p_2$
$\alpha'_{21} = q_1q_3q_5 + q_1q_2p_3q_4q_5$
$\alpha'_{22} = q_1q_4q_5(q_3 + q_2p_3)$
$\alpha'_{23} = q_1q_3p_4p_5 + q_1q_2p_3p_4 + q_1p_2p_3 + p_1$
$\alpha'_{31} = 0$
$\alpha'_{32} = 0$
$\alpha'_{33} = 1$

Με όμοιο τρόπο βρίσκουμε την πιθανότητα $P(N_{5,3} = 0)$:

- $\pi_0 = (1, 0, 0, 0)$
- $U'(C_0) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$
- Ο πίνακας μετάβασης:

$$\Lambda_t = \begin{bmatrix} q_t & p_t & 0 & 0 \\ q_t & 0 & p_t & 0 \\ 0 & 0 & q_t & p_t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Έτσι,

$$\prod_{t=1}^{5} \Lambda_t = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ \alpha_9 & \alpha_{10} & \alpha_{11} & \alpha_{12} \\ \alpha_{13} & \alpha_{14} & \alpha_{15} & \alpha_{16} \end{bmatrix}$$

Κάνοντας μια αντικατάσταση προκύπτει ότι:

$$\begin{aligned} P(N_{5,3} = 0) &= (1, 0, 0, 0) \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ \alpha_9 & \alpha_{10} & \alpha_{11} & \alpha_{12} \\ \alpha_{13} & \alpha_{14} & \alpha_{15} & \alpha_{16} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \\ &= (\alpha_1, \alpha_2, \alpha_3, \alpha_4) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \\ &= \alpha_1 + \alpha_2 + \alpha_3 \end{aligned}$$

όπου:

$$\begin{aligned}
\alpha_1 &= q_2q_5(q_4+q_3p_4)+2q_1p_2q_3q_4q_5\\
\alpha_2 &= q_4q_5(q_2+q_1p_2q_3)\\
\alpha_3 &= q_3p_4p_5(q_2+q_1p_2)+p_2q_6(p_3p_4+q_1p_3q_4+p_1q_3q_4)\\
\alpha_4 &= q_2p_3p_4p_5+p_2q_4p_5(q_1p_3+p_1q_3)+p_2p_4(q_1p_3+p_1q_3)+p_1p_2p_3\\
\alpha_5 &= q_1q_3q_5(q_4+q_2q_4+p_4)\\
\alpha_6 &= q_1q_3q_4p_5(1+q_2)\\
\alpha_7 &= q_1q_3p_4(p_5+q_2q_5)+q_4q_5(q_1p_2p_3+p_1q_2q_3)\\
\alpha_8 &= q_1p_5(q_2q_3p_4+p_2p_3q_4)+p_1q_2q_3(q_4p_5+p_4)+p_3(q_1p_2p_4+p_1q_2)+p_1p_2\\
\alpha_9 &= 0\\
\alpha_{10} &= 0\\
\alpha_{11} &= q_1q_2q_3q_4q_5\\
\alpha_{12} &= q_1q_2q_3(q_4p_5+p_4)+q_1p_2(q_3+1)+p_1
\end{aligned}$$

Τέλος με μια απλή αντικατάσταση στον τύπο

$$P(L_5=2)=P(N_{5,3}=0)-P(N_{5,2}=0),$$

προκύπτει το ζητούμενο.

Όμως, μπορούμε να αποκτήσουμε άμεσα τη συνάρτηση πιθανότητας της μεταβλητής L_n, κατασκευάζοντας τον κατάλληλο πίνακα μετάβασης και εφαρμόζοντας το θεώρημα των *Fu* και *Koutras* (1994).
Αν πάρουμε, για παράδειγμα, 3 δοκιμές *Bernoulli*, τότε οι δυνατές τιμές του x είναι $x=0,1,2,3$

Ο χώρος καταστάσεων:
$\Omega=\{(0,0),(1,-1),(1,0),(1,1),(2,-1),(2,0),(2,1),(3,-1)\}$.

Μια διαμέριση αυτού:
$C_0=\{(0,0)\},C_1=\{(1,-1),(1,0),(1,1)\},C_2=\{(2,-1),(2,0),(2,1)\},C_3=\{(3,-1)\}$.

και τέλος ο πίνακας μετάβασης:

$$\Lambda_t(L) = \begin{bmatrix} q_t & p_t & 0 & 0 & & & & & \\ & 0 & q_t & 0 & p_t & & & & \\ & 0 & q_t & p_t & 0 & & & & \\ & 0 & q_t & 0 & p_t & & & & \\ & & & & 0 & q_t & 0 & 0 & p_t \\ & & & & 0 & q_t & p_t & 0 & 0 \\ & & & & 0 & q_t & 0 & p_t & 0 \\ & & & & 0 & q_t & 0 & 0 & p_t \\ & & & & & & & & 1 \end{bmatrix}$$

Ας σημειώσουμε ότι η παράμετρος i παίρνει την τιμή -1 στην περίπτωση όπου το $m \geq k$.

3.3 Δεσμευμένη συνάρτηση κατανομής της L_n δοθέντος του αριθμού των επιτυχιών στην ακολουθία

Χρησιμοποιώντας τη σχέση που συνδέει τις μεταβλητές L_n και $N_{n,k}$ μπορούμε να βρούμε την $P(L_n < k|S_n = n_1), k = 0, \ldots, n_1$. Αν η μεταβλητή S_n μας δίνει το συνολικό αριθμό των επιτυχιών σε μια ακολουθία έχουμε ότι:

$$P(L_n < k|S_n = n_1) = P(N_{n,k} = 0|S_n = n_1)$$

και έτσι θα στρέψουμε το ενδιαφέρον μας στην εύρεση της δεσμευμένης πιθανότητας της $N_{n,k}$ δοθέντος του αριθμού των επιτυχιών στην ακολουθία. Πριν αναφέρουμε το αντίστοιχο θεώρημα, θα πρέπει να ορίσουμε τα εξής:

Θεωρούμε μια μη ομογενή Μαρκοβιανή αλυσίδα δύο καταστάσεων (Y_t). Έστω ότι στα n πειράματα έχουμε n_1 επιτυχίες.

1. Θέτουμε $Y_t = (x, i)$, όπου x είναι ο αριθμός των επιτυχιών στις t πρώτες δοκιμές και i είναι ο αριθμός των συνεχόμενων επιτυχιών που βρίσκονται στο τέλος των πρώτων t δοκιμών μετρώντας προς τα πίσω.
   Από τα παραπάνω προκύπτει ότι:
   - $i = h, h = 0, \ldots, k-1$ αν υπάρχουν h συνεχόμενες επιτυχίες μετρώντας προς τα πίσω από την t δοκιμή και $N_{n,k} = 0$.
   - $i = S$ αν $N_{n,k} \geq 1$ και η t-οστή δοκιμή είναι S και $i = F$ αν η $t-$οστή δοκιμή είναι F.
   - $Y_t = \Delta$ αν $Y_j = (n_1, i), j = n_1+1, \ldots, t-1$, Δ είναι η απορροφητική κατάσταση.

2. Το χώρο καταστάσεων ως: $\Omega = (0,0) \cup \{(x,i) : x = 1,\ldots,n_1; i = 0,1,\ldots,k-1,S,F\} \cup \Delta$, όπου $i \leq x$.

3. Τις διαμερίσεις που δημιουργούνται από: $C_0 = \{(0,0)\}, C_\Delta = \{\Delta\}, C_{n_1,SF} = \{(n_1,S),(n_1,F)\}, C_{n_1,k}\{(n_1,i)\}, C_{x,SF} = \{(x,S),(x,F)\}, C_{x,k} = \{(x,i)\}, x = 1,\ldots,n_1-1, i = 0,1,\ldots,k-1$ και $i \leq x$. Το πλήθος των στοιχείων του Ω είναι $m = \frac{k(k+1)}{2} + (n_1-k+1)(k+2)+1$.

Χρησιμοποιώντας τους παραπάνω συμβολισμούς, η *Lou* (1996) έδωσε μια μαθηματική έκφραση για την κατανομή της τ.μ. $N_{n,k}$.

Θεώρημα 3.3 *Θεωρούμε μια ακολουθία από μη ομογενείς Μαρκοβιανά εξαρτημένες δοκιμές Bernoulli. Για συγκεκριμένες τιμές των n, n_1 και $k \leq n_1$, το τυχαίο διάνυσμα $(S_n, N_{n,k})$ μπορεί να εμφυτευθεί σε Μαρκοβιανή αλυσίδα, με πίνακα μετάβασης $\Lambda(t)$. Γνωρίζοντας την αρχική κατάσταση $P(Y_0 = (0,0) = 1)$, ισχύουν τα παρακάτω:*

- Για $n_1 = 0, P(S_n = 0, N_{n,k} = 0) = 1$ και για $1 = k \leq n_1 \leq n, P(S_n = 0, N_{n,k} = 0) = 0$.

