



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

**Ανάλυση Επιβίωσης με Ανταγωνιστικούς Κινδύνους και Μοντέλα Multi-state**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Δημήτρη Παπαδημητρίου

10<sup>ο</sup> Εξάμηνο

A.M. : ge12006

Επιβλέπουσα Καθηγήτρια: Χρυσίς Καρώνη

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2017



## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Ανάλυση Επιβίωσης με Ανταγωνιστικούς Κινδύνους και Μοντέλα Multi-state**

Δημήτρης Παπαδημητρίου  
Α.Μ.: ge12006

Επιβλέπουσα Καθηγήτρια: Χρυσή Καρώνη

Τριμελής εξεταστική επιτροπή:

Χρυσή Καρώνη  
Καθηγήτρια Ε.Μ.Π.

Χρήστος Κουκουβίνος  
Καθηγητής Ε.Μ.Π.

Φίλια Βόντα  
Αναπλ. Καθηγήτρια Ε.Μ.Π.

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2017



## Ευχαριστίες

Για τη διεκπεραίωση της παρούσας Διπλωματικής Εργασίας θα ήθελα να ευχαριστήσω την καθηγήτρια του Ε.Μ.Π., κα Χρυσήδα Καρώνη, για την καθοδήγηση και την πολύτιμη συμβολή της.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένειά μου για την συνεχή στήριξή της καθ' όλη την διάρκεια των σπουδών μου, και ιδίως τον πατέρα μου που με έμαθε να σκέφτομαι.



## Περιεχόμενα

<b>Ευχαριστίες</b> .....	<b>i</b>
<b>Περίληψη</b> .....	<b>iv</b>
<b>Abstract</b> .....	<b>v</b>
<b>1 Ανάλυση Επιβίωσης</b> .....	<b>1</b>
1.1 Εισαγωγή .....	1
1.2 Μαθηματική Μοντελοποίηση .....	2
1.3 Εκτίμηση Μεγεθών και Στατιστική Συμπερασματολογία .....	3
1.4 Μοντέλα Παλινδρόμησης για την Συνάρτηση Διακινδύνευσης .....	7
1.5 Εφαρμογή του Μοντέλου του Cox .....	11
1.5.1 Πλήρες Μοντέλο .....	11
1.5.2 Προσαρμογή εναλλακτικού μοντέλου .....	13
1.5.3 Έλεγχος της υπόθεσης αναλογικότητας .....	14
<b>2 Ανταγωνιστικοί Κίνδυνοι</b> .....	<b>17</b>
2.1 Εισαγωγή .....	17
2.2 Μοντελοποίηση των προβλημάτων ανταγωνιστικών κινδύνων .....	18
2.3 Μη-παραμετρικές μέθοδοι εκτίμησης .....	20
2.4 Λανθάνοντες Χρόνοι Αποτυχίας .....	22
2.5 Προσαρμογή ημι-παραμετρικών μοντέλων .....	24
2.5.1 Μέθοδος Reduced Rank .....	25
2.5.2 Πιθανότητες Μετάβασης .....	28
2.6 Ημι-παραμετρικό μοντέλο των Fine και Gray .....	30
2.7 Εφαρμογή προβλήματος Ανταγωνιστικών Κινδύνων .....	35
2.7.1 Περιγραφή του προβλήματος .....	35
2.7.2 Μη-παραμετρική ανάλυση .....	37
2.7.3 Σύγκριση του προβλήματος ανταγωνιστικών κινδύνων με το αντίστοιχο single-event .....	40
2.7.4 Ημι-παραμετρική ανάλυση .....	42
2.7.5 Μέθοδοι μείωσης του συνολικού αριθμού παραμέτρων .....	47
2.7.6 Ημι-παραμετρική εκτίμηση των πιθανοτήτων μετάβασης .....	50

2.7.7 Προσαρμογή μοντέλων παλινδρόμησης Fine και Gray στις συναρτήσεις διακινδύνευσης .....	62
2.7.8 Σύγκριση των ανά αιτία και Fine και Gray μοντέλων παλινδρόμησης ....	68
<b>3 Μοντέλα Multi-state .....</b>	<b>71</b>
3.1 Μοντελοποίηση .....	71
3.2 Δύο απλοποιημένα μοντέλα multi-state .....	74
3.3 Μαρκοβιανές Αλυσίδες Συνεχούς Χρόνου .....	75
3.3.1 Εκτιμήτριες Aalen-Johansen .....	76
3.4 Uni-Directional Μοντέλα Illness-Death .....	79
3.4.1 Ανά μετάβαση συναρτήσεις διακινδύνευσης .....	79
3.4.2 Πιθανότητες Μετάβασης .....	80
3.4.3 Εκτιμήσεις των πιθανοτήτων μετάβασης .....	82
3.5 Εφαρμογή σε Uni-directional μοντέλο illness-death .....	84
3.5.1 Περιγραφή του προβλήματος .....	84
3.5.2 Μη-παραμετρική εκτίμηση των πιθανοτήτων μετάβασης .....	87
3.5.3 Ημι-παραμετρική ανάλυση .....	94
3.5.4 Εκτιμήτριες Breslow .....	99
3.5.5 Εκτιμήτριες Aalen-Johansen .....	102
<b>4 Παράρτημα .....</b>	<b>130</b>
4.1 Μελέτη των Klein και Moeschberger για τη μεταμόσχευση του μυελού των οστών .....	130
4.2 Μελέτη του Lagakos για τον καρκίνο του πνεύμονα .....	133
<b>Βιβλιογραφία .....</b>	<b>136</b>



## Περίληψη

Η ανάλυση επιβίωσης πραγματεύεται την ανάλυση δεδομένων που αφορούν στον χρόνο μέχρι την επέλευση ενός γεγονότος. Οι βασικές προσδιορίσιμες ποσότητες των δεδομένων αυτών είναι οι συναρτήσεις επιβίωσης, διακινδύνευσης και σωρευτικής διακινδύνευσης. Περιγράφονται μη-παραμετρικές μέθοδοι εκτίμησης των παραπάνω συναρτήσεων και εισάγεται το ημι-παραμετρικό μοντέλο αναλογικής διακινδύνευσης του Cox το οποίο και προσαρμόζεται στην συνάρτηση διακινδύνευσης.

Τα δεδομένα ανταγωνιστικών κινδύνων προκύπτουν όταν οι υπό παρακολούθηση μονάδες μπορούν να εμφανίσουν περισσότερα του ενός, αμοιβαίως αποκλειόμενα, γεγονότα. Μελετώνται δύο μη-παραμετρικές θεωρήσεις και εφαρμόζονται δύο ημι-παραμετρικά μοντέλα παλινδρόμησης τα οποία, αν και στηρίζονται σε διαφορετικές υποθέσεις, βασίζονται στην μοντελοποίηση της συνάρτησης διακινδύνευσης κάνοντας χρήση του μοντέλου του Cox. Ακόμα, παρουσιάζονται διάφορες μέθοδοι μείωσης του συνολικού αριθμού των παραμέτρων που χρειάζεται να εκτιμηθούν στα παραπάνω μοντέλα.

Εισάγονται τα μοντέλα multi-state, τα οποία περιγράφουν διαδικασίες όπου οι υπό παρακολούθηση μονάδες, κάθε χρονική στιγμή, βρίσκονται σε μία από διάφορες πιθανές καταστάσεις. Τα γεγονότα ενδιαφέροντος μας είναι οι μεταβάσεις μεταξύ των καταστάσεων. Η ανάλυση τους γίνεται χρησιμοποιώντας δεξιά συνεχείς μη-ομοιογενείς στοχαστικές διαδικασίες πεπερασμένων καταστάσεων και συνεχούς χρόνου. Με την υπόθεση της ισχυρής Μαρκοβιανής ιδιότητας, οι πιθανότητες μετάβασης των παραπάνω αλυσίδων μπορούν να εκτιμηθούν χρησιμοποιώντας τις εκτιμήτριες Aalen-Johansen. Οι εκτιμήτριες αυτές υπολογίζονται χρησιμοποιώντας μία μη-παραμετρική και μία ημι-παραμετρική μέθοδο βασισμένη στην προσαρμογή μοντέλων Cox στις ανά μετάβαση συναρτήσεις διακινδύνευσης. Ιδιαίτερη έμφαση δίνεται στα μοντέλα illness-death τα οποία έχουν μία απορροφητική και μία μεταβατική κατάσταση.

Τέλος, χρησιμοποιώντας τρεις βιβλιοθήκες της R, εφαρμόζονται οι μη-παραμετρικές και ημι-παραμετρικές μεθοδολογίες που αναφέρθηκαν παραπάνω σε ένα πρόβλημα ανταγωνιστικών κινδύνων με δύο αιτίες αποτυχίας καθώς και σε ένα μοντέλο illness-death.

**Λέξεις κλειδιά:** Ανταγωνιστικοί κίνδυνοι, μοντέλα multi-state, στρωματοποιημένα ανά αιτία παλινδρόμηση κατά Cox, μοντέλο των Fine και Gray, μοντέλο illness-death, μαρκοβιανές αλυσίδες συνεχούς χρόνου, εκτιμήτριες Aalen-Johansen.

## Abstract

Survival Analysis is concerned with the analysis of data that pertain to the time span until the occurrence of an event. The basic, determinable quantities of these data are the survival, hazard and cumulative hazard functions. Non-parametric methods for estimating these functions are described and the semi-parametric proportional hazards Cox model is introduced and subsequently fitted to the hazard function.

Competing risks data arise when the units under observation can exhibit more than one, mutually exclusive, events. Two non-parametric theories are studied and two semi-parametric regression models are fitted which, albeit depending upon different assumptions, are based in modelling the hazard function using the Cox model. Moreover, several methods for reducing the total number of parameters that need to be estimated in the aforementioned models, are presented.

Multi-state models, which describe situations where the units under observation can be found at any time in one of several possible states, are introduced. The events of interest in this case are the transitions between states. The analysis of multi-state models is implemented using right continuous non-homogeneous stochastic processes with a finite number of states in continuous time. Under the assumption of the strong Markov property the transition probabilities of the chains can be estimated using the Aalen-Johansen estimators. These estimators are computed using one non-parametric and one semi-parametric method based on fitting Cox models to the transition hazard functions. Particular emphasis is put on the illness-death model which has one absorbing and one transient state.

Finally, using three R libraries, the aforementioned non-parametric and semi-parametric methodologies are applied on a competing risks problem with two events of failure as well as on an illness-death model.

**Keywords:** Competing risks, multi-state models, stratified cause-specific Cox regression, Fine and Gray model, illness-death model, continuous time markov chains, Aalen-Johansen estimators.

# 1 Ανάλυση Επιβίωσης

## 1.1 Εισαγωγή

Η ανάλυση δεδομένων διάρκειας ζωής αφορά την παρακολούθηση του χρόνου μέχρι να συμβεί ένα γεγονός που μας ενδιαφέρει, το οποίο καλούμε συνήθως αποτυχία/failure. Τέτοια παραδείγματα μπορεί να είναι ο χρόνος από την μόλυνση ενός ατόμου με HIV μέχρι την εμφάνιση του AIDS, ο χρόνος από την μεταμόσχευση μυελού των οστών μέχρι την επανεμφάνιση λευχαιμίας, ο χρόνος λειτουργίας μέχρι την βλάβη μιας μηχανής ή και ο χρόνος μέχρι την χρεοκοπία μιας εταιρείας. Στα παραπάνω παραδείγματα, το AIDS, η επανεμφάνιση λευχαιμίας, η βλάβη της μηχανής και η χρεοκοπία αποτελούν τις αποτυχίες που μας ενδιαφέρει να μελετήσουμε ενώ τα δεδομένα που έχουμε στην διάθεση μας καλούνται μονάδες ενδιαφέροντος. Η ανάλυση επιβίωσης και αξιοπιστίας είναι ένας κλάδος της στατιστικής που χρησιμοποιείται κατά κόρον από τις επιστήμες της βιοστατιστικής (biostatistics), του ελέγχου ποιότητας (industrial statistics) και της αναλογιστικής επιστήμης (actuarial science) (Machin et al. , 2006).