- Για $1 < k \leq n_1 \leq n$,

$$P(S_n = n_1, N_{n,k} = 0) = \pi_0 \left(\prod_{t=1}^{n} \Lambda_t\right) U(C_{n_1,k}),$$

  όπου $\pi_0 = (1,0,\ldots,0)$ είναι διάνυσμα διάστασης $1 \times m$ και $C_{n_1,k} = \{(n_1,i); i = 0,1,\ldots,k-1\}$,

- Για $1 < k \leq n_1 \leq n$,

$$P(N_{n,k} = 0 | S_n = n_1) = \frac{\pi_0 \left(\prod_{t=1}^{n} \Lambda_t\right) U(C_{n_1,k})}{\pi_0 \left(\prod_{t=1}^{n} \Lambda_t\right) U(C_{n_1})},$$

όπου $C_{n_1} = C_{n_1,k} \cup C_{n_1,SF}$.

Το επόμενο πόρισμα προκύπτει άμεσα από το Θεώρημα 4.3 και αναφέρεται στη δεσμευμένη συνάρτηση κατανομής της τυχαίας μεταβλητής L_n.

Πόρισμα *Έχοντας συγκεκριμένες τιμές για τις ποσότητες n, n_1, k, $(1 < k \leq n_1 \leq n)$, η δεσμευμένη συνάρτηση κατανομής της L_n δοθέντος του αριθμού των επιτυχιών $S_n = n_1$* είναι:

$$P(L_n < k | S_n = n_1) = \frac{\pi_0 \left(\prod_{t=1}^{n} \Lambda_t\right) U'(C_{n_1,k})}{\pi_0 \left(\prod_{t=1}^{n} \Lambda_t\right) U'(C_{n_1})}.$$

Παράδειγμα 3.3 Έστω ότι έχουμε μια ακολουθία απο $n = 5$ δοκιμές *Bernoulli* με αριθμό επιτυχιών $n_1 = 2$ και $k = 2$. Τότε:

Ο χώρος καταστάσεων είναι:
$\Omega = \{(0,0),(1,0),(1,1),(2,0),(2,1),(2,S),(2,F),\Delta\}$.

Οι διαμερίσεις: $C_2 = \{(2,i) : i = 0,1,S,F\}$ και $C_{2,2} = \{(2,0),(2,1)\}$.

Ακολουθώντας τους κανόνες για την κατασκευή του πίνακα μετάβασης, έχουμε για τις συγκεκριμένες τιμές των n, n_1, k και για $t = 1,2,\ldots,5$

$$
\begin{array}{cc}
 & \begin{array}{cccccccc} (0,0) & (1,0) & (1,1) & (2,0) & (2,1) & (2,S) & (2,F) & \Delta \end{array} \\
\begin{array}{c} (0,0) \\ (1,0) \\ (1,1) \\ (2,0) \\ (2,1) \\ (2,S) \\ (2,F) \\ \Delta \end{array} &
\begin{bmatrix}
P_{FF}(t) & 0 & P_{FS}(t) & 0 & 0 & 0 & 0 & 0 \\
0 & P_{FF}(t) & 0 & 0 & P_{FS}(t) & 0 & 0 & 0 \\
0 & P_{SF}(t) & 0 & 0 & 0 & P_{SS}(t) & 0 & 0 \\
0 & 0 & 0 & P_{FF}(t) & 0 & 0 & 0 & P_{FS}(t) \\
0 & 0 & 0 & P_{SF}(t) & 0 & 0 & 0 & P_{SS}(t) \\
0 & 0 & 0 & 0 & 0 & 0 & P_{SF}(t) & P_{SS}(t) \\
0 & 0 & 0 & 0 & 0 & 0 & P_{FF}(t) & P_{FS}(t) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Επίσης, μπορούμε να καταγράψουμε όλες τις πιθανές ακολουθίες που δημιουργούνται μεταθέτοντας κάθε φορά τις 2 επιτυχίες, και να παρατηρήσουμε σε ποιες περιπτώσεις έχουμε μέγιστο μήκος ροής επιτυχιών μικρότερο του 2. Έτσι,

$\{SSFFF\}$, $\{FFFSS\}$, $\{FSSFF\}$, $\{FFSSF\}$, $\underline{\{SFSFF\}}$, $\underline{\{FSFFS\}}$
$\underline{\{SFFSF\}}$ $\underline{\{SFFFS\}}$, $\underline{\{FSFSF\}}$,$\underline{\{FFSFS\}}$

Στη συνέχεια ακολουθώντας μία άλλη μέθοδο ο *Eryilmaz* (2005α) έδωσε ένα τύπο για την κατανομή της L_n καθώς επίσης και κάποια φράγματα για τη συνάρτηση κατανομής $P(L_n < k)$, στην περίπτωση που έχουμε Μαρκοβιανά εξαρτημένες δοκιμές.

3.4 Μη ομογενείς Μαρκοβιανά εξαρτημένες μεταβλητές

Έστω ότι έχουμε μια Μαρκοβιανή αλυσίδα δύο καταστάσεων $\{Y_n\}$ και συμβολίζουμε με η_n το μήκος της ροής επιτυχιών που βρίσκεται στο τέλος τον n-οστού βήματος. Αποδεικνύεται ότι η η_n είναι μη ομογενής Μαρκοβιανή αλυσίδα με πιθανότητες μετάβασης:

$$P(\eta_n = i|\eta_{n-1} = i-1) = P(Y_n = 1|Y_{n-1} = 1) = p_{11}^{(n)},\ 2 \le i \le n;\ n \ge 2$$

$$P(\eta_n = 0|\eta_{n-1} = i) = P(Y_n = 0|Y_{n-1} = 1) = p_{10}^{(n)},\ 1 \le i \le n;\ n \ge 2$$

$$P(\eta_n = 0|\eta_{n-1} = 0) = P(Y_n = 0|Y_{n-1} = 0) = p_{00}^{(n)},\quad n \ge 2$$

$$P(\eta_n = 1|\eta_{n-1} = 0) = P(Y_n = 1|Y_{n-1} = 0) = p_{01}^{(n)},\quad n \ge 2.$$

Στο παρακάτω λήμμα βρίσκουμε την κατανομή της η_n.

Λήμμα 3.2 *Για την η_n, $n \ge 1$, ισχύει ότι:*

$$P(\eta_n = k) = \begin{cases} \alpha(n), & k = 0 \\ p_{01}^{(n-k+1)}\alpha(n-k)\prod_{i=1}^{k-1} p_{11}^{(n-i+1)}, & k = 1, 2, \ldots, n-1 \\ p_1 \prod_{i=1}^{n-1} p_{11}^{(n-i+1)}, & k = n \end{cases}$$

όπου $\prod_{\Phi} = 1, \alpha(j) = P(Y_j = 0)$ και

$\alpha(j) = (p_{00}^{(j)} - p_{10}^{(j)})a(j-1) + p_{10}^{(j)},\ j \ge 2,\ \alpha(1) = p_0.$

Απόδειξη Για $k = 0$ έχουμε από τον ορισμό της η_n ότι:

$$P(\eta_n = 0) = P(Y_n = 0) = \alpha_n.$$

Για $k = 1, 2, \ldots, n-1$:

$$\begin{aligned} P(\eta_n = k) &= P(Y_n = 1, Y_{n-1} = 1, \ldots, Y_{n-k+1} = 1, Y_{n-k} = 0) \\ &= P(Y_{n-k})P(Y_{n-k+1} = 1|Y_{n-k} = 0) \\ &\quad \times P(Y_{n-k+2}|Y_{n-k+1}, Y_{n-k} = 0) \cdots P(Y_n = 1|Y_{n-1} = 1, \ldots, Y_{n-k} = 0) \end{aligned}$$

και λόγω Μαρκοβιανής εξάρτησης

$$= P(Y_{n-k})P(Y_{n-k+1} = 1|Y_{n-k} = 0)P(Y_{n-k+2} = 1|Y_{n-k+1} = 1)$$

$$
\begin{aligned}
&\cdots P(Y_n = 1 | Y_{n-1} = 1) \\
= \quad & \alpha(n-k) p_{01}^{(n-k+1)} \prod_{i=1}^{k-1} p_{11}^{(n-i+1)}.
\end{aligned}
$$

<u>Για $k = n$</u> :

$$P(\eta_n = n) = P(Y_n = 1, Y_{n-1} = 1, \ldots, Y_1 = 1)$$

και χρησιμοποιώντας, όπως στην προηγούμενη περίπτωση, το πολλαπλασιαστικό θεώρημα προκύπτει ότι :

$$P(\eta_n = n) = p_1 \prod_{i=1}^{n-1} p_{11}^{(n-i+1)}.$$

$\diamond$

Πριν περάσουμε στο θεώρημα που μας δίνει την κατανομή της τυχαίας μεταβλητής L_n συναρτήσει της η_n, θα πρέπει να δούμε το παρακάτω λήμμα το οποίο θα μας βοηθήσει στην απόδειξη του πρώτου.

Λήμμα 3.3 *Για $n \geq 2$ ισχύει :*

$$P(\eta_{n-1} < k, \eta_n < k)$$

$$
= \begin{cases}
p_{00}^{(n)} P(\eta_{n-1} = 0), & k = 1 \\
\sum_{j=0}^{k-2} P(\eta_{n-1} = j) + p_{10}^{(n)} P(\eta_{n-1} = k-1), & k = 2, 3, \ldots, n-1 \\
1 - p_1 \prod_{j=1}^{n-1} p_{11}^{(n-j+1)}, & k = n
\end{cases}
$$

Απόδειξη <u>Για $k = 1$</u>:

$$
\begin{aligned}
P(\eta_{n-1} < k, \eta_n < k) &= P(\eta_{n-1} = 0, \eta_n = 0) \\
&= P(\eta_n = 0 | \eta_{n-1} = 0) P(\eta_{n-1} = 0) \\
&= p_{00}^{(n)} P(\eta_{n-1} = 0).
\end{aligned}
$$

Για $k = 2, 3, \ldots, n-1$:

$$\begin{aligned}
& P(\eta_{n-1} < k, \eta_n < k) \\
= \ & P(\eta_{n-1} = 0, \eta_n = 0) + \sum_{j=1}^{k-1} P(\eta_{n-1} = j, \eta_n = 0) \\
& + P(\eta_{n-1} = 0, \eta_n = 1) + \sum_{j=2}^{k-1} P(\eta_{n-1} = j-1, \eta_n = j) \\
= \ & P(\eta_{n-1} = 0) + \sum_{j=1}^{k-2} P(\eta_{n-1} = j) + p_{10}^{(n)} P(\eta_{n-1} = k-1).
\end{aligned}$$

Για $k = n$:

$$\begin{aligned}
P(\eta_{n-1} < k, \eta_n < k) &= P(\eta_n < n) \\
&= 1 - P(\eta_n \geq n) \\
&= 1 - P(\eta_n = n) \\
&= 1 - p_1 \prod_{j=1}^{n-1} p_{11}^{(n-j+1)}
\end{aligned}$$

από το προηγούμενο λήμμα.

$\diamond$

Θεώρημα 3.3 *Για την κατανομή της τυχαίας μεταβλητής L_n ισχύει ότι:*

Για $k = 1, 2, \ldots, n-1$:

$$\begin{aligned}
& P(L_n < k) \\
= \ & \sum_{i_k=0}^{k-1} \sum_{i_{k+1}=0}^{k-1} \cdots \sum_{i_n=0}^{k-1} \left\{ P(\eta_k = i_k) \prod_{m=k+1}^{n} P(\eta_m = i_m | \eta_{m-1} = i_{m-1}) \right\}.
\end{aligned}$$

Για $k = n$

$$P(L_n < k) = 1 - p_1 \prod_{i=1}^{n-1} p_{11}^{(n-i+1)}.$$

Απόδειξη Εκφράζοντας, με τη βοήθεια των ορισμών, την τυχαία μεταβλητή L_n συναρτήσει των η_i, έχουμε για την πιθανότητα του γεγονότος $\{L_n < k\}$ και για $k = 1, 2, \ldots, n-1$ ότι:

$$\begin{aligned}
P(L_n < k) &= P\left\{\max_{k\leq i\leq n} \eta_i < k\right\} \\
&= P(\eta_k < k, \eta_{k+1} < k, \ldots, \eta_n < k) \\
&= \sum_{i_k=0}^{k-1}\sum_{i_{k+1}=0}^{k-1}\ldots\sum_{i_n=0}^{k-1} P(\eta_k = i_k, \eta_{k+1} = i_{k+1}, \ldots, \eta_n = i_n) \\
&= \sum_{i_k=0}^{k-1}\sum_{i_{k+1}=0}^{k-1}\ldots\sum_{i_n=0}^{k-1} P(\eta_n = i_n|\eta_{n-1} = i_{n-1})P(\eta_{n-1} = i_{n-1}|\eta_{n-2} = i_{n-2}) \\
&\quad \times \ldots P(\eta_{k+1} = i_{k+1}|\eta_k = i_k)P(\eta_k = i_k)
\end{aligned}$$

και για $k = n$

$$\begin{aligned}
P(L_n < n) &= 1 - P(L_n \geq k) \\
&= 1 - P(L_n = n) \\
&= 1 - P(Y_n = 1, Y_{n-1} = 1, \ldots, Y_1 = 1) \\
&= 1 - p_1 \prod_{i=1}^{n-1} p_{11}^{(n-i+1)}.
\end{aligned}$$

◇

Λόγω της πολυπλοκότητας του υπολογισμού της ακριβούς κατανομής της τυχαίας μεταβλητής L_n για μεγάλες τιμές του n, ο *Eryilmaz* (2005α) έδωσε φράγματα για την πιθανότητα $P(L_n < k)$. Οι πιθανότητες $\theta_1(i,k) := P(\eta_i \geq k)$ και $\theta_2(i,k) := P(\eta_i \geq k, \eta_{i+1} \geq k)$ αποτελούν τη βάση αυτών. Πιο συγκεκριμένα έχουμε το παρακάτω θεώρημα.

Θεώρημα 3.4 *Για $1 < k < n$ ισχύουν οι παρακάτω ανισότητες για τη συνάρτηση κατανομής της L_n,*

$$max\left(0, 1 - \sum_{i=k}^{n}\theta_1(i,k) + \sum_{i=k}^{n-1}\theta_2(i,k)\right) \leq P(L_n < k) \leq \min_{k\leq i\leq n}(1 - \theta_1(i,k)).$$

Απόδειξη Αν θεωρήσουμε ως $A_i \equiv \{\eta_i < k\}$ τότε ακολουθώντας την τακτική του Θεωρήματος 3.3, έχουμε ότι:

$$P(L_n < k) = P\left\{\bigcap_{i=k}^{n} A_i\right\}.$$

Κάτω φράγμα

Χρησιμοποιώντας την παραλλαγή $Worsley$ της ανισότητας $Bonferroni$, έχουμε ότι:

$$\begin{aligned} P\left\{\bigcap_{i=k}^{n} A_i\right\} &\geq 1 - \sum_{i=k}^{n} P(A_i^c) + \sum_{i=k}^{n-1} P(A_i^c A_{i+1}^c) \\ &= 1 - \sum_{i=k}^{n} \theta_1(i,k) + \sum_{i=k}^{n-1} \theta_2(i,k). \end{aligned}$$

Άνω φράγμα

Μπορούμε εύκολα να δούμε ότι:

$$\begin{aligned} P\left\{\bigcap_{i=k}^{n} A_i\right\} \leq & \quad \min_{k \leq i \leq n} (P\{A_i\}) \\ = & \quad \min_{k \leq i \leq n} (1 - \theta_1(i,k)). \end{aligned}$$

$\diamond$

Κεφάλαιο 4

Συνεχόμενα k-από-τα-n:F συστήματα

4.1 Εισαγωγικές έννοιες

Αφού έχουμε μελετήσει την τυχαία μεταβλητή L_n και τις διάφορες μεθόδους με τις οποίες μπορούμε να βρούμε την κατανομή της, στο κεφάλαιο αυτό θα επικεντρωθούμε στις εφαρμογές της στην αξιοπιστία συστημάτων και πιο συγκεκριμένα στα συνεχόμενα k-από-τα-n:F συστήματα (συστήματα αποτυχίας).

Ορισμός Υποθέτουμε ότι έχουμε n συνιστώσες οι οποίες είναι γραμμικά συνδεδεμένες με τέτοιο τρόπο, ώστε το σύστημα να αποτυγχάνει αν και μόνο αν τουλάχιστον k συνεχόμενες συνιστώσες αποτυγχάνουν. Ένα τέτοιο σύστημα ονομάζεται γραμμικό συνεχόμενο k-από-n:F σύστημα.

Ορισμός Αν $X_1, X_2, \ldots, X_n$ είναι δυαδικές (0-1) τυχαίες μεταβλητές που παριστάνουν τις καταστάσεις των συνιστωσών ενός συστήματος n συνιστωσών, ορίζεται η δυαδική τυχαία μεταβήτή $\phi(\tilde{X}) = \phi(X_1, \ldots, X_n)$, η οποία παριστάνει την κατάσταση του συστήματος, $\phi(\tilde{X}) = 1$, αν το σύστημα λειτουργεί ; 0 αν δε λειτουργεί.

Ορισμός Αξιοπιστία ενός συστήματος ονομάζεται η πιθανότητα $P\left[\phi(\tilde{X}) = 1\right]$.