Τα περισσότερα δεδομένα του παραπάνω κλάδου, ωστόσο, όπως και αυτά που θα χρησιμοποιήσουμε σε αυτή την εργασία, προέρχονται από ιατρικές και φαρμακευτικές μελέτες. Κεντρικό ρόλο στην ανάλυση μας παίζει η τυχαία μεταβλητή  $T$  η οποία υποδηλώνει τον λεγόμενο χρόνο επιβίωσης, την διάρκεια ζωής δηλαδή των μονάδων ενδιαφέροντος μας. Στην συνέχεια θα αναφερθούμε στις κεντρικές υποθέσεις που κάνουμε για την παραπάνω τυχαία μεταβλητή, αφού όμως αναφερθούμε πρώτα σε δύο είδη δεδομένων που συναντώνται πολύ συχνά στην ανάλυση επιβίωσης.

**1. Δεξιά αποκομμένα/Right censored:** Πολλές φορές δεν έχουμε στην διάθεση μας τον ακριβή χρόνο της διάρκειας ζωής μιας μονάδας που μας ενδιαφέρει να μελετήσουμε, αλλά γνωρίζουμε ότι έχει ξεπεράσει το συνολικό χρόνο που ήταν προς παρακολούθηση, έστω  $c$ . Τέτοια παραδείγματα μπορεί να είναι ο τερματισμός ενός πειράματος παρακολούθησης της διάρκειας λειτουργίας ενός λαμπτήρα ή ο τερματισμός της κλινικής μελέτης ενός ασθενούς προτού εκδηλώσει τα συμπτώματα του. Σε κάθε τέτοια περίπτωση, γνωρίζουμε ότι  $T > c$ , ότι δηλαδή η διάρκεια ζωής της μονάδας ενδιαφέροντος είναι μεγαλύτερη της τιμής αποκοπής  $c$ , πληροφορία που θα χρησιμοποιήσουμε στην περαιτέρω ανάλυση μας. Η παρατηρούμενη χρονική τιμή για κάθε μονάδα ενδιαφέροντος, δηλαδή, είναι η  $\min\{T, C\}$ , όπου  $T$  είναι η τυχαία μεταβλητή που αναφέρεται στον χρόνο αποτυχίας των μονάδων (χρόνος επιβίωσης) και  $C$  είναι η μεταβλητή που εκφράζει τον μηχανισμό δεξιών αποκοπών που εφαρμόζεται. Μάλιστα, όσον αφορά την μεταβλητή  $C$

(μηχανισμός αποκοπής/censoring mechanism), θεωρούμε καθ' όλη τη διάρκεια του κειμένου αυτού ότι είναι ανεξάρτητη της  $T$ .

**2. Αριστερά κολοβές/Left truncated:** Μια ακόμη μορφή ατελών δεδομένων που συνδέονται άρρηκτα με την ανάλυση επιβίωσης είναι τα δεδομένα στα οποία η παρακολούθηση της διάρκειας ζωής κάποιων μονάδων ενδιαφέροντος ξεκινάει μια μεταγενέστερη χρονική στιγμή  $t = L$  από τη στιγμή που έγινε η εμφάνιση του γεγονότος που μας ενδιαφέρει να μελετήσουμε  $t = 0$ . Τυπικά παραδείγματα είναι η παρακολούθηση εργοστασιακών εξαρτημάτων τα οποία έχουν ήδη δοκιμαστεί για κάποιες ώρες, χωρίς να υπάρχουν σχετικές πληροφορίες, αλλά και η έναρξη παρακολούθησης ενός ασθενή με καρκίνο, αφού εμφανίστηκε στον οργανισμό του.

Έχουμε, δηλαδή, πληροφορίες για την διάρκεια ζωής των μονάδων ενδιαφέροντος μας μόνο για τις χρονικές στιγμές  $t > L$ . Θα θεωρούμε τον χρόνο  $L$  γνωστό για όλα τα αριστερά κολοβά δεδομένα που θα συναντήσουμε στην συνέχεια καθώς η ανάλυση δεδομένων στα οποία ο χρόνος  $L$  δεν είναι γνωστός είναι αρκετά πιο περίπλοκη και δεν θα μας απασχολήσει σε αυτό το κείμενο (Moore, 2016).

## 1.2 Μαθηματική Μοντελοποίηση

Έστω λοιπόν  $T$  η διάρκεια ζωής μιας μονάδας ενδιαφέροντος, για παράδειγμα ενός ασθενή, ενός εξαρτήματος ή μιας εταιρείας. Μοντελοποιούμε την  $T$  ως μια συνεχή και θετική τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας  $f(t)$ . Ορίζουμε επίσης, την συνάρτηση κατανομής  $F(t) = \int_0^t f(u)du$  αλλά και την πολύ σημαντική στα δεδομένα διάρκειας ζωής, συνάρτηση αξιοπιστίας ή συνάρτηση επιβίωσης:

$$S(t) = 1 - F(t) = P(T > t) = \int_t^{\infty} f(u)du$$

η οποία εκφράζει την πιθανότητα η διάρκεια ζωής μιας μονάδας ενδιαφέροντος να υπερβαίνει τον χρόνο  $t$ .

Επίσης:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

Επίσης, ορίζουμε την συνάρτηση διακινδύνευσης/hazard function:

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}$$

Η παραπάνω συνάρτηση έχει θεμελιώδη ρόλο στην ανάλυση επιβίωσης και εκφράζει την τάση προς αποτυχία μιας μονάδας ενδιαφέροντος στο χρονικό διάστημα  $(t, t + dt]$  με δεδομένη την επιβίωση του μέχρι την χρονική στιγμή  $t$ .

Αν, επιπλέον, η  $F(t)$  είναι συνεχής, τότε προκύπτει η παραπάνω εξίσωση:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d \log(1 - F(t))}{dt}$$

Ακόμα, ορίζουμε και την σωρευτική συνάρτηση διακινδύνευσης/cumulative hazard, ως εξής:

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln S(t)$$

ή αλλιώς:

$$S(t) = \exp(-\Lambda(t))$$

Η ερμηνεία της συνάρτησης  $\Lambda(t)$  είναι αρκετά πιο πολύπλοκη από αυτές που έχουμε δει ως τώρα. Έστω ότι παρακολουθούμε μια μονάδα ενδιαφέροντος από την  $t = 0$ . Έστω, τώρα, ότι την χρονική στιγμή  $T_1$  η μονάδα αποτυγχάνει. Την αντικαθιστούμε με μια άλλη, ακριβώς όμοια μονάδα, η οποία είναι σε λειτουργία από την  $t = 0$  αλλά είναι ακόμα σε λειτουργία και δεν έχει αποτύχει την  $T_1$ . Τότε, συνεχίζουμε την παρακολούθηση της μονάδας αυτής μέχρι την χρονική στιγμή αποτυχίας της  $T_2$ , και ακολουθούμε την ίδια διαδικασία. Η συνάρτηση  $\Lambda(t)$  εκφράζει τον αναμενόμενο αριθμό επανατοποθετήσεων τέτοιων μονάδων, με τον τρόπο που περιγράψαμε, μέχρι και την χρονική στιγμή  $t$  (Andersen and Keiding, 2011).

### 1.3 Εκτίμηση Μεγεθών και Στατιστική Συμπερασματολογία

Υπάρχουν πολλά παραμετρικά μοντέλα που θα μπορούσαμε να προσαρμόσουμε σε μια τυχαία μεταβλητή χρόνου επιβίωσης, όπως των Exponential, Gamma, Weibull, Log-Normal και Log-Logistic κατανομών. Ωστόσο, σε αυτήν την εργασία θα μας απασχολήσουν μόνο τα μη-παραμετρικά/non-parametric μοντέλα επιβίωσης στα οποία θα εμφανίζονται κατά κανόνα δεξιά αποκομμένες παρατηρήσεις. Έστω τυχαίο δείγμα  $n$  μονάδων παρακολούθησης, μερικές εκ των οποίων, έστω  $k$ , δεν είναι δεξιά αποκομμένες και έχουν διατεταγμένους χρόνους αποτυχίας  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ . Συμβολίζουμε επίσης με  $n_j$  τις μονάδες ενδιαφέροντος που λειτουργούσαν αμέσως πριν την  $t_{(j)}$  και με  $d_j$  τον αριθμό των μονάδων που αποτυγχάνουν την χρονική στιγμή  $t_{(j)}$ ,  $\forall j = 1, 2, \dots, k$ .

Τότε, ορίζουμε την διακριτή μορφή της συνάρτησης διακινδύνευσης και επιβίωσης, ως εξής:

$$\hat{\lambda}(t_{(j)}) = P(T = t_{(j)} | T > t_{(j-1)}) = \frac{d_{(j)}}{n_{(j)}}$$

Επομένως, προκύπτει:

$$\hat{S}(t_{(j)}) = P(T > t_{(j)}) = P(T > t_{(1)})P(T > t_{(2)}|T > t_{(1)})\dots P(T > t_{(j)}|T > t_{(j-1)})$$

ή αλλιώς:

$$\hat{S}(t) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_{(j)}}{n_{(j)}}\right)$$

Η παραπάνω μη-παραμετρική εκτιμήτρια της συνάρτησης επιβίωσης λέγεται Kaplan-Meier (Aalen et al. , 2008).

Το τυπικό της σφάλμα δίνεται από τον τύπο του Greenwood:

$$se(\hat{S}(t)) = \hat{S}(t) \sqrt{\sum_{t_{(j)} \leq t} \frac{d_{(j)}}{n_{(j)}(n_{(j)} - d_{(j)})}}$$

Χρησιμοποιώντας την μη-παραμετρική εκτιμήτρια συνάρτησης επιβίωσης Kaplan-Meier μπορούμε να κατασκευάσουμε την επίσης μη-παραμετρική εκτιμήτρια σωρευτικού κινδύνου Nelson-Aalen, ως εξής:

$$\hat{\Lambda}(t) = -\ln\left(\prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_{(j)}}{n_{(j)}}\right)\right) = -\sum_{j:t_{(j)} \leq t} \ln\left(1 - \frac{d_{(j)}}{n_{(j)}}\right) \simeq \sum_{t_{(j)} \leq t} \frac{d_{(j)}}{n_{(j)}}$$

Η αντίστοιχη διασπορά της παραπάνω εκτιμήτριας είναι:

$$\hat{V}(\hat{\Lambda}) = \sum_{j:t_{(j)} \leq t} \frac{d_{(j)}}{n_{(j)}^2}$$

### Εφαρμογή:

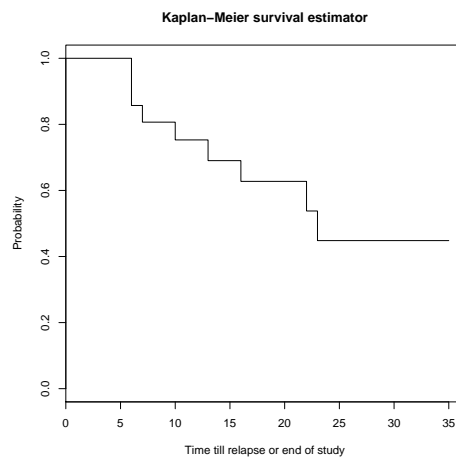
Στον Πίνακα 1 παρουσιάζονται οι πρώτες έξι παρατηρήσεις από τα δεδομένα από μια μελέτη του 1963 για την επίδραση ενός φαρμάκου (6-MP) που χορηγείται για την αντιμετώπιση της λευχαιμίας σε 21 ασθενείς (Klein and Moeschberger, 1997).