Την αξιοπιστία ενός συνεχόμενου k-από-τα-n:F συστήματος θα την συμβολίζουμε ως $R_L(k, n)$.

Για παράδειγμα, αν έχουμε ένα γραμμικό συνεχόμενο 3-από-τα-6:F σύστημα, τότε λέμε ότι αν ο αριθμός των συνεχόμενων αποτυχιών είναι μικρότερος του 3, τότε το σύστημα λειτουργεί, ενώ σε αντίθετη περίπτωση αποτυγχάνει.

Ο κύριος σκοπός αυτού του κεφαλαίου είναι η σύνδεση της αξιοπιστίας ενός τέτοιου συστήματος με την κατανομή της τυχαίας μεταβλητής L_n. Είναι σαφές ότι:

$$R_L(k,n) = P(L_n < k-1).$$

Βέβαια, λόγω της σχέσης των μεταβλητών L_n και $N_{n,k}$, μπορούμε να γράψουμε ισοδύναμα την ισότητα,

$$R_L(k,n) = P(N_{n,k} = 0)$$

Συμπεραίνουμε ότι, όταν θέλουμε να υπολογίσουμε την αξιοπιστία ενός k-από-τα-n:F συστήματος, υπολογίζουμε τη συνάρτηση κατανομής της L_n.

Πολλοί ερευνητές από τις αρχές του 1980 έχουν ασχοληθεί με τα συνεχόμενα k-από-τα-n:F συστήματα. Στις μέρες μας τα διάφορα μηχανικά συστήματα όπως υπολογιστές, αεροπλανοφόρα, αυτοκίνητα, απαιτείται να έχουν υψηλή αξιοπιστία. Γι΄ αυτό η εκτίμηση της αξιοπιστίας είναι πολύ σημαντική στην κατασκευή, το σχεδιασμό και τη λειτουργία των μηχανικών συστημάτων. Για να δούμε μια πιο σαφή εικόνα για τη χρήση αυτών, καλό θα ήταν να αναφερθούν δύο παραδείγματα, όπως έχουν γραφτεί από τους *Chiang* και *Niu* (1981).

Παράδειγμα Έστω ότι έχουμε μια ακολουθία από n σταθμούς μικροκυμάτων οι οποίοι μεταφέρουν πληροφορίες από ένα μέρος A σε ένα μέρος B. Οι σταθμοί μικροκυμάτων είναι τοποθετημένοι μεταξύ των χώρων A και B σε ίσες αποστάσεις. Καθένας σταθμός μπορεί να μεταφέρει πληροφορίες σε απόσταση η οποία περιλαμβάνει k σταθμούς μικροκυμάτων. Είναι προφανές ότι ένα τέτοιο σύστημα αποτυγχάνει εάν και μόνο εάν τουλάχιστον k συνεχόμενοι σταθμοί μικροκυμάτων δεν λειτουργούν. Οι αξιοπιστίες των σταθμών ίσως να είναι διαφορετικές επειδή είναι διαφορετικές και οι περιβαλλοντικές συνθήκες. Έτσι οι διαδικασίες λειτουργίας ή όχι των σταθμών θεωρούνται στατιστικά ανεξάρτητες, αλλά όχι κατ΄ ανάγκην ισόνομες.

Παράδειγμα Έχουμε ένα σύστημα που μεταφέρει πετρέλαιο με σωλήνες από ένα μέρος A σε ένα μέρος B και υπάρχουν n σταθμοί τροφοδότησης. Οι σταθμοί τροφοδότησης είναι ισοδύναμα τοποθετημένοι μεταξύ των χώρων A και B. Κάθε σταθμός τροφοδότησης μπορεί να μεταφέρει το πετρέλαιο σε μια απόσταση ίση με k σταθμούς τροφοδότησης. Εάν ένας σταθμός τροφοδότησης χαλάσει, η ροή του πετρελαίου δεν θα διακοπεί, επειδή οι επόμενοι $k-1$ σταθμοί μπορούν να μεταφέρουν το φορτίο. ωστόσο, όταν χαλάσουν τουλάχιστον k συνεχόμενοι σταθμοί τροφοδότησης, η ροή του πετρελαίου σταματά και το σύστημα αποτυγχάνει να εκπληρώσει το σκοπό για τον οποίο έχει κατασκευαστεί. Εδώ οι σταθμοί δεν λειτουργούν ανεξάρτητα ο ένας από τον άλλον γιατί όταν χαλάσει ένας σταθμός το φορτίο του θα περάσει στο γειτονικό του οπότε το φορτίο του γειτονικού του σταθμού θα αυξηθεί.

4.2 Αξιοπιστία ενός συστήματος με ανεξάρτητες και ισόνομες συνιστώσες

Πριν παραθέσουμε διάφορους τύπους για τον υπολογισμό της αξιοπιστίας ενός συνεχόμενου k-από-τα-n:F συστήματος, θα πρέπει να αντιστοιχίσουμε διάφορα σύμβολα με τις έννοιές τους, για να γίνει ευκολότερη η κατανόηση της παραγράφου. Πιο αναλυτικά,

- n: ο αριθμός των συνιστωσών ενός συστήματος
- k: ο μικρότερος αριθμός των διαδοχικών συνιστωσών που λειτουργούν (αποτυγχάνουν) ενός συστήματος που λειτουργεί (αποτυγχάνει)
- p: η αξιοπιστία μιας συνιστώσας σε ένα σύστημα με ανεξάρτητες και ισόνομες συνιστώσες
- q: η πιθανότητα αποτυχίας μιας συνιστώσας σε ένα σύστημα με ανεξάρτητες και ισόνομες συνιστώσες, $q = 1 - p$
- p_i: η αξιοπιστία της i συνιστώσας, $i = 1, 2, \ldots, n$
- q_i: η πιθανότητα αποτυχίας της i συνιστώσας, $i = 1, 2, \ldots, n$
- $Lin/Con/k/n : F$: γραμμικό συνεχόμενο k-από-τα-n σύστημα αποτυχίας
- $R_L(k, n)$: η αξιοπιστία ενός $Lin/Con/k/n : F$ συστήματος
- $Q_L(k, n)$: η πιθανότητα αποτυχίας ενός $Lin/Con/k/n : F$ συστήματος, $Q_L(k, n) = 1 - R_L(k, n)$.

Οι *Chiang* και *Niu* (1981) έδωσαν έναν κλειστό τύπο για τον υπολογισμό της αξιοπιστίας ενός $Lin/Con/2/n : F$ συστήματος,

$$R_L(2, n) = \sum_{j=0}^{n} N(j, 2, n) q^j p^{n-j}$$

όπου,

$$N(j, 2, n) = \begin{cases} 0, & \lfloor \frac{n+1}{2} \rfloor < j \leq n, \\ \\ \begin{pmatrix} n - j + 1 \\ j \end{pmatrix}, & 0 \leq j \leq \lfloor \frac{n+1}{2} \rfloor. \end{cases}$$

Η διαδικασία απόδειξης έχει ως εξής:

Αν συμβολίσουμε με j τον αριθμό των συνιστωσών που αποτυγχάνουν, τότε έχουμε τις εξής περιπτώσεις:

- Έστω $j > \lfloor \frac{n+1}{2} \rfloor$.
Τότε θα υπάρχει σίγουρα ένα ζευγάρι από αποτυχημένες συνιστώσες το οποίο βρίσκεται δίπλα δίπλα και αυτό έχει σαν αποτέλεσμα το σύστημά μας να μην λειτουργεί και κατ΄ επέκταση να ισχύει $N(j, 2, n) = 0$

- Έστω $j \leq \lfloor \frac{n+1}{2} \rfloor$.
Σε αυτή την περίπτωση το σύστημά μας μπορεί να λειτουργήσει και για να συμβεί αυτό θα πρέπει να υπάρχει το λιγότερο μία επιτυχημένη συνιστώσα μεταξύ οποιωνδήποτε δύο αποτυχημένων. Αν έχουμε συνολικά n συνιστώσες εκ των οποίων οι j δεν λειτουργούν, τότε οι επιτυχημένες $n - j$ χωρίζουν την ακολουθία σε $n - j + 1$ κελιά. Κάθε ένα κελί όμως θέλουμε να περιλαμβάνει το πολύ μία αποτυχημένη συνιστώσα, γιατί αλλιώς το σύστημά μας θα αποτύχει. Εφόσον κάποια κελιά θα έχουν μία αποτυχία και κάποια δε θα έχουν καμία και εφόσον οι αποτυχίες είναι j τότε από τα συνολικά $n - j + 1$ κελιά, τα j θα έχουν μία αποτυχημένη συνιστώσα. Έτσι λοιπόν ο αριθμός των τρόπων που μπορώ να διαλέξω j κελιά από τα $n - j + 1$ είναι $\binom{n-j+1}{j}$.

Όταν έχουμε $2 \leq k \leq n - 1$ τότε ο τύπος γίνεται:

$$R_L(k, n) = \sum_{j=0}^{M} N(j, k, n) p^{n-j} q^j,$$

όπου M συμβολίζουμε το συνολικό αριθμό των αποτυχημένων συνιστωσών που πρέπει να υπάρχουν σε ένα $Lin/Con/k/n : F$ σύστημα, έτσι ώστε αυτό να λειτουργεί και δίνεται από την σχέση:

$$M = \begin{cases} n - \frac{n+1}{k} + 1, & \text{όταν } n+1 \text{ είναι πολλαπλάσιο του } k \\ \\ n - \lfloor \frac{n+1}{k} \rfloor, & \text{όταν } n+1 \text{ δεν είναι πολλαπλάσιο του } k \end{cases}$$

Άλλοι ερευνητές κατάφεραν να δώσουν αναδρομικές εκφράσεις για την $N(n, k, j)$, όπως για παράδειγμα ο *Derman et al.* (1974) η σκέψη του οποίου διατυπώνεται στο παρακάτω λήμμα.

Λήμμα 4.1 *Αν με $C(j,m,r)$ συμβολίσουμε τον αριθμό των τρόπων με τους οποίους μπορούμε να τοποθετήσουμε j σφαίρες σε r ξεχωριστά κελιά με τον περιοριμό ότι το πολύ m σφαίρες θα πρέπει να περιέχονται σε κάθε ένα κελί, έχουμε ότι:*

$$
\begin{aligned}
C(j,1,r) &= \begin{cases} \binom{r}{j}, & 1 \leq j \leq r \\ \\ 0, & j > r \end{cases} \\
C(j,2,r) &= \sum_{i=\max(j-r,0)}^{\min(\lfloor j/2 \rfloor, r)} \binom{r}{i} \binom{r-i}{j-2i} \\
C(j,m,r) &= \sum_{i=0}^{r} \binom{r}{i} C(j-mi, m-1, r-i), \quad m \geq 3.
\end{aligned}
$$

Η απόδειξη είναι όμοια με αυτή του Λήμματος 1.5.

◇

Χρησιμοποιώντας το Λήμμα 4.1, προκύπτει ο παρακάτω αναδρομικός τύπος για τον υπολογισμό του αριθμού $N(j,k,n)$:

$$
\begin{aligned}
N(j,k,n) &= C(j,k-1,n-j+1) \\
&= \sum_{i=0}^{n-j+1} \binom{n-j+1}{i} N(j-i(k-1), k-2, n-j-i+1), \\
& \qquad k > 3.
\end{aligned}
$$

Μία άλλη προσέγγιση για τον υπολογισμό του $N(j,k,n)$ είναι ο αναδρομικός τύπος που αποδείχθηκε από τους *Bollinger* και *Salvia* (1982) και πιο συγκεκριμένα έχουμε ότι:

$$
N(j,k,n) = \begin{cases} 0, & j = n \geq k \\ \\ \binom{n}{j}, & 0 \leq j \leq k-1 \\ \\ \sum_{i=1}^{k} N(j-i+1, k, n-i), & k \leq j < n \end{cases}
$$

Εκτός όμως από τους αναδρομικούς τύπους, οι *Lambiris* και *Papastavridis*

(1985) καθώς και ο *Goulden* (1987) έδωσαν ένα κλειστό τύπο για τον υπολογισμό της ποσότητας $N(j,k,n)$ χρησιμοποιώντας τη γεννήτρια συνάρτηση

$$g(z) = (1 + z + \ldots + z^{k-1})^{n-j+1},$$

η οποία χρησιμοποιήθηκε στην απόδειξη του Λήμματος 1.5.

Επομένως, για την ποσότητα $N(j,k,n)$ ισχύει η ισότητα:

$$N(j,k,n) = \sum_{i=0}^{\lfloor j/k \rfloor} (-1)^i \binom{n-j+1}{i} \binom{n-ik}{n-j},$$

σημειώνοντας ότι το i δεν μπορεί να ξεπεράσει την ποσότητα $\lfloor j/k \rfloor$. Τέλος, επιπρόσθετοι τύποι για τον υπολογισμό της αξιοπιστίας ενός τέτοιου συστήματος, αποδείχθηκαν από τους *Goulden* (1987) και *Hwang* (1986) και είναι:

$$\begin{aligned} R_L(k,n) &= \sum_{i=0}^{\lfloor n/(k+1) \rfloor} (-1)^i p^i q^{ki} \left[\binom{n-ki}{i} - q^k \binom{n-k(i+1)}{i} \right], \\ R_L(k,n) &= \sum_{i=0}^{\lfloor (n+1)/(k+1) \rfloor} (-1)^i p^{i-1} q^{ki} \left[\binom{n-ki+1}{i} - q \binom{n-ki}{i} \right]. \end{aligned}$$

Επίσης, στη συγκεκριμένη περίπτωση όπου $n \leq 2k$, έχουμε τον κλειστό τύπο:

$$R_L(k,n) = \begin{cases} 1, & 0 \leq n < k \\ 1 - q^k - (n-k)pq^k, & k \leq n \leq 2k. \end{cases}$$

4.3 Αξιοπιστία συστημάτων με ανεξάρτητες συνιστώσες

Ο τύπος που αποδείχθηκε από τους *Chiang* και *Niu* (1981) για την αξιοπιστία ενός $Lin/Con/k/n : F$ συστήματος με ανεξάρτητες και ισόνομες συνιστώσες, μπορεί να εφαρμοστεί και στην περίπτωση όπου αυτές δεν είναι αναγκαία και ισόνομες. Συμβολίζουμε με L τον δείκτη της πρώτης αποτυχίας και με M το δείκτη της πρώτης επιτυχίας μετά από τον L. Για παράδειγμα, αν σε ένα σύστημα η πρώτη συνιστώσα λειτουργεί και η δεύτερη αποτυγχάνει, τότε θα έχουμε $L = 2$.

Με βάση τα παραπάνω παρατηρούμε ότι, το σύστημα αποτυγχάνει σίγουρα όταν $M - L \geq k$ ενώ λειτουργεί όταν $M - L < k$ και το υποσύστημα που περιλαμβάνει τις συνιστώσες $M+1, M+2, \ldots, n$ λειτουργεί. Έτσι, ένας αναδρομικός τύπος που μας δίνει την αξιοπιστία ενός $Lin/Con/k/n : F$ συστήματος είναι:

$$\begin{aligned}
R_L(k,n) &= \sum_l \sum_m P(\text{το σύστημα λειτουργεί}|L=l, M=m)P(L=l, M=m) \\
&= \sum_{l=1}^{n-k+1} \sum_{m=l+1}^{l+k-1} R_L(k,(m+1,n))p_m \left(\prod_{i=1}^{l-1} p_i\right) \left(\prod_{j=l}^{m-1} q_j\right) + \prod_{i=1}^{n-k+1} p_i,
\end{aligned}$$

όπου $\prod_{i=a}^{b} \equiv 1$ αν $b < a$, $R_L(k,n) = 0$ για $n < 0$, και $R_L(k,n) = 1$ για $0 \leq n < k$.

Μια άλλη αναδρομική σχέση έδωσε ο *Hwang* (1982) χρησιμοποιώντας το γεγονός E_i : {η συνιστώσα i είναι η τελευταία που λειτουργεί}. Δεδομένου ότι τα γεγονότα E_i είναι ξένα μεταξύ τους, για να λειτουργήσει το σύστημα θα πρέπει να συμβαίνει ένα από τα γεγονότα $E_{n-k+1}, \ldots, E_n$ και έτσι:

$$\begin{aligned}
R_L(k,n) &= \sum_{i=n-k+1}^{n} P(E_i)R_L(k,i-1) \\
&= \sum_{i=n-k+1}^{n} R_L(k,i-1)p_i \prod_{j=i+1}^{n} q_j,
\end{aligned}$$

με αρχική συνθήκη ότι $R_L(k,n) = 1$ για $n < k$, (όταν οι συνιστώσες του συστήματος είναι λιγότερες από αυτές που απαιτούνται για να μη λειτουργεί το σύστημα οπότε είναι λογικό να λειτουργεί πάντα) και $\prod_{i=a}^{b} x_i \equiv 1$ για $a > b$.

Ο *Hwang* (1982) έδωσε μια καλύτερη προσέγγιση δουλεύοντας με το γεγονός F_i : {το σύστημα αποτυγχάνει για πρώτη φορά στην i-οστή συνιστώσα}. Προκύπτει ότι, ο δείκτης i εκφράζει τον μικρότερο ακέραιο έτσι ώστε k συνεχόμενες συνιστώσες $\{i-k+1, i-k+2, \ldots, i-1, i\}$ να αποτυγχάνουν. Αν $i > k$ τότε το F_i μας δείχνει ότι η συνιστώσα $i-k$ λειτουργεί και το υποσύστημα με συνιστώσες $(1, \ldots, i-k-1)$ πρέπει να λειτουργεί.