Πίνακας 1: Κλινική μελέτη φαρμάκου 6-MP.

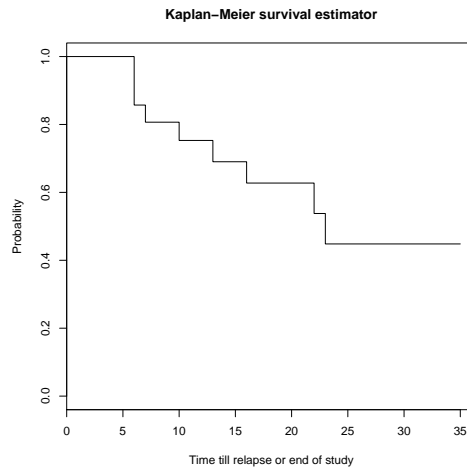
Patient ID	Time	Censoring
1	10	1
2	7	1
3	32	0
4	23	1
5	22	1
6	6	1

Δίνεται ο χρόνος, σε μήνες, από την χορήγηση του φαρμάκου μέχρι το τέλος της παρακολούθησης του κάθε ασθενή. Οι παραπάνω χρονικές διάρκειες, δηλαδή, εκφράζουν είτε τον χρόνο επανεμφάνισης της λευχαιμίας (Censoring=1) είτε την αποχώρηση του ασθενή από την κλινική μελέτη (Censoring=0). Συνεπώς, στα δεδομένα μας παρουσιάζονται δεξιά αποκομμένες παρατηρήσεις τις οποίες θα συμπεριλάβουμε, όπως αναφέραμε και παραπάνω, στην ανάλυση μας. Τέλος, θεωρούμε ότι καμία εκ των παρατηρήσεων μας δεν είναι αριστερά κολοβή.

Στα Σχήματα 1 και 2, σχεδιάζουμε με την βοήθεια της R, χρησιμοποιώντας το πακέτο survival (Therneau, 2017) τις βηματικές εκτιμήτριες Kaplan-Meier και Nelson-Aalen, αντίστοιχα, για τα δεδομένα του Πίνακα 1.



Σχήμα 1: Εκτιμήτρια Kaplan-Meier της συνάρτησης επιβίωσης.



Σχήμα 2: Εκτιμήτρια Nelson-Aalen της συνάρτησης σωρευτικής διακινδύνευσης.

**Συνάρτηση Πιθανοφάνειας παρουσία δεξιά αποκομμένων παρατηρήσεων:**

Στην συνέχεια, θα μελετήσουμε τον ορισμό της συνάρτησης πιθανοφάνειας παρουσία δεξιά αποκομμένων παρατηρήσεων.

Έστω οι παρατηρήσεις  $t_i : i = 1, 2, \dots, n$  για την διάρκεια ζωής μιας μονάδας ενδιαφέροντος και έστω η συνάρτηση:

$$\delta_i = \begin{cases} 1, & \text{i-οστή παρατήρηση μη-αποκομμένη} \\ 0, & \text{i-οστή παρατήρηση δεξιά αποκομμένη} \end{cases}$$

Τότε, η συνάρτηση πιθανοφάνειας τέτοιων δεδομένων αποτελείται από δύο συνιστώσες, την  $L_U$  για τις μη-αποκομμένες παρατηρήσεις και την  $L_C$  για τις δεξιά αποκομμένες. Συνδυαστικά, τελικά προκύπτει η συνάρτηση πιθανοφάνειας:

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^n \lambda(t_i)^{\delta_i} S(t_i)$$

αλλά και η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας:

$$l = \sum_{i=1}^n (\delta_i \ln f(t_i) + (1 - \delta_i) \ln S(t_i))$$



## 1.4 Μοντέλα Παλινδρόμησης για την Συνάρτηση Διακινδύνευσης

Η συνάρτηση διακινδύνευσης έχει θεμελιώδη σημασία για την ανάλυση δεδομένων διάρκειας ζωής καθώς εκφράζει την πιθανότητα αποτυχίας μιας μονάδας ενδιαφέροντος σε ένα απειροστό χρονικό διάστημα δια το μήκος του διαστήματος αυτού. Η μοντελοποίηση της συνάρτησης αυτής αποτελεί κομβικό ζήτημα στην ανάλυση επιβίωσης και αξιοπιστίας. Έχει δημιουργηθεί πληθώρα τέτοιων μοντέλων, τόσο παραμετρικών όσο και ημι-παραμετρικών. Στην παρούσα εργασία θα ασχοληθούμε με το ημι-παραμετρικό μοντέλο αναλογικής διακινδύνευσης του Cox, το οποίο ορίζεται ως εξής:

Έστω ότι για κάθε μονάδα ενδιαφέροντος που έχουμε στην διάθεση μας είναι γνωστές οι τιμές  $p$  μεταβλητών  $z_1, z_2, \dots, z_p$  τις οποίες θέλουμε να συμπεριλάβουμε στο μοντέλο που θα κατασκευάσουμε για την συνάρτηση διακινδύνευσης. Για παράδειγμα, σε μια κλινική μελέτη για την επίδραση ενός φαρμάκου σε ασθενείς που πάσχουν από άνοια μας ενδιαφέρει να μελετήσουμε το κατά πόσο μεταβλητές όπως η ηλικία, το βάρος, τα γενετικά χαρακτηριστικά των ατόμων, είναι παράγοντες οι οποίοι σχετίζονται με τον χρόνο επιβίωσης των ασθενών. Οι μεταβλητές  $z_1, z_2, \dots, z_p$  λέγονται συμμεταβλητές/covariates και έστω  $\vec{z}$  το  $p \times 1$  διάνυσμα τους. Σε κάθε μονάδα ενδιαφέροντος  $i = 1, 2, \dots, n$  αντιστοιχεί και ένα διάνυσμα  $\vec{z}_i$ .

### Μοντέλο αναλογικής διακινδύνευσης:

Ας ορίσουμε αρχικά το μοντέλο παλινδρόμησης της αναλογικής διακινδύνευσης/proportional hazards:

$$\lambda(t|\vec{z}) = g(\vec{z})\lambda_0(t)$$

όπου η συνάρτηση  $\lambda(t|\vec{z})$  είναι η συνάρτηση διακινδύνευσης, η  $g(\vec{z})$  παίρνει θετικές τιμές και η  $\lambda_0(t)$  λέγεται βασική συνάρτηση διακινδύνευσης/baseline hazard function και ισούται με την συνάρτηση διακινδύνευσης όταν  $\vec{z} = \vec{0}$ .

Το μοντέλο παλινδρόμησης αυτό, παίρνει το όνομα του από την εξής πολύ σημαντική ιδιότητα του (Καρώνη, 2009).

Έστω δύο μονάδες ενδιαφέροντος με διανύσματα συμμεταβλητών  $\vec{z}_1$  και  $\vec{z}_2$ , αντιστοίχα. Τότε ο λόγος των αντίστοιχων συναρτήσεων διακινδύνευσης τους

$$\frac{\lambda_1(t|\vec{z})}{\lambda_2(t|\vec{z})} = \frac{g(\vec{z}_1)}{g(\vec{z}_2)}$$

είναι σταθερός και ανεξάρτητος του χρόνου  $t$ .

Η βασική συνάρτηση διακινδύνευσης παίζει μεγάλο ρόλο στην παραπάνω ανάλυση. Η μορφή της διαχωρίζει τα παραπάνω μοντέλα παλινδρόμησης αναλογικής διακινδύνευσης σε δύο μεγάλες κατηγορίες, τα παραμετρικά και τα ημι-παραμετρικά. Πράγματι, θα μπορούσαμε να προσαρμόσουμε στην  $\lambda_0(t)$  διάφορες κατανομές όπως την Weibull ή την Gumbel. Στο κείμενο αυτό, ωστόσο, δεν θα ασχοληθούμε με παραμετρικά μοντέλα αναλογικής διακινδύνευσης. Θα μας απασχολήσουν μόνο τα ημι-παραμετρικά, στα οποία η  $\lambda_0(t)$  παραμένει ακαθόριστη.

### Μοντέλο του Cox:

Στην ειδική περίπτωση που θέσουμε  $g(\vec{z}) = e^{(\vec{b}^T \vec{z})}$ , προκύπτει το πολύ γνωστό στην βιβλιογραφία της ανάλυσης επιβίωσης, μοντέλο του Cox (Cox, 1972):

$$\lambda(t|\vec{z}) = \lambda_0(t)e^{\vec{b}^T \vec{z}}$$

Το  $p \times 1$  διάνυσμα  $\vec{b}$  περιέχει τους συντελεστές των συμμεταβλητών  $z_i$  για μονάδα  $i = 1, 2, \dots, n$  που μελετούμε. Οι συντελεστές αυτοί εκφράζουν την ποσοτική επίδραση καθεμιάς εκ των συμμεταβλητών στην συνάρτηση διακινδύνευσης, όπως έχει μοντελοποιηθεί από το μοντέλο του Cox. Σκοπός μας είναι να εκτιμήσουμε τους παραπάνω συντελεστές ώστε να διαπιστώσουμε ποιες συμμεταβλητές είναι στατιστικά σημαντικές στο μοντέλο παλινδρόμησης, ποιες δεν είναι αλλά και ποια είναι η ακριβής επίδραση των πρώτων στην συνάρτηση διακινδύνευσης και, έμμεσα, στην διάρκεια ζωής των μονάδων που μελετούμε. Ο προσδιορισμός των παραπάνω συντελεστών γίνεται μέσω της συνάρτησης Μερικής Πιθανοφάνειας που εισήγαγε ο Cox.

### Συνάρτηση Μερικής Πιθανοφάνειας:

Έστω  $n$  μονάδες ενδιαφέροντος,  $k$  εκ των οποίων δεν είναι δεξιά αποκομμένες. Αυτές έχουν διατεταγμένους σε αύξουσα σειρά χρόνους αποτυχίας  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$  και θεωρούμε σε πρώτο στάδιο ότι  $d_{(j)} = 1 \quad \forall j = 1, 2, \dots, k$ . Αν  $R_{(j)}$  είναι το σύνολο των μονάδων που βρίσκονται σε κίνδυνο αποτυχίας αμέσως πριν την χρονική στιγμή  $t_{(j)}$ , τότε προκύπτει η συνάρτηση μερικής πιθανοφάνειας:

$$L(\vec{b}) = \prod_{j=1}^k \frac{e^{(\vec{b}^T \vec{z}_j)}}{\sum_{i \in R_j} e^{(\vec{b}^T \vec{z}_i)}}$$

Μεγιστοποιώντας την παραπάνω συνάρτηση ή ισοδύναμα τον λογάριθμο της, προσδιορίζουμε τους συντελεστές των συμμεταβλητών,  $\vec{b}$ , και είμαστε σε θέση να προσαρμόσουμε ένα μοντέλο παλινδρόμησης Cox στην συνάρτηση διακινδύνευσης.