Λόγω του ότι τα F_i είναι ξένα μεταξύ τους και ότι για να αποτύχει το σύστημα θα πρέπει να συμβαίνει ένα από τα γεγονότα έχουμε:

$$\begin{aligned}
Q_L(k,n) &= \sum_{i=k}^{n} R_L(k,i-k-1)p_{i-k} \prod_{j=i-k+1}^{i} q_j \\
&= Q_L(k,n-1) + R_L(k,n-k-1)p_{n-k} \prod_{j=n-k+1}^{n} q_j,
\end{aligned}$$

με αρχική συνθήκη $Q_L(k,n) = 0$ για $n < k$ και $p_0 \equiv 1$.

Επειδή όμως έχουμε ορίσει ως $Q_L(k,n) = 1 - R_L(k,n)$, προκύπτει άμεσα από την προηγούμενη ισότητα η παρακάτω αναδρομική σχέση για την αξιοπιστία:

$$R_L(k,n) = R_L(k,n-1) - R_L(k,n-k-1)p_{n-k} \prod_{j=n-k+1}^{n} q_j.$$

Τέλος, οι *Zuo* και *Kuo* (1990) απέδειξαν έναν κλειστό τύπο για την περίπτωση όπου $k \leq 2n$:

$$R_L(k,n) = \begin{cases} 1, & 0 \leq n < k \\ \\ 1 - \sum_{i=1}^{n-k+1} \left(p_{i+k} \prod_{j=i}^{i+k-1} q_j \right), & p_{n+1} \equiv 1; k \leq n \leq 2k. \end{cases}$$

4.4 Αξιοπιστία συστήματος με την μέθοδο εμφύτευσης σε Μαρκοβιανή αλυσίδα

Όπως έχουμε αναφέρει για την μέθοδο αυτή απαιτείται η ύπαρξη ενός χώρου καταστάσεων και του πίνακα μετάβασης. Βάσει λοιπόν του ορισμού ενός $Lin/Con/k/n : F$ συστήματος, ορίζουμε το χώρο καταστάσεων ως $S = \{0, 1, \ldots, k\}$. Αν i είναι μια κατάσταση, τότε αυτή σημαίνει ότι το μήκος της τελευταίας ροής αποτυχιών είναι ίσο με i, για $0 \leq i \leq k-1$, ενώ η κατάσταση k μας δίνει ότι μήκος της τελευταίας ροής αποτυχιών είναι μεγαλύτερο ή ίσο του k.
Ο πίνακας μετάβασης:

$$\Lambda_t = \begin{bmatrix} p_t & q_l & & & \\ p_t & & q_t & & \\ \vdots & & & \ddots & \\ p_t & & & & q_t \\ & & & & 1 \end{bmatrix}$$

Πριν δούμε πως προκύπτει αυτός ο πίνακας, καλό είναι να αναφέρουμε ότι ένα υποσύστημα που περιλαμβάνει τις συνιστώσες $1, \ldots, t$ βρίσκεται στην κατάσταση 1 όταν η συνιστώσα t αποτυγχάνει, η $t-1$ λειτουργεί και δεν υπάρχουν k συνεχόμενες αποτυχίες πριν από την συνιστώσα $t-1$. Επίσης, αυτό που πρέπει να γνωρίζουμε είναι ότι η κατάσταση i ενός υποσυστήματος, όταν αυτό αποτυγχάνει, είναι ο αριθμός των διαδοχικών αποτυχιών στο τέλος του υποσυστήματος, $(0 \leq i < k)$. Είναι ίσο με το k όταν το σύστημα αποτυγχάνει λόγω των k ή και περισσότερων διαδοχικών αποτυχιών στο τέλος της ακολουθίας των συνιστωσών του συστήματος. Έτσι, αν πάρουμε για παράδειγμα την δεύτερη σειρά του πίνακα μετάβασης μπορούμε να πούμε ότι αν το σύστημα με $(t-1)$ συνιστώσες βρίσκεται στην κατάσταση 1, η πιθανότητα για το υποσύστημα με t συνιστώσες να βρίσκεται στην κατάσταση 0 είναι p_t και η πιθανότητα για το υποσύστημα να βρίσκεται στην κατάσταση 2 είναι q_t.
Η αξιοπιστία $R_L(k,n)$ ενός $Lin/Con/k/n : F$ συστήματος δίνεται από τον τύπο $R_L(k,n) = 1 - \alpha_k(n)$ όπου την ποσότητα α_k μπορούμε να την υπολογίσουμε

μέσω των αναδρομικών σχέσεων:

$$\begin{aligned}\alpha_0(t) &= p_t(1-\alpha_k(t-1)),\\ \alpha_j(t) &= q_t\alpha_{j-1}(t-1), \quad j=1,2,\ldots,k-1,\\ \alpha_k(t) &= q_t\alpha_{k-1}(t-1)+\alpha_k(t-1),\end{aligned}$$

με αρχική κατάσταση $\alpha_0(0)=1$ και $\alpha_j(0)=0$ για $j\neq 0$.

4.5 Εφαρμογές στην αξιοπιστία ενός $Lin/Con/2/5:F$ συστήματος

Στην παράγραφο αυτή θα εφαρμόσουμε τους τύπους για την αξιοπιστία ενός $Lin/Con/2/5:F$ συστήματος λαμβάνοντας κάθε φορά υπόψη τη σχέση που έχουν μεταξύ τους οι συνιστώσες αυτού. Την αξιοπιστία θα την υπολογίσουμε είτε μέσω τύπων που αναφέρθηκαν σε προηγούμενες παραγράφους αυτού του κεφαλαίου, είτε χρησιμοποιώντας τη τυχαία μεταβλητή L_n.

Αρχικά, πρέπει να παρατηρήσουμε ότι η πιθανότητα $P(L_5\leq 1)$ είναι η πιθανότητα το μέγιστο μήκος ροής σε μια ακολουθία από 5 δοκιμές να είναι μικρότερο ή ίσο του 1. Αν στο εκάστοτε αποτέλεσμα εναλλάξουμε τα p με τα q, τότε η πιθανότητα $P(L_5\leq 1)$ θα είναι η πιθανότητα να μην εμφανιστεί ροή αποτυχιών μήκους μεγαλύτερου ή ίσου του 2, στις 5 αυτές δοκιμές, με άλλα λόγια δηλαδή μας δίνει την αξιοπιστία ενός $Lin/Con/k/n:F$ συστήματος.

α) *Με ανεξάρτητες και ισόνομες συνιστώσες*

Χρησιμοποιώντας τον κλειστό τύπο των *Chiang* και *Niu* (1981), θα υπολογίσουμε την πιθανότητα:

$$R_L(2,5)=\sum_{j=0}^{3}\binom{5-j+1}{j}q^jp^{5-j}.$$

Έτσι,

<u>Για $j=0$</u>

$$\binom{5-0+1}{0}q^0p^5=\binom{6}{0}p^5=p^5$$

Για $j = 1$

$$\binom{5-1+1}{1} qp^5 = 5qp^4$$

Για $j = 2$

$$\binom{5-2+1}{2} q^2p^3 = 6q^2p^3$$

Για $j = 3$

$$\binom{5-3+1}{3} q^3p^2 = q^3p^2$$

Αθροίζοντας λοιπόν τα παραπάνω έχουμε για την αξιοπιστία:

$$R_L(2,5) = p^5 + 5qp^4 + 6q^2p^3 + q^3p^2.$$

Στο ίδιο αποτέλεσμα θα καταλήξουμε αν χρησιμοποιήσουμε την τυχαία μεταβλητή L_n. Όπως έχουμε αναφέρει, για να πάρουμε το επιθυμητό αποτέλεσμα αρκεί να βρούμε την πιθανότητα $P(L_n \leq k-1)$. Εφαρμόζοντας τον τύπο του *Godbole* (1990) για $n = 5$ και $k = 2$, έχουμε ότι:

$$P(L_5 \leq 1) = \sum_{y=2}^{5} q^y p^{5-y} \sum_{j=0}^{\lfloor (5-y)/2 \rfloor} (-1)^j \binom{y+1}{j} \binom{5-2j}{y}$$

Για $y = 2$ $\longrightarrow 0 \leq j \leq 1$

$$q^2p^3 \left[\binom{3}{0}\binom{5}{2} - \binom{3}{1}\binom{3}{2} \right] = q^2p^3$$

Για $y = 3$ $\longrightarrow 0 \leq j \leq 1$

$$q^3p^2 \left[\binom{4}{0}\binom{5}{3} - \binom{4}{1}\binom{3}{3} \right] = 6q^3p^2$$

Για $y = 4$ $\longrightarrow j = 0$

$$q^4p \binom{5}{0}\binom{5}{4} = 5q^4p$$

Για $\underline{y = 5} \longrightarrow j = 0$

$$q^5 \begin{pmatrix} 6 \\ 0 \end{pmatrix} \begin{pmatrix} 5 \\ 5 \end{pmatrix} = q^5$$

Αθροίζοντας λοιπόν έχουμε το αποτέλεσμα:

$$P(L_5 \leq 1) = q^2p^3 + 6q^3p^2 + 5q^4p + q^5.$$

Εναλλάσσοντας τα p με τα q προκύπτει ο τύπος των $Chiang$ και Niu (1981), που μας δίνει την αξιοπιστία ενός γραμμικού 2-από-τα-5 συστήματος αποτυχίας.