Συγκεκριμένα, έχουμε:

$$l(\vec{b}) = \ln(L(\vec{b})) = \sum_{j=1}^k \vec{b}^T z_j - \sum_{j=1}^k \ln\left(\sum_{i \in R_j} e^{\vec{b}^T z_i}\right)$$

με μερικές παραγώγους ως προς τις  $r = 1, 2, \dots, p$  συνιστώσες του  $p \times 1$  διανύσματος συντελεστών  $\vec{b}$ :

$$\frac{\partial l}{\partial b_r} = \sum_{j=1}^k \vec{z}_{jr} - \sum_{j=1}^k \left[ \frac{\sum_{i \in R_j} \vec{z}_{ir} e^{\vec{b}^T z_i}}{\sum_{i \in R_j} e^{\vec{b}^T z_i}} \right]$$

Στη συνέχεια, επιλύουμε το παρακάτω σύστημα ως προς τις συνιστώσες του  $\vec{b}$

$$\frac{\partial l}{\partial b_r} = 0 \quad \forall r = 1, 2, \dots, p$$

χρησιμοποιώντας αριθμητικές μεθόδους.

Μπορούμε να προσδιορίσουμε και το τυπικό σφάλμα των συντελεστών των συμμεταβλητών ως  $se(\hat{\vec{b}}) = (I^{-1}(\vec{b}))^{1/2}$ , όπου  $I$  είναι ο πίνακας πληροφορίας κατά Fisher,  $I_{rs}(\vec{b}) = -E\left[\frac{\partial^2 L(\vec{b})}{\partial b_r \partial b_s}\right]$ .

Τα στοιχεία  $(r, s)$  αυτά του πίνακα κατά Fisher στην περίπτωση του μοντέλου του Cox είναι:

$$-\frac{\partial^2 l}{\partial b_r \partial b_s} = \sum_{j=1}^k \sum_{i \in R_j} \vec{z}_{ir} \left[ \vec{z}_{is} - \frac{\sum_{l \in R_j} \vec{z}_{ls} e^{\vec{b}^T z_l}}{\sum_{l \in R_j} e^{\vec{b}^T z_l}} \right] \frac{e^{\vec{b}^T z_i}}{\sum_{l \in R_j} e^{\vec{b}^T z_l}}$$

Επισημαίνεται, τέλος, ότι η βασική συνάρτηση διακινδύνευσης  $\lambda_0(t)$  δεν εμφανίζεται πουθενά στην παραπάνω διαδικασία μεγιστοποίησης της συνάρτησης πιθανοφάνειας του μοντέλου του Cox. Αυτό εξηγεί το γεγονός ότι η παραπάνω συνάρτηση πιθανοφάνειας καλείται Μερική.

Επίσης, μπορούμε να προσδιορίσουμε τη μη-παραμετρική εκτιμήτρια Breslow της συνάρτησης  $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ , ως:

$$\hat{\Lambda}_0(t) = \sum_{t_{(j)} \leq t} \frac{d_j}{\sum_{i \in R_j} e^{\hat{\vec{b}}^T z_i}}$$

Γυρνώντας πίσω στην συνάρτηση Μερικής Πιθανοφάνειας, αν  $\exists j : d_{(j)} > 1$  τότε κάνοντας χρήση της προσέγγισης Breslow (Breslow method of handling ties), ο όρος  $\frac{e^{\vec{b}^T \vec{z}_j}}{\sum_{i \in R_j} e^{\vec{b}^T \vec{z}_i}}$  αντικαθίσταται από τον όρο  $\frac{e^{\vec{b}^T \vec{x}_j}}{(\sum_{i \in R_j} e^{\vec{b}^T \vec{z}_i})^{d_j}}$ , όπου  $\vec{x}_j = \sum_{k=1}^{d_j} \vec{z}_k$  με  $\vec{z}_k$  το διάνυσμα συμμεταβλητών της μονάδας  $k$ , με  $d_j$  αποτυχίες την χρονική στιγμή  $t_{(j)}$ .

### Διαγνωστικός Έλεγχος Schoenfeld:

Τέλος, μετά την προσαρμογή ενός μοντέλου Cox χρησιμοποιούμε τα υπόλοιπα Schoenfeld  $\hat{r}_j = \vec{z}_j - \hat{E}(\vec{z}|R_j)$  ως ένα διαγνωστικό έλεγχο για την υπόθεση των αναλογικών διακινδυνεύσεων (Proportional hazards assumption). Βασικό ρόλο στο ημι-παραμετρικό μοντέλο αναλογικής διακινδύνευσης του Cox παίζει το γεγονός ότι οι συντελεστές των συμμεταβλητών  $\vec{b}$  είναι υποτίθενται ότι είναι σταθεροί ως προς τον χρόνο. Υποθέτουμε δηλαδή ότι  $b_i(t) = b_i \forall t$ . Τότε, προσεγγιστικά, μπορεί να δειχθεί ότι  $E(\hat{r}_{ij}) \simeq b_i(t_{(j)}) - \hat{b}_i$  όπου  $b_i(t_{(j)})$  είναι ο συντελεστής της συμμεταβλητής  $i \in \{1, 2, \dots, p\}$  την χρονική στιγμή  $t_{(j)}$ .

Συνεπώς, για να ισχύει η αρχική μας υπόθεση αναλογικότητας κινδύνων αρκεί οι συντελεστές των συμμεταβλητών του μοντέλου να είναι ανεξάρτητοι του χρόνου, το οποίο μεταφράζεται στο ότι η γραφική παράσταση  $\hat{r}_{ij} + \hat{b}_i$  vs  $t_{(j)}$  πρέπει να προσεγγίζει μια οριζόντια γραμμή.

## 1.5 Εφαρμογή του μοντέλου του Cox

### 1.5.1 Πλήρες Μοντέλο

Θα μελετήσουμε και θα προσαρμόσουμε ένα μοντέλο Cox στα δεδομένα 228 ασθενών με προχωρημένο καρκίνο του πνεύμονα όπως μελετήθηκαν από το North Central Treatment Group (Loprinzi et al. , 1994). Τα δεδομένα αυτά σε μια πιο αναλυτική μορφή υπάρχουν στα σύνολα δεδομένων της R στην βιβλιοθήκη survival (Therneau, 2017) με το όνομα cancer. Μελετούμε τον χρόνο (time), σε μέρες, από την εισαγωγή των ασθενών στην κλινική μελέτη μέχρι το τέλος της παρακολούθησης του ασθενή. Η μεταβλητή status υποδηλώνει εάν η παρατήρηση είναι δεξιά αποκομμένη (status=1-δηλαδή ο ασθενής αποχώρησε από την μελέτη-) ή όχι (status=2-δηλαδή ο ασθενής απεβίωσε-).

Πέρα από αυτές τις μεταβλητές, στην μοντελοποίηση της συνάρτησης διακινδύνευσης θα χρησιμοποιήσουμε τέσσερις συμμεταβλητές, οι οποίες είναι:

- (1) age: ηλικία των ατόμων όταν προσήλθαν στη μελέτη
- (2) sex: δίτιμη μεταβλητή για το φύλο των ασθενών, male=1 και female=2
- (3) ph.ecog: ένα test με όνομα ECOG που παίρνει τιμές από το 0=good μέχρι 5=dead
- (4) ph.karno: το test Karnofsky με τιμές μεταξύ του 0=bad και 100=good

Καταγράφουμε τις παρατηρήσεις της προαναφερθείσας έρευνας για τους πρώτους έξι ασθενείς (Πίνακας 2):

Πίνακας 2: Πρώτες έξι παρατηρήσεις του συνόλου δεδομένων cancer στην βιβλιοθήκη survival της R.

ID	time	status	age	sex	ph.ecog	ph.karno
1	306	2	74	1	1	90
2	455	2	68	1	0	90
3	1010	1	56	1	0	90
4	210	2	57	1	1	90
5	883	2	60	1	0	100
6	1022	1	74	1	1	50

Για τα δεδομένα του Πίνακα 2, με την βοήθεια της R (Fox and Weisberg, 2011) προσαρμόζουμε στην συνάρτηση διακινδύνευσης το μοντέλο αναλογικής διακινδύνευσης Cox, το οποίο έχει την εξής μορφή:

$$\lambda(t|\vec{z}) = \lambda_0(t) \exp(b_1 \text{age} + b_2 \text{sex} + b_3 \text{ph.ecog} + b_4 \text{ph.karno})$$

Προχωρούμε στην προσαρμογή του ημι-παραμετρικού μοντέλου αναλογικής διακινδύνευσης Cox στην συνάρτηση διακινδύνευσης των δεδομένων μας (Πίνακας 3):

Πίνακας 3: Προσαρμογή του μοντέλου του Cox.

covariates	coef	exp(coef)	se(coef)	z-value	p-value
age	0.01284	1.01292	0.00940	1.37	0.17216
sex	-0.57196	0.56442	0.16922	-3.38	0.00072
ph.ecog	0.63213	1.88161	0.17610	3.59	0.00033
ph.karno	0.01255	1.01263	0.00952	1.32	0.18748

Likelihood ratio test=31.2 on 4 df, p=2.82e-06

Στον Πίνακα 3, για κάθε συμμεταβλητή δίνεται η τιμή του συντελεστή της (coef) αλλά και το τυπικό σφάλμα αυτού (se(coef)) όπως προέκυψαν από την μεγιστοποίηση της μερικής πιθανοφάνειας κατά Cox.

Καταλήγουμε, λοιπόν, στο εξής μοντέλο παλινδρόμησης κατά Cox για την συνάρτηση διακινδύνευσης:

$$\lambda(t|\vec{z}) = \lambda_0(t) \exp(0.01284age - 0.57196sex + 0.63213ph.ecog + 0.01255ph.karno)$$

Στην συνέχεια, λαμβάνουμε τις τιμές  $e^{\hat{b}_i}$ ,  $\forall i = 1, 2, 3, 4$  καθεμία εκ των οποίων εκφράζει την επίδραση που θα είχε η αύξηση μιας μονάδας της αντίστοιχης συμμεταβλητής στην συνάρτηση διακινδύνευσης, δεδομένου ότι οι υπόλοιπες συμμεταβλητές παραμένουν σταθερές.

Για παράδειγμα, για μια τυχαία μονάδα ενδιαφέροντος, αν οι συμμεταβλητές age, sex και ph.karno διατηρηθούν σταθερές και έχουμε αύξηση κατά μια μονάδα στην μεταβλητή ph.ecog τότε η νέα συνάρτηση διακινδύνευσης θα είναι:

$$\lambda_{new}(t|\vec{z}) = \lambda(t|\vec{z}) \exp(\hat{b}_3) = 1.88161\lambda(t|\vec{z})$$

Θα παρατηρηθεί, δηλαδή, μια αύξηση στην συνάρτηση διακινδύνευσης της τάξης του 88.161%.

Επίσης, δίνεται η τιμή του Likelihood Ratio Test:

$$H_0 : M_0 - \text{χωρίς συμμεταβλητές}$$

$$H_1 : M_1 - 4 \text{ συμμεταβλητές}$$

με ελεγχοςυνάρτηση  $-2(\hat{l}_0 - \hat{l}_1) \sim X_4^2$ , όπου  $\hat{l}_0$  και  $\hat{l}_1$  οι μεγιστοποιημένες συναρτήσεις λογαριθμοποιημένης πιθανοφάνειας των μοντέλων  $M_0$  και  $M_1$ , αντίστοιχα. Εν προκειμένω, προκύπτει η τιμή της ελεγχοςυνάρτησης  $X^2 = 31.2$ .

Με  $p\text{-value} = P(X_4^2 > X^2) = 2.8210^{-6}$ , απορρίπτουμε την μηδενική υπόθεση και καταλήγουμε στο συμπέρασμα ότι η προσαρμογή των τεσσάρων συμμεταβλητών βελτιώνει το παραπάνω μοντέλο.

Τέλος, λαμβάνουμε τις τιμές της ελεγχοσυνάρτησης του ελέγχου του Wald (z-value) και το p-value αυτού (p-value) για τον συντελεστή κάθε συμμεταβλητής. Έχουμε:

Wald Test:

$$H_0 : b_i = 0 \quad \forall i = 1, 2, 3, 4$$

$$H_1 : b_i \neq 0 \quad \forall i = 1, 2, 3, 4$$

Τότε, η ελεγχοσυνάρτηση  $z = \frac{\hat{b}_i}{se(\hat{b}_i)}$  η οποία σύμφωνα με τα ασυμπτωτικά θεωρήματα που αφορούν τις εκτιμήτριες μέγιστης πιθανοφάνειας, ακολουθεί την κατανομή  $N(0, 1)$  με  $p\text{-value} = P(|Z| > |z|)$ . Παρατηρώντας τον Πίνακα 3 διαπιστώνουμε ότι οι συμμεταβλητές sex και rh.ecog είναι στατιστικά σημαντικές για το μοντέλο μας σε επίπεδο σημαντικότητας του ελέγχου 95% με τιμές p-value αρκετά μικρότερες του 0.05. Επίσης, οι επιδράσεις τους στην συνάρτηση διακινδύνευσης είναι 0.56442 και 1.88161, αντίστοιχα. Αντίστοιχα με τον συντελεστή  $\exp(\hat{b}_3)$ , την ερμηνεία του οποίου δώσαμε παραπάνω, δεδομένων των άλλων παραγόντων σταθερών, η συνάρτηση διακινδύνευσης μιας γυναίκας (sex=2) είναι περίπου η μισή (56.442%) ενός άνδρα (sex=1).

## 1.5.2 Προσαρμογή εναλλακτικού μοντέλου

Στην συνέχεια, θα εφαρμόσουμε το μοντέλο παλινδρόμησης του Cox στην συνάρτηση διακινδύνευσης χρησιμοποιώντας ως συμμεταβλητές μόνο τις στατιστικά σημαντικές μεταβλητές sex και rh.ecog.

Πίνακας 4: Προσαρμογή μοντέλου Cox με δύο συμμεταβλητές.

	coef	exp(coef)	se(coef)	z-value	p-value
sex	0.576	0.168	-0.175	-3.30	0.00098
rh.ecog	1.627	0.112	0.375	4.34	1.4e-05
Likelihood ratio test=29 on 2 df, p=5.13e-07					

Το αποτέλεσμα που λαμβάνουμε είναι το αναμενόμενο, και από το Likelihood Ratio Test αλλά και από τους ελέγχους Wald των συντελεστών των συμμεταβλητών μας προκύπτει ότι το παραπάνω μοντέλο Cox προσαρμόζεται ικανοποιητικά στα δεδομένα μας.

Στην συνέχεια ελέγχουμε ποιο από τα μοντέλα

$$H_0 : M_0-2 \text{ συμμεταβλητές}$$

$$H_1 : M_1-4 \text{ συμμεταβλητές}$$

προσαρμόζεται καλύτερα στα δεδομένα μας.

Συνδυάζοντας τις τιμές των Likelihood Ratio ελέγχων που είδαμε παραπάνω και για τα δύο μοντέλα που μας ενδιαφέρουν σε σύγκριση με το μοντέλο χωρίς καμία συμμεταβλητή (null), προκύπτει τελικά  $-2(\hat{l}_0 - \hat{l}_1) = X^2 = 2.2$ , ώστε  $X^2 \sim X_2^2$ . Οπότε, προκύπτει  $p\text{-value} = P(X_2^2 > X^2) = 0.3328711$ . Το παραπάνω  $p\text{-value}$  είναι αρκετά μεγάλο και άρα δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση (μοντέλο με sex και ph.ecog) έναντι της εναλλακτικής (μοντέλο και με τις τέσσερις συμμεταβλητές).

### 1.5.3 Έλεγχος της υπόθεσης αναλογικότητας

#### Πλήρες Μοντέλο:

Θα εφαρμόσουμε τον διαγνωστικό έλεγχο αναλογικής διακινδύνευσης που αναφέραμε παραπάνω. Κατά την προσαρμογή ενός μοντέλου αναλογικής διακινδύνευσης (Cox) έχουμε θεωρήσει ότι οι συντελεστές των συμμεταβλητών μας είναι ανεξάρτητοι του χρόνου. Κάνοντας χρήση των υπολοίπων Schoenfeld είμαστε σε θέση να ελέγξουμε την παραπάνω υπόθεση.

Για τον έλεγχο:

$$H_0 : b_i(t) = b_i, \forall t \text{ και } \forall i = 1, 2, 3, 4$$

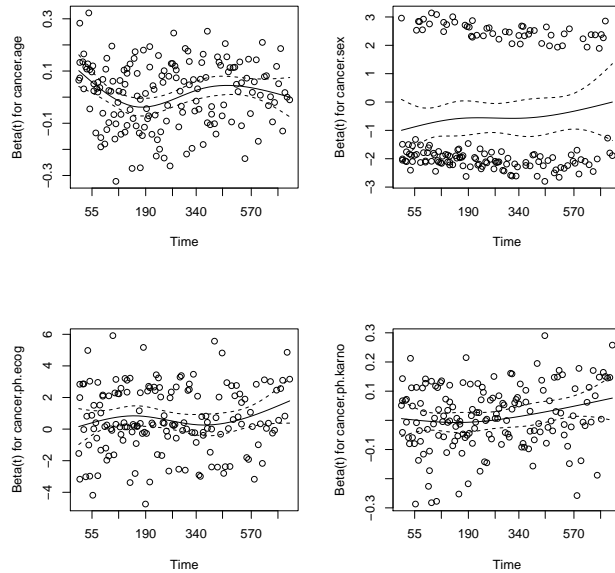
$$H_1 : b_i(t) \neq b_i, \forall t \text{ και } \forall i = 1, 2, 3, 4$$

Έχουμε:

Πίνακας 5: Υπόλοιπα Schoenfeld του μοντέλου με τις τέσσερις συμμεταβλητές.

	rho	chisq	p-value
age	0.0169	0.0505	0.8223
sex	0.0879	12.482	0.2639
ph.ecog	0.0604	0.5181	0.4717
ph.karno	0.1846	44.799	0.0343
Global	NA	80.446	0.0900





Σχήμα 3: Υπόλοιπα Schonfeld του μοντέλου με τις τέσσερις συµµεταβλητές.

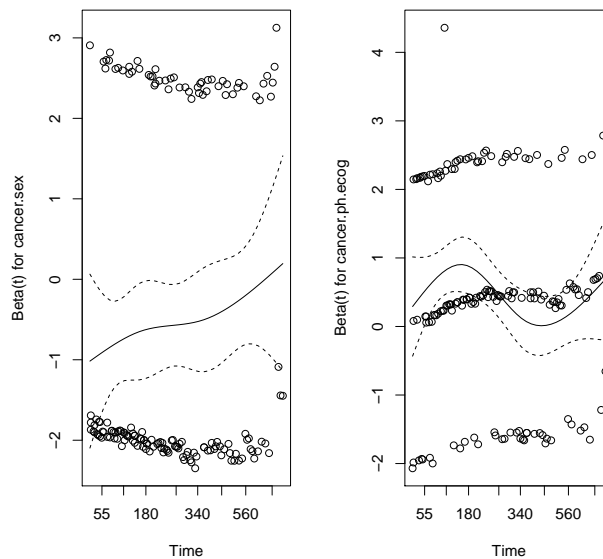
Στον Πίνακα 5 και στο Σχήμα 3 παρουσιάζεται τόσο ο ποσοτικός όσο και ο γραφικός έλεγχος, αντίστοιχα, της υπόθεσης αναλογικότητας των συντελεστών του μοντέλου του Cox που προσαρµόσαµε στην συνάρτηση διακινδύνευσης. Παρατηρούµε ότι, µε εξαίρεση την συµµεταβλητή *ph.karno*, οι υπόλοιπες έχουν αρκετά µεγάλα *p-value* και καµπύλες που προσεγγίζουν αρκετά κοντά τρεις οριζόντιες γραµµές. Άρα, δεν µπορούµε να απορρίψουµε σε επίπεδο σηµαντικότητας του ελέγχου 95% ότι οι συντελεστές των υπόλοιπων τριών συµµεταβλητών είναι ανεξάρτητοι του χρόνου. Εποµένως, η υπόθεση αναλογικότητας των κινδύνων είναι αρκετά ισχυρή στα παραπάνω δεδοµένα.

### Εναλλακτικό Μοντέλο:

Εφαρμόζουμε την ίδια ανάλυση στο μοντέλο παλινδρόμησης Cox με τις δύο συμμεταβλητές (age και ph.ecog) το οποίο, όπως είδαμε, έχει καλύτερη προσαρμογή στα δεδομένα μας.

Πίνακας 6: Υπόλοιπα Schoenfeld του μοντέλου με τις δύο συμμεταβλητές.

	rho	chisq	p-value
sex	0.124	2.47	0.1163
ph.ecog	-0.118	2.30	0.1291
Global	NA	4.64	0.0985



Σχήμα 4: Υπόλοιπα Schonfeld του μοντέλου με τις δύο συμμεταβλητές.

Και πάλι, από τις τιμές των p-value των ελέγχων αναλογικότητας διακινδύνευσης του Πίνακα 6 δεν μπορούμε να απορρίψουμε την μηδενική μας υπόθεση αναλογικότητας των κινδύνων.

Συμπερασματικά, η προσαρμογή και των δύο ημι-παραμετρικών μοντέλων αναλογικής διακινδύνευσης που προσαρμόσαμε παραπάνω στηρίζεται σημαντικά από τους διαγνωστικούς ελέγχους υπολοίπων Schoenfeld.

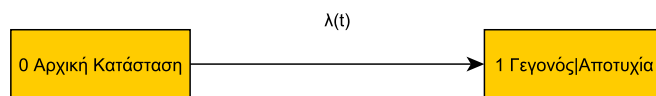
## 2 Ανταγωνιστικοί Κίνδυνοι

### 2.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο μελετήσαμε την διάρκεια ζωής μιας μονάδας ενδιαφέροντος μέχρι τελικά να αποτύχει, δηλαδή να συμβεί το γεγονός που μας ενδιαφέρει να μελετήσουμε, ή να διακοπεί η παρακολούθησή της (δεξιά αποκοπή). Αναφέραμε δύο τυπικές κατηγορίες ατελών παρατηρήσεων που συναντώνται σε τέτοιου είδους μελέτες, εκτιμήσαμε τις βασικές συναρτήσεις που εμπλέκονται στην ανάλυση επιβίωσης και μοντελοποιήσαμε την συνάρτηση διακινδύνευσης χρησιμοποιώντας το ημι-παραμετρικό μοντέλο αναλογικής διακινδύνευσης του Cox.

Βασική υπόθεση των παραπάνω ήταν η θεώρηση μιας μοναδικής αιτίας αποτυχίας των μονάδων ενδιαφέροντος μας (single-event ανάλυση επιβίωσης).

Σχηματικά, η μελέτη μας αφορούσε το εξής (Σχήμα 5):



Σχήμα 5: Ανάλυση Επιβίωσης με μοναδική αιτία αποτυχίας.