β) *Ανεξάρτητες συνιστώσες*

Στην περίπτωση αυτή θα χρησιμοποιήσουμε την αναδρομική σχέση του $Hwang$ (1982) και έτσι θα έχουμε:

$$\begin{aligned} R_L(2,5) &= \sum_{i=4}^{5} p_i \left(\prod_{j=i+1}^{5} q_j \right) R_L(2, i-1) \\ &= p_4q_5R_L(2,3) + p_5R_L(2,4). \end{aligned}$$

Για να μπορέσουμε να υπολογίουμε την $R_L(2,5)$ πρέπει να βρούμε αντίστοιχα τις $R_L(2,3)$ και $R_L(2,4)$. Έτσι,

$$R_L(2,3) = p_2q_3R_L(2,1) + p_3R_L(2,2)$$

όπου

$$\begin{aligned} R_L(2,2) &= p_1q_2R_L(2,0) + p_2R_L(2,1) \\ &= p_1q_2 + p_2. \end{aligned}$$

Γι' αυτό έχουμε ότι:

$$R_L(2,3) = p_2q_3 + p_3p_1q_2 + p_3p_2$$

Επίσης,

$$\begin{aligned} R_L(2,4) &= p_3q_4R_L(2,2) + p_4R_L(2,3) \\ &= p_3q_4(p_1q_2 + p_2) + p_4(p_2q_3 + p_3p_1q_2 + p_3p_2). \end{aligned}$$

Κάνοντας μια αντικατάσταση προκύπτει ότι:

$$R_L(2,5) = p_4q_5(p_2q_3 + p_3p_1q_2 + p_3p_2) + p_5[p_3q_4(p_1q_2 + p_2) + p_4(p_2q_3 + p_3p_1q_2 + p_3p_2)].$$

γ) *Ανταλλαξιμότητα*

Για την εύρεση της αξιοπιστίας θα χρησιμοποιήσουμε τον τύπο της συνάρτησης κατανομής της τυχαίας μεταβλητής L_n, που απέδειξαν οι *Eryilmaz* και *Demir* (2007) και στην περίπτωση του $Lin/Con/2/5 : F$ έχουμε ότι:

$$P(L_5 < 2) = \sum_{l=0}^{5} \sum_{j=0}^{\min\lfloor l/2, 5-l+1 \rfloor} \sum_{i=0}^{5-l} (-1)^j (-1)^i \binom{5-l+1}{j} \binom{5-2j}{5-l} \binom{5-l}{i} \lambda_{l+i}.$$

<u>Για $l = 0$</u>: $j = 0$ και $i = 0, \ldots, 5$

$$\begin{aligned} & \sum_{i=0}^{5} (-1)^j \binom{6}{0} \binom{5}{5} \binom{5}{i} \lambda_i \\ = \; & \lambda_0 - 5\lambda_1 + 10\lambda_2 - 10\lambda_3 + 5\lambda_4 - \lambda_5 \end{aligned}$$

<u>Για $l = 1$</u>: $j = 0$ και $i = 0, \ldots, 4$

$$\begin{aligned} & \sum_{i=0}^{4} (-1)^j \binom{5}{0} \binom{5}{4} \binom{4}{i} \lambda_{1+i} \\ = \; & 5\lambda_1 - 20\lambda_2 + 30\lambda_3 - 20\lambda_4 + 5\lambda_5 \end{aligned}$$

<u>Για $l = 2$</u>: $j = 0, 1$ και $i = 0, \ldots, 3$

$$\begin{aligned} & \sum_{j=0}^{1} \sum_{i=0}^{3} (-1)^j (-1)^i \binom{4}{j} \binom{5-2j}{3} \binom{3}{i} \lambda_{2+i} \\ = \; & 6\lambda_2 - 18\lambda_3 + 18\lambda_4 - 6\lambda_5 \end{aligned}$$

<u>Για $l = 3$</u>: $j = 0, 1$ και $i = 0, \ldots, 2$

$$\begin{aligned} & \sum_{j=0}^{1} \sum_{i=0}^{2} (-1)^j (-1)^i \binom{3}{j} \binom{5-2j}{2} \binom{2}{i} \lambda_{3+i} \\ = \; & \lambda_3 - 2\lambda_4 + \lambda_5 \end{aligned}$$

Για $l = 4$: $j = 0, 1$ και $i = 0, \ldots, 1$

$$\begin{aligned}
& \sum_{j=0}^{1}\sum_{i=0}^{1}(-1)^j(-1)^i \binom{2}{j}\binom{5-2j}{1}\binom{1}{i}\lambda_{4+i} \\
= \quad & \lambda_4 - \lambda_5
\end{aligned}$$

Για $l = 5$: $j = 0, 1, 2$ και $i = 0$

$$\sum_{j=0}^{2}(-1)^j \binom{1}{j}\binom{5-2j}{0}\binom{0}{0}\lambda_5 = 0$$

Έτσι, προκύπτει ότι:

$$P(L_5 < 2) = \lambda_0 - 4\lambda_2 + 11\lambda_3 - 14\lambda_4 + 6\lambda_5.$$

Για να υπολογίσουμε την αξιοπιστία, θα πρέπει να ορίσουμε τις μεταβλητές λ_i ως εξής:

$$\lambda_i = P(X_1 = \ldots = X_i = 1),$$

όπου $X_i = 1$ όταν η i συνιστώσα αποτυγχάνει, $X_i = 0$ όταν η i συνιστώσα λειτουργεί.

δ) *Μέθοδος Μαρκοβιανής εμφύτευσης*

Θα υπολογίσουμε την αξιοπιστία του συστήματος κάνοντας χρήση των παρακάτω αναδρομικών σχέσεων:

$$\begin{aligned}
\alpha_0(l) &= p_l(1 - \alpha_k(l-1)), \\
\alpha_j(l) &= q_l\alpha_{j-1}(l-1), \quad j = 1, 2, \ldots, k-1, \\
\alpha_k(l) &= q_l\alpha_{k-1}(l-1) + \alpha_k(l-1),
\end{aligned}$$

με αρχικές αυνθήκες $\alpha_0 = 1$ και $\alpha_j(0) = 0$ για $j \neq 0$.

Στην περίπτωσή μας θέλουμε να υπολογίσουμε την ποσόστητα $\alpha_2(5)$:

$$\alpha_2(5) = q_5\alpha_1(4) + \alpha_2(4)$$

Υπολογισμος της ποσόστητας $\alpha_1(4)$:

$$\begin{aligned}\alpha_1(4) &= q_4\alpha_0(3) && (24)\\ \alpha_0(3) &= p_3(1-\alpha_2(2)) && (25)\\ \alpha_2(2) &= q_2\alpha_1(1)-\alpha_2(1) && (26)\\ \alpha_1(1) &= q_1\alpha_0(0)=q_1 && (27)\\ \alpha_2(1) &= 0 && (28)\end{aligned}$$

Αντικαθιστώντας την σχέση (28) στην σχέση (27), την (27) στην (26) κ.ο.κ. προκύπτει:

$$\begin{aligned}\alpha_1(4) &= q_4(p_3-p_3q_2q_1)\\ &= q_4p_3-q_4p_3q_2q_1\end{aligned}$$

Υπολογισμός της ποσότητας $\alpha_2(4)$:

$$\begin{aligned}\alpha_2(4) &= q_4\alpha_1(3)+\alpha_2(3) && (29)\\ \alpha_1(3) &= q_3\alpha_0(2) && (30)\\ \alpha_0(2) &= p_2(1-\alpha_2(1)) && (31)\\ \alpha_2(1) &= 0 && (32)\end{aligned}$$

Υπολογισμός της ποσότητας $\alpha_2(3)$:

$$\begin{aligned}\alpha_2(3) &= q_3\alpha_1(2)+\alpha_2(2) && (33)\\ \alpha_1(2) &= q_2\alpha_0(1) && (34)\\ \alpha_0(1) &= p_1(1-\alpha_2(0)) && (35)\\ \alpha_2(0) &= 0 && (36)\end{aligned}$$