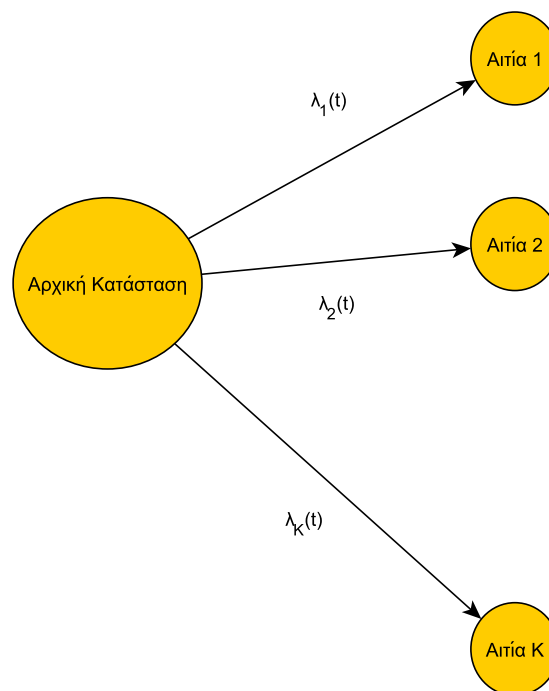
Η έννοια των ανταγωνιστικών κινδύνων συναντάται σε μελέτες όπου οι μονάδες ενδιαφέροντος βρίσκονται σε κίνδυνο αποτυχίας/θανάτου από περισσότερες της μιας, αμοιβαίως αποκλειόμενες αιτίες. Για παράδειγμα, σε κλινικές μελέτες για τον καρκίνο, ο θάνατος λόγω καρκίνου είναι το γεγονός που μας ενδιαφέρει να μελετήσουμε ενώ ο θάνατος από άλλες αιτίες, λόγω χάρη λόγω χειρουργικών επιπλοκών ή προχωρημένης ηλικίας αποτελούν ανταγωνιστικούς κινδύνους.

Υπάρχουν αρκετές προσεγγίσεις για την θεωρητική περιγραφή και ανάπτυξη του θέματος αυτού. Στην παρούσα εργασία θα μελετήσουμε τρεις εξ' αυτών, θα αναφέρουμε τα προτερήματα και τους περιορισμούς καθεμίας και θα εφαρμόσουμε δύο από αυτές σε δεδομένα μιας κλινικής μελέτης ασθενών που πάσχουν από squamous cell carcinoma/καρκίνωμα πλακωδών κυττάρων (Lagakos, 1978).

## 2.2 Μοντελοποίηση των προβλημάτων ανταγωνιστικών κινδύνων

Έστω  $K$  αμοιβαίως αποκλειόμενα αίτια αποτυχίας/θανάτου και έστω η συνεχής και θετική τυχαία μεταβλητή της διάρκειας ζωής  $T$ . Κάθε ένα από τα αίτια αυτά  $D = \{1, 2, \dots, K\}$  οδηγεί την μονάδα ενδιαφέροντος μας στον θάνατο και η μελέτη της σταματάει.

Σχηματικά, η ανάλυση μας σχετίζεται με το πρόβλημα του Σχήματος 6:



Σχήμα 6: Πρόβλημα ανταγωνιστικών κινδύνων με  $K$  αιτίες αποτυχίας.

Δεξιά αποκομμένες παρατηρήσεις εμφανίζονται και στα προβλήματα ανταγωνιστικών κινδύνων. Για παράδειγμα, ένας ασθενής μπορεί να αποχωρήσει από μια κλινική μελέτη προτού του συμβεί κάποια από τις  $K$  πιθανές και ανταγωνιζόμενες αιτίες θανάτου.

Αρχικά, ας ορίσουμε κάποιες συναρτήσεις οι οποίες θα μας φανούν πολύ χρήσιμες στην συνέχεια (Putter et al. , 2007):

Ανά αιτία (cause-specific) συνάρτηση διακινδύνευσης:

$$\lambda_k(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt, D = k | T \geq t)}{dt}$$

η οποία εκφράζει το στιγμιαίο ρυθμό αποτυχίας από την αιτία  $k$  παρουσία των υπολοίπων αιτιών.

Ανά αιτία (cause-specific) συνάρτηση σωρευτικής διακινδύνευσης:

$$\Lambda_k(t) = \int_0^t \lambda_k(s) ds$$

η οποία εκφράζει τον συνολικό αναμενόμενο αριθμό αποτυχιών από την αιτία  $k$  μέχρι την χρονική στιγμή  $t$  αν το γεγονός της αποτυχίας μπορεί να επαναλαμβανεται.

Επίσης, χρησιμοποιώντας το Θεώρημα Ολικής Πιθανότητας πάνω στις  $K$  αιτίες οι οποίες είναι αμετάβλητες αποκλειόμενες αφετέρου αποτελούν ολόκληρο το σύνολο των πιθανών και δυνατών αιτιών αποτυχίας της μονάδας ενδιαφέροντος μας, προκύπτουν τα εξής για τις (Prentice and Kalbfleisch, 1978):

$$\lambda_{total}(t) = \sum_{k=1}^K \lambda_k(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}$$

$$\Lambda_{total}(t) = \sum_{k=1}^K \Lambda_k(t) = \int_0^t \lambda_{total}(u) du$$

Οι παραπάνω συναρτήσεις διακινδύνευσης και σωρευτικής διακινδύνευσης είναι αυτές που θα προέκυπταν στο αντίστοιχο πρόβλημα ανάλυσης επιβίωσης με μοναδική αιτία αποτυχίας, θεωρώντας ότι δεν υπάρχει διαχωρισμός μεταξύ  $K$  αιτιών αποτυχίας.

Επίσης, κατ' αναλογία της ανάλυσης επιβίωσης με μοναδική αιτία αποτυχίας, όπου  $S(t) = \exp(-\Lambda(t))$ , θα έχουμε ότι:

$$S(t) = \exp\left(-\sum_{k=1}^K \Lambda_k(t)\right) = \exp(-\Lambda_{total}(t))$$

Η παραπάνω συνάρτηση εκφράζει την πιθανότητα η μονάδα ενδιαφέροντος μας να μην έχει αποτύχει από καμία αιτία μέχρι την χρονική στιγμή  $t$ .

Τέλος, ορίζουμε την Συνάρτηση Αθροιστικής Επίπτωσης ή Cumulative Incidence Function (CIF):

$$\begin{aligned} I_k(t) &= P(T \leq t, D = k) = \int_0^t \lambda_k(u) S(u) du \\ &= \int_0^t \lambda_k(u) \exp\left(-\sum_{k=1}^K \int_0^u \lambda_k(v) dv\right) du \end{aligned}$$

η οποία εκφράζει την πιθανότητα μια μονάδα ενδιαφέροντος να έχει αποτύχει πριν ή κατά την χρονική στιγμή  $t$  από την αιτία  $k \in D$ .

### 2.3 Μη-παραμετρικές μέθοδοι εκτίμησης

Στην συνέχεια, θα εφαρμόσουμε μη-παραμετρικές μεθόδους για να εκτιμήσουμε τις παραπάνω συναρτήσεις από τα δεδομένα μας. Η πιο συνηθισμένη μέθοδος που χρησιμοποιούμε για την εκτίμηση των παραπάνω μεγεθών και η οποία μάλιστα βασίζεται στις μη-παραμετρικές εκτιμήτριες που αναφέραμε στο Κεφάλαιο 1, είναι η παρακάτω (Marubini and Valsecchi, 1994).

Για κάθε αιτία αποτυχίας  $j \in D$  θεωρούμε όλες τις παρατηρήσεις οι οποίες έχουν αποτύχει από τις υπόλοιπες  $K - 1$  αιτίες, ως δεξιά αποκομμένες. Πλέον, δηλαδή, η ανάλυση της αιτίας  $j$  μετατρέπεται σε μια single-event ανάλυση επιβίωσης με αποτυχίες προερχόμενες από μια μοναδική αιτία αποτυχίας, παρουσία δεξιά αποκομμένων παρατηρήσεων. Οι εκτιμήτριες, δηλαδή, που θα χρησιμοποιήσουμε για τα ανά αιτία (cause-specific) μεγέθη θα είναι ίδιες με αυτές του Κεφαλαίου 1. Ωστόσο, θα έχουν μεγαλύτερα σφάλματα καθώς βασίζονται σε μη-ρεαλιστικές υποθέσεις. Πράγματι, η παραπάνω θεώρηση συμπεριφέρεται σε μονάδες ενδιαφέροντος που έχουν αποτύχει από άλλες, εκτός της  $j$ , αμοιβαίως αποκλειόμενες αιτίες, ως μονάδες οι οποίες κάποια στιγμή στο μέλλον θα αποτύχουν από την αιτία  $j$ . Στις παραπάνω συναρτήσεις, λόγω των προαναφερθεισών υποθέσεων, θα αναφερόμαστε ως Naive Kaplan-Meier, για να δείξουμε την μεροληψία τους. Μάλιστα, όσο περισσότερες οι αιτίες  $K$  τόσο μεγαλύτερη η μεροληψία των εκτιμητριών μας.

Έστω δείγμα  $n$  ασθενών με τους αντίστοιχους χρόνους αποτυχίας ή χρόνους δεξιάς αποκοπής τους και  $t_1 < t_2 < \dots < t_N$  οι διατεταγμένες σε αύξουσα σειρά  $N < n$  μη-αποκομμένες παρατηρήσεις χρόνων αποτυχίας, ανεξαρτήτως αιτίας. Επίσης, συμβολίζουμε με  $d_{kj}$  τον αριθμό των μονάδων ενδιαφέροντος που αποτυγχάνουν από την αιτία  $k \in D$  την χρονική στιγμή  $t_j$  όπου  $j \in \{1, 2, \dots, N\}$  και έστω  $d_j = \sum_{k=1}^K d_{kj}$  ο συνολικός αριθμός αποτυχιών από όλες τις αιτίες την χρονική στιγμή  $t_j$ . Επιπλέον, με  $n_j$  συμβολίζουμε τον αριθμό των μονάδων ενδιαφέροντος που δεν έχουν αποτύχει από καμία αιτία μέχρι και την  $t_j$ . Έτσι, προκύπτει η μη-παραμετρική εκτιμήτρια της συνάρτησης επιβίωσης, ανεξαρτήτως αιτίας αποτυχίας, Kaplan-Meier:

$$\hat{S}(t) = \hat{P}(T > t_j) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

Επίσης, ορίζουμε, όμοια με το Κεφάλαιο 1, την διακριτή μορφή της ανά αιτία συνάρτησης διακινδύνευσης:

$$\lambda_k(t_j) = P(T = t_j, D = k | T > t_{j-1})$$

και την εκτιμήτρια αυτής (Geskus, 2011):

$$\hat{P}(T = t_j, D = k | T > t_{j-1}) = \hat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j}$$

Χρησιμοποιώντας τις παραπάνω εκτιμήσεις, μπορούμε να γράψουμε την συνάρτηση Kaplan-Meier, χωρίς να γίνεται διάκριση μεταξύ των διαφορετικών αιτιών αποτυχίας, και με έναν ακόμα τρόπο:

$$\hat{S}(t) = \prod_{j:t_{(j)} \leq t} \left(1 - \sum_{k=1}^K \hat{\lambda}_k(t_j)\right)$$

Επίσης, μπορούμε να ορίσουμε την πιθανότητα αποτυχίας την χρονική στιγμή  $t_j$  από την αιτία  $k$ , ως:

$$p_k(t_j) = P(T = t_j, D = k)$$

και συνεχίζοντας προκύπτει ότι:

$$\begin{aligned} \hat{p}_k(t_j) &= \hat{P}(T = t_j, D = k, T > t_{j-1}) = \hat{P}(T = t_j, D = k | T > t_{j-1}) \hat{P}(T > t_{j-1}) \\ &= \hat{\lambda}_k(t_j) \hat{S}(t_{j-1}) \end{aligned}$$

Η διακριτή μορφή της συνάρτησης αθροιστικής επίπτωσης (CIF) δίνεται από τον τύπο:

$$\hat{I}_k(t) = \hat{P}(T \leq t, D = k) = \sum_{j:t_j \leq t} \hat{p}_k(t_j)$$

Τέλος, αν στα παραπάνω μοντέλα θεωρήσουμε ότι η διάρκεια ζωής των μονάδων ενδιαφέροντος και ο μηχανισμός δεξιάς αποκοπής είναι ανεξάρτητοι, η συνάρτηση πιθανοφάνειας παίρνει την εξής μορφή: (Andersen et al. , 2002)

$$L = \prod_{k=1}^k \prod_{i=1}^n \lambda_k(t_k)^{\delta_{ik}} \exp\left(-\int_0^{t_j} \lambda_k(u) du\right)$$

όπου

$$\delta_{ik} = \begin{cases} 1, & \text{η αιτία θανάτου της } i\text{-μονάδας ήταν η } k \\ 0, & \text{σε κάθε άλλη περίπτωση} \end{cases}$$

Παρατηρούμε ότι η μορφή της παραπάνω συνάρτησης εξαρτάται μόνο από τις  $K$  ανά αιτία συναρτήσεις διακινδύνευσης των  $K$  ανταγωνιζόμενων αιτιών. Είναι φανερό λοιπόν, η σημασία των συναρτήσεων  $\lambda_k(t)$  στα μοντέλα ανταγωνιστικών κινδύνων. Στις επόμενες παραγράφους θα επικεντρωθούμε στην μοντελοποίηση των παραπάνω συναρτήσεων, οι οποίες όπως είδαμε έχουν θεμελιώδη σημασία στην ανάλυση επιβίωσης.

## 2.4 Λανθάνοντες Χρόνοι Αποτυχίας

Θα περιγράψουμε την πρώτη ιστορικά προσέγγιση που έγινε στην ανάλυση και μοντελοποίηση δεδομένων ανταγωνιστικών κινδύνων, η οποία αν και αποτελεί λογική συνέχεια της ανάλυσης του πρώτου κεφαλαίου, έχει, όπως θα δούμε, πολλούς πρακτικούς περιορισμούς. Η βασική υπόθεση που κάνουμε είναι η θεώρηση  $K$  λανθανόντων χρόνων αποτυχίας/latent failure times, μια για κάθε πιθανή αιτία. Οι προαναφερθείσες  $K$  τιμές αποτελούν τους υποθετικούς χρόνους αποτυχίας των μονάδων ενδιαφέροντος θεωρώντας ότι όλες οι υπόλοιπες ανταγωνιζόμενες αιτίες αποτυχίας δεν ήταν παρούσες. Η παραπάνω προσέγγιση, δηλαδή, βασίζεται στην υπόθεση ότι όλες οι αιτίες αποτυχίας είναι ανεξάρτητες μεταξύ τους, γεγονός που σημαίνει ότι και οι λανθάνοντες χρόνοι αποτυχίας είναι επίσης ανεξάρτητοι. Συμβολίζουμε τους χρόνους αυτούς ως  $\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_K$  και στο πείραμα μας παρατηρούμε είτε τον  $T = \min\{\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_K\}$  είτε τον χρόνο δεξιάς αποκοπής  $C = c$  καθώς επίσης και την ακριβή αιτία αποτυχίας  $k \in D = \{1, 2, \dots, K\}$  (Arjas and Eerola, 1993).

Όπως στο προηγούμενο κεφάλαιο, έτσι και στην ανάλυση ανταγωνιστικών κινδύνων, θεωρούμε ότι ο μηχανισμός των δεξιά αποκομμένων παρατηρήσεων  $C$  είναι ανεξάρτητος των  $K$  γεγονότων.



Στην παραπάνω θεώρηση, λοιπόν, ορίζουμε την κοινή συνάρτηση επιβίωσης των αιτιών αποτυχίας αλλά και τις αντίστοιχες περιθώριες (Fine and Lindqvist, 2014):

$$\begin{aligned}\bar{S}(t_1, t_2, \dots, t_K) &= P(\tilde{T}_1 > t_1, \dots, \tilde{T}_K > t_K) \\ S^{(k)}(t) &= P(\tilde{T}_k > t) = \bar{S}(0, 0, \dots, t_k, \dots, 0) \quad \forall k = 1, 2, \dots, K\end{aligned}$$

Επίσης, οι ανά αιτία συναρτήσεις διακινδύνευσης δίνονται από τον τύπο (Tsiatis, 1975):

$$\lambda_k(t) = -\frac{\partial \log \bar{S}(t, t, \dots, t)}{\partial t_k}$$

Επομένως, με βάση και την παραπάνω μορφή της συνάρτησης πιθανοφάνειας του Andersen, δοσμένης της κοινής συνάρτησης επιβίωσης μπορούμε να βρούμε τις  $\lambda_k(t)$  και άρα να προσδιορίσουμε την συνάρτηση πιθανοφάνειας κάθε μοντέλου ανταγωνιστικών κινδύνων όπου οι χρόνοι δεξιάς αποκοπής είναι ανεξάρτητοι με τους χρόνους αποτυχίας.

### Μειονεκτήματα της μεθόδου:

Ωστόσο, όπως έδειξαν οι Tsiatis και Andersen, στην πράξη, η παραπάνω θεώρηση έχει δύο μεγάλα μειονεκτήματα που δεν μπορούν να αγνοηθούν.

**1)** Η θεμελιώδης υπόθεση ότι οι λανθάνοντες χρόνοι επιβίωσης είναι ανεξάρτητοι έχει δύο πολύ βασικά προβλήματα. Αφενός, περιορίζει πολύ τα προβλήματα τα οποία μπορούμε να αναλύσουμε καθώς για παράδειγμα στα βιοϊατρικά δεδομένα η ανεξαρτησία μεταξύ των αιτιών θανάτου ενός ασθενή σε μια κλινική μελέτη είναι εξαιρετικά σπάνιο φαινόμενο. Αφετέρου, δεν υπάρχει τρόπος να διαπιστώσουμε από τα δεδομένα μας αν η πολύ ισχυρή υπόθεση της ανεξαρτησίας των λανθανόντων χρόνων ισχύει ή όχι.

**2)** Για κάθε από κοινού συνάρτησης επιβίωσης με ανεξάρτητους λανθάνοντες χρόνους, υπάρχει μια διαφορετικά κοινή συνάρτηση επιβίωσης στην οποία οι εμπλεκόμενοι χρόνοι θα μπορούσαν να έχουν οποιαδήποτε εξάρτηση μεταξύ τους οι οποίες να δίνουν τις ίδιες ακριβώς ανά αιτία συναρτήσεις διακινδύνευσης (Tsiatis, 1975). Οι επιπτώσεις του παραπάνω είναι ότι δεν υπάρχει τρόπος να εκτιμήσουμε από τα δεδομένα μας ούτε την από κοινού συνάρτηση επιβίωσης αλλά ούτε και τις περιθώριες της.

Συμπερασματικά, η παραπάνω πρώτη, αυτή, θεώρηση για τα μοντέλα ανταγωνιστικών κινδύνων είναι πρακτικά αδύνατον να καρποφορήσει καθώς βασίζεται σε πολύ ισχυρές, μη ελέγξιμες από τα δεδομένα μας, υποθέσεις.

## 2.5 Προσαρμογή ημι-παραμετρικών μοντέλων

Όπως αναφέραμε και στην προηγούμενη παράγραφο, η μεγαλύτερη έμφαση στην ανάλυση των μοντέλων ανταγωνιστικών κινδύνων δόθηκε στην μοντελοποίηση των ανά αιτία συναρτήσεων διακινδύνευσης κάνοντας όσο το δυνατόν λιγότερες υποθέσεις εργασίας. Πράγματι, στα δύο ημι-παραμετρικά μοντέλα που θα παρουσιάσουμε στην συνέχεια, η ανεξαρτησία των αιτιών αποτυχίας δεν αποτελεί υπόθεση στην οποία βασιζόμαστε και ούτε καλούμαστε να την ελέγξουμε.

Έστω το  $p \times 1$  διάνυσμα συμμεταβλητών  $z_1, z_2, \dots, z_p$  και  $b_1, b_2, \dots, b_p$  οι συντελεστές τους. Για κάθε αιτία  $k \in D$  προσαρμόζουμε στην αντίστοιχη ανά αιτία συνάρτηση διακινδύνευσης ένα ημι-παραμετρικό μοντέλο αναλογικής διακινδύνευσης του Cox.

Η παραπάνω μεθοδολογία ονομάζεται στρωματοποιημένη/stratified παλινδρόμηση κατά Cox (Bakoyannis and Touloumi, 2012).

Οπότε, για κάθε πιθανή μετάβαση του Σχήματος 6, θα έχουμε:

$$\lambda_k(t|\vec{z}) = \lambda_{k,0}(t) \exp(\vec{b}_k^T \vec{z})$$

Η μοντελοποίηση γίνεται κατ' αναλογία με παραπάνω. Θεωρούμε, δηλαδή, ότι για κάθε  $k \in D$  οι αποτυχίες που οφείλονται σε άλλες αιτίες, εκτός την  $k$ , αποτελούν δεξιά αποκομμένες παρατηρήσεις του πρόβληματος επιβίωσης με μοναδική αιτία αποτυχίας που προκύπτει. Στα παραπάνω, μετασχηματισμένα δεδομένα προσαρμόζουμε ένα μοντέλο αναλογικής διακινδύνευσης Cox κάνοντας χρήση των τεχνικών που περιγράψαμε στο Κεφάλαιο 1.

Οι αντίστοιχες συναρτήσεις επιβίωσης και αθροιστικής επίπτωσης (CIF) αποκτούν της εξής μορφή:

$$S(t) = \exp\left(-\sum_{k=1}^K \int_0^t \lambda_{k,0}(u) \exp(\vec{b}_k^T \vec{z}) du\right)$$

$$I_k(t) = \int_0^t \lambda_{k,0}(u) \exp(\vec{b}_k^T \vec{z}) \exp\left(-\sum_{k=1}^K \int_0^u \lambda_{k,0}(v) \exp(\vec{b}_k^T \vec{z}) dv\right) du$$

Η συνάρτηση  $\lambda_{k,0}(t)$  λέγεται ανά αιτία βασική (cause-specific baseline) και είναι πάλι, μια ευέλικτη, μη-παραμετρική συνάρτηση. Ένα πολύ σημαντικό πλεονέκτημα αυτής της θεώρησης είναι ότι βασίζεται στα μοντέλα παλινδρόμησης Cox για τα οποία η βιβλιογραφία της ανάλυσης επιβίωσης όσο και οι αντίστοιχες βιβλιοθήκες της R έχουν πλούσιο υλικό.

Υπό το παραπάνω πλαίσιο μπορούν να γίνουν δύο έλεγχοι υποθέσεων οι οποίοι ενδεχομένως να απλοποιήσουν σημαντικά τα παραπάνω  $K$  μοντέλα  $p$  παραμέτρων έκαστος.

- 1.. Μπορεί να γίνει έλεγχος της υπόθεσης της ισότητας κάποιων εκ των συντελεστών των αντίστοιχων συμμεταβλητών σε διαφορετικές αιτίες αποτυχίας.
2. Επίσης, είναι δυνατόν να ελεγχθεί η υπόθεση αναλογικότητας διακινδύνευσης μεταξύ συναρτήσεων διακινδύνευσης διαφορετικών αιτιών, θεωρώντας ανάλογες ανά αιτία βασικές συναρτήσεις.