Αντικαθιστώντας την σχέση (36) στην (35), την (35) στην (34) κ.ο.κ. έχουμε ότι:

$$\alpha_2(3) = q_2q_3p_1 + q_1q_2$$

Με αντικατάσταση των ποσοτήτων προκύπτει ότι:

$$\alpha_2(5) = q_5(p_3q_4 - q_1q_2p_3q_4) + q_4q_3p_2 + q_3p_1q_2 + q_1q_2$$

Έτσι λοιπόν η αξιοπιστία ισούται:

$$\begin{aligned} R_L(2,5) &= 1 - \alpha_2(5) \\ &= 1 - q_5(p_3q_4 - q_1q_2p_3q_4) + q_4q_3p_2 + q_3p_1q_2 + q_1q_2. \end{aligned}$$

Έχουμε βέβαια την δυνατότητα να φτάσουμε στο ιδιο αποτέλεσμα χρησιμοποιώντας την τυχαία μεταβλητή L_n. Αυτό μπορεί να πραγματοποιηθεί χρησιμοποιώντας τον τύπο των Fu και $Koutras$ (1994) και πιο συγκεκριμένα:

$$\begin{aligned} P(N_{n,k}=0) &= P(L_n < k) \\ &= \pi_0 \left(\prod_{t=1}^{n} \Lambda_t \right) U'(C_0). \end{aligned}$$

Έτσι λοιπόν εφαρμόζοντας τον τύπο στην περίπτωση που έχουμε $n = 5$ και $k = 2$:

$$\begin{aligned} P(N_{5,2}=0) &= (1,0,0) \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\ &= \alpha_{11} + \alpha_{12}, \end{aligned}$$

όπου:
$\alpha_{11} = q_2q_4q_5 + q_2q_3p_4q_5 + q_1p_2q_3q_5$
$\alpha_{12} = q_2q_3q_4p_5 + q_2p_3q_4p_5 + p_1q_2q_3q_4p_5 + q_1p_2q_3q_4p_5$
$\alpha_{13} = q_2q_3q_4p_5 + q_2p_3p_4 + q_1p_2q_3p_4p_5 + q_1p_2p_3 + p_1p_2$
$\alpha_{21} = q_1q_3q_5 + q_1q_2p_3q_4q_5$

$\alpha_{22} = q_1 q_4 q_5 (q_3 + q_2 p_3)$
$\alpha_{23} = q_1 q_3 p_4 p_5 + q_1 q_2 p_3 p_4 + q_1 p_2 p_3 + p_1$
$\alpha_{31} = 0$
$\alpha_{32} = 0$
$\alpha_{33} = 1$

Εναλλάσσοντας τα p_i με τα q_i, $i = 1, 2, 3, 4, 5$, παίρνουμε την αξιοπιστία ενός γραμμικού 2-από-τα-5 συστήματος αποτυχίας.

Summary

LONGEST SUCCESS RUN AND APLLICATIONS

Consider a sequence of n two state (success-failure) trials with outcomes arranged on a line. Success run is a sequence of consecutive successes preceded and followed by a failure or by nothing. The number of the successes in the success run is referred to as its length.

The use of runs in many scientific fields was very effective. They are used in many areas such as hypotheses testing, quality control, meteorology, biology and system reliability. Specifically, the study of the number of success runs, the number of runs of a specified length using several counting schemes and the longest success run length have attracted the interest of many researchers.

In this work we concentrate on the study of the statistic, L_n, which represents the length of the longest success run in a sequence of n binary trials. The study of the random variables $G_{n,k}$ and $N_{n,k}$, representing the number of success runs of length greater than or equal to k and the number of success runs of length k, respectively, is also important for this study. From its definition it is clear that the longest success run is an upper bound of the number of consecutive successes that appear in the sequence. For example, in the following binary sequence of 10 trials 1100111101, we have $L_{10} = 4$, $G_{10,2} = 2$, $N_{10,2} = 3$ and $G_{10,5} = N_{10,5} = 0$.

Our study gives an overview of results referring to the distribution of the random variable L_n defined on sequences of *Bernoulli* trials (independent and identically distributed), Poisson trials (independent), Markov dependent trials, exchangeable trials and sequences with outcomes from a Polya-Eqqenberger sampling scheme, as a case of particular importance of exchangeability. The methods that have been used to obtain the distribution of L_n are also presented ; i.e. combinatorial analysis, recursive schemes, generating functions and the Markov chain imbedding technique.

The distribution of L_n is used to study a consecutive-k-out-of-n:F system; i.e. a system that fails if and only if k consecutive failures appear among its components. Several results concerning the reliability of such a system are

also presented. Numerical examples are given for comparison reasons and to illustrate the theoretical results.

Αναφορές

[1] N. Balakrishnan, M.V. Koutras (2002). Runs and scans with applications, *Wiley, New York*

[2] R.C. Bollinger (1982). Direct computation for consecutive k-out-of-n:F systems, *IEEE Transactions on Reliability, r-31(5), 444-446*

[3] R.C. Bollinger and A.A. Salvia (1982). Consecutive k-out-of-n:F networks, *IEEE Transactions on Reliability, R-31(1), 53-55*

[4] D.T. Chiang and S.C. Niu (1981). Reliability of consecutive k-out-of-n:F system, *IEEE Transactions on Reliability, R-30(1): 87-89*

[5] C.Derman, G.J. Lieberman and S.M. Ross (1974). On the consecutive k-out-of-n:F system, *IEEE Transactions on Reliability, R-31(1): 57-63*

[6] S. Eryilmaz (2005a). The longest run statistic associated with exchangeable binary variables, *Turk. Eng. Environ. Sci. 29: 105-111*

[7] S. Eryilmaz (2005b). On the distribution and expectation of success runs in nonhomogeneous Markov dependent trials, *Applied Mathematics and Computation, 175: 119-130*

[8] S. Eryilmaz and S. Demir (2007). Success run in a sequence of exchangeable binary data, *Biometrics, 51, 512-523*

[9] J.C. Fu and M.V. Koutras (1994). Distribution theory of runs: a Markov chain approach, *Journal of the American Statistical Association, 89, 1050-1058*

[10] E.O George and D. Bowman (1995). A full likelihood prodecure for analyzing exchangeable binary data, *Biometrics, 51, 512-523*

[11] A.P. Godbole (1990). Specific formula for some success run distributions, *Statistics and Probability Letters 10, 119-124*

[12] I.P. Goulden (1987). Generating functions and reliabilities for consecutive k-out-of-n:F systems, *Utilitas Mathematic, 32(1): 141-147*

[13] F.K. Hwang (1982). Fast solutions for consecutive k-out-of-n:F system, *IEEE Transactions on Reliability, R-31(5): 447-448*

[14] F.K. Hwang (1986). Simplified reliabilities for consecutive k-out-of-n:F system, *SIAM Journal on Algebraic and Discrete Methods, 7(2): 258-264*

[15] W.Kuo and M.J. Zuo (2003). Optimal reliability modeling, *New Jersey*

[16] M. Lambiris and S.G. Papastavridis (1985). Exact reliability formulas for linear and circular consecutive k-out-of-n:F systems, *IEEE Transactions on Reliability, R-34(2), 124-126*

[17] W.Y.W. Lou (1996). On runs and longest run tests: A method of finite Markov chain imbedding, *Journal of the American Statistical Association, Vol 91, No. 436*

[18] F.S. Makri, A.N. Philippou and Z.M. Psillakis (2007a). Shortest and longest length of success runs in binary sequences, *Journal of Statistical Planning and Inference, 137: 2226-2239*

[19] F.S. Makri, A.N. Philippou and Z.M. Psillakis (2007b). Success run statistics defined on a urn model, *Advances in Applied Probability, 39:991-1019*

[20] F.S. Makri and Z.M. Psillakis (2009). On success runs of length exceeded a threshold, *Methodology and Computing in Applied Probability, DOI 10.1007/s11009-009-9147-1*

[21] F. Mosteler (1941). Note on an application of runs to quality control charts, *Annals of Mathematical Statistics, 12, 228-232*

[22] M. Muselli (1996). Simple expressions for success run distributions, *Statistics and Probability Letters, 10, 119-124*

[23] A.N. Philippou and F.S. Makri (1986). Successes runs and longest runs, *Statistics and Probability Letters, 4, 101-105*

[24] A.N. Philippou and A.A. Muwafi (1982). Waiting for the kth consecutive success and the Fibonacci sequence of order k, *The Fibonacci Quarterly 20, 28-32*

[25] A. Wald and J. Wolfowitz (1940). Another generalized Fibonacci sequence, *The Fibonacci Quarterly, 5, 209-222*

[26] J. Wolfowitz (1943). On the theory of runs with some applications to quality control, *Annals of Mathematical Statistics, 14, 280-288*

[27] M. Zuo and W. Kuo (1990). Design and performance analysis of consecutive k-out-of-n structure, *Naval Research Logistics, 37: 203-230*