Πέρα των παραπάνω μεθόδων μείωσης των παραμέτρων ενός προβλήματος ανταγωνιστικών κινδύνων, έχει αναπτυχθεί και μια πολύ σημαντική αλγοριθμική μέθοδος, reduced rank, η οποία μειώνει τον αριθμό των  $p \times K$  συνολικών παραμέτρων, σε λιγότερες και άρα πιο διαχειρίσιμες.

### 2.5.1 Μέθοδος Reduced Rank

Έστω ένα πρόβλημα ανταγωνιστικών κινδύνων στο οποίο μοντελοποιούμε τις ανά αιτία συναρτήσεις διακινδύνευσης χρησιμοποιώντας μοντέλα παλινδρόμησης κατά Cox με  $K$  ανταγωνιζόμενες αιτίες αποτυχίας και  $p$  συμμεταβλητές έκαστος. Έτσι, ο πίνακας των παραμέτρων ενός τέτοιου μοντέλου -full rank matrix- είναι ένας πίνακας  $p \times K$ . Ο σκοπός των μεθόδων reduced rank σε μοντέλα αναλογικής διακινδύνευσης (εν προκειμένω σε μοντέλα Cox) είναι η μείωση του αριθμού των παραμέτρων που πρέπει να εκτιμηθούν διατηρώντας ταυτόχρονα την στρωματοποίηση τους (στρωματοποιημένη παλινδρόμηση κατά Cox/stratified Cox regression). Οι νέοι παράμετροι προς εκτίμηση, χρησιμοποιώντας την προαναφερθείσα μέθοδο, θα είναι  $R \times (p + K - R)$  αντί για  $p \times K$ , όπου  $R \leq \min(p, K)$  είναι η τάξη (rank) του νέου πίνακα των παραμέτρων παλινδρόμησης (Fiocco et al., 2006). Έστω ότι υπάρχει ένα σύνολο από  $R$  γραμμικούς μετασχηματισμούς ανεξάρτητων μεταβλητών μέσω των οποίων μπορεί να εκφραστεί η εξάρτηση κάθε μεταβλητή απόκρισης της παραπάνω μοντελοποίησης. Από αυτό συνεπάγεται ότι υπάρχουν  $R$  διανύσματα  $\vec{a}_r : r = 1, 2, \dots, R$  ώστε οι στατιστικές συναρτήσεις  $\vec{a}_1^T \vec{z}, \dots, \vec{a}_R^T \vec{z}$  να σχηματίζουν ένα επαρκές σύνολο παραμέτρων παλινδρόμησης για όλες τις μεταβλητές απόκρισης (Fiocco et al., 2008). Η παραπάνω θεώρηση είναι ισοδύναμη με τον μετασχηματισμό του  $p \times K$  πίνακα  $B = [\vec{b}_1 | \vec{b}_2 | \dots | \vec{b}_K]$  ως  $B = AC^T$ . Ο μετασχηματισμός αυτός δεν είναι μοναδικός, ωστόσο. Πράγματι, για κάθε πίνακα  $M$ ,  $R \times R$ , με  $\det(M) \neq 0$  ώστε  $MM^T = I_R$ , είναι  $B = AMM^{-1}C^T = (AM)(CM^{-1,T})^T$  είναι επίσης ένας μετασχηματισμός της παραπάνω μορφής.

Για την επίλυση του προβλήματος αυτού έχουν προταθεί από ερευνητές αρκετοί μετασχηματισμοί οι οποίοι επιβάλουν επιπλέον συνθήκες ώστε τελικά να προκύψει ένα μοναδικό σύνολο παραμέτρων για τους πίνακες  $A$  και  $C$ . Ωστόσο, τέτοιοι μετασχηματισμοί δεν θα μας απασχολήσουν σε αυτό το κείμενο, καθώς στην συνέχεια θα περιγράψουμε την αλγοριθμική επίλυση του παραπάνω προβλήματος. Οι πίνακες  $A = [\vec{a}_1 | \dots | \vec{a}_R]$  και  $C = [\vec{c}_1 | \dots | \vec{c}_R]$  έχουν διαστάσεις  $p \times R$  και  $K \times R$ , αντίστοιχα. Με βάση τα παραπάνω, οι ανά αιτία συναρτήσεις βασισμένες στην παλινδρόμηση κατά Cox, γίνονται:

$$\lambda_k(t|\vec{z}) = \lambda_{k,0}(t) \exp\left(\sum_{r=1}^R c_{kr} \vec{a}_r^T \vec{z}\right) \quad [1]$$

### Πλεονεκτήματα:

Δύο μεγάλα πλεονεκτήματα της παραπάνω μεθόδου είναι ότι για μικρότερες τάξεις (rank) μοντέλων είναι εφικτό να αντιμετωπίσουμε τις σπάνιες αιτίες αποτυχίας (αιτίες αποτυχίας από τις οποίες οι μονάδες ενδιαφέροντος αποτυγχάνουν σε πολύ μικρά ποσοστά) αλλά και να αποφευχθεί ο κίνδυνος της υπερ-πληθώρας μη στατιστικά σημαντικών παραμέτρων στα παραπάνω μοντέλα (over-fitting).

### Παράδειγμα:

Αναφέρουμε συνοπτικά, δύο από τις πιο απλές περιπτώσεις μοντέλων reduced rank:

#### Για $R=1$ :

$$\lambda_k(t|\vec{z}) = \lambda_{k,0}(t) \exp(c_k \vec{a}^T \vec{z})$$

Στο παραπάνω μοντέλο, όλες οι συμμεταβλητές έχουν την ίδια επίδραση σε κάθε μετάβαση με εξαίρεση τον συντελεστή αναλογικότητας  $c_k$  που εν γένει είναι διαφορετικός για κάθε αιτία αποτυχίας. Ο παράγοντας  $\vec{a}^T \vec{z}$  μπορεί να θεωρηθεί ως ένα προγνωστικό score για έναν ασθενή με διάνυσμα συμμεταβλητών  $\vec{z}$  το οποίο καθορίζει το πόσο πιθανό είναι ο ασθενής να βιώσει το γεγονός. Η παράμετρος  $c_k$  καθορίζει την βαρύτητα του παραπάνω προγνωστικού score για την μετάβαση/αιτία αποτυχίας  $k$ .

#### Για $R=2$ :

$$\lambda_k(t|\vec{z}) = \lambda_{k,0}(t) \exp(c_{k1} \vec{a}_1^T \vec{z} + c_{k2} \vec{a}_2^T \vec{z})$$

Στο μοντέλο αυτό, κάθε ασθενής έχει δύο προγνωστικά scores ( $\vec{a}_1^T \vec{z} + \vec{a}_2^T \vec{z}$ ) και δύο συντελεστές  $c_{k1}$  και  $c_{k2}$  οι οποίοι καθορίζουν την επίδραση των score αυτών στην  $k$ - μετάβαση/αιτία αποτυχίας.

### Αλγόριθμος της μεθόδου Reduced Rank:

Παρακάτω θα περιγράψουμε τον βασικό βηματικό αλγόριθμο (Fiocco et al. , 2005) που χρησιμοποιείται και από τα υπολογιστικά πακέτα, όπως η R, για την εκτίμηση των παραμέτρων παλινδρόμησης ενός μοντέλου reduced rank.

- 1) Ως αρχική εκτίμηση για τον  $C$  παίρνουμε τυχαίο πίνακα  $K \times R$  με  $\text{rank } R$ . Αποδεικνύεται ότι αν διαλέξουμε για αρχική προσέγγιση του  $C$  τον  $K \times (R - 1)$  πίνακα  $C_{R-1}$  με  $\text{rank } R - 1$  και μια επιπρόσθετη τυχαία στήλη η σύγκλιση του αλγορίθμου επιταχύνεται σημαντικά.
- 2) Για δεδομένο  $C$  με στοιχεία  $\hat{c}_{kr}$ , θα εκτιμήσουμε τον  $A$ . Πράγματι, κάνοντας χρήση της σχέσης [1] προκύπτει:

$$\lambda_k(t|\vec{z}) = \lambda_{k,0}(t) \exp\left(\sum_{r=1}^R \vec{a}_r^T \vec{W}_{kr}\right)$$

Πλέον, τον ρόλο των συμμεταβλητών παίζει το διάνυσμα  $\vec{W}_{kr} = \hat{c}_{kr}\vec{z}$  και χρησιμοποιώντας την παλινδρόμηση κατά Cox για κάθε μετάβαση  $k = 1, 2, \dots, K$  ξεχωριστά, μπορούμε να εκτιμήσουμε τα διανύσματα  $\vec{a}_r^T$ , τους συντελεστές δηλαδή των συμμεταβλητών.

- 3) Στην συνέχεια, έχοντας εκτιμήσει τις παραμέτρους  $\hat{\vec{a}}_r$  των προαναφερθέντων στρωματοποιημένων μοντέλων παλινδρόμησης Cox πάνω στις ανά αιτία συναρτήσεις διακινδύνευσης, χρησιμοποιώντας την σχέση [1] θα προκύψει ότι:

$$\lambda_k(t|\vec{z}) = \lambda_{k,0}(t) \exp\left(\sum_{r=1}^R c_{kr} \vec{V}_r\right)$$

όπου πλέον τα διανύσματα  $\vec{V}_r = \hat{\vec{a}}_r\vec{z}$  παίζουν τον ρόλο των συμμεταβλητών των παραπάνω μοντέλων και τα  $c_{kr}$  τον ρόλο των προς εκτίμηση παραμέτρων. Έτσι, με την ίδια λογική του βήματος 2, έχουμε μια καλύτερη εκτίμηση του πίνακα  $C$ .

- 4) Επαναλαμβάνουμε τα βήματα 2 και 3 μέχρι οι συνολικές/αθροιστικές λογαριθμοποιημένες πιθανοφάνειες των συντελεστών που εκτιμούμε στα παραπάνω βήματα -των  $\vec{a}_r^T \forall r = 1, 2, \dots, R$  στο βήμα 2 και των  $c_{kr} \forall k = 1, \dots, K$  και  $\forall r = 1, \dots, R$  στο βήμα 3 -να μην διαφέρουν περισσότερο από μια προκαθορισμένη τιμή ανοχής.

### Παρατηρήσεις:

1. Η σύγκλιση του παραπάνω αλγορίθμου δεν είναι εγγυημένη.
2. Κατά την εκτέλεση του αλγορίθμου είναι πιθανό, αν και αρκετά σπάνιο, οι λογαριθμοποιημένες πιθανοφάνειες των  $A$  και  $C$  να συγκλίνουν στο μηδέν ή να αποκλίνουν ταυτόχρονα. Το παραπάνω οδηγεί σε αριθμητική αστάθεια των πινάκων  $A$  και  $C$  αν και ο πίνακας  $B$  σε κάθε περίπτωση, λόγω μοναδικότητας, θα σταθεροποιείται.

## 2.5.2 Πιθανότητες Μετάβασης

Στην παράγραφο αυτή, θα μελετήσουμε τα μοντέλα ανταγωνιστικών κινδύνων από μια άλλη, πιθανοθεωρητική, σκοπιά. Μια σημαντική παράμετρος των παραπάνω μοντέλων είναι ο χρόνος, τον οποίο θεωρούμε συνεχή. Μετά το πέρας της ανάλυσης μας θα είμαστε σε θέση να εκτιμήσουμε μεγέθη όπως την πιθανότητα η μονάδα ενδιαφέροντος μας να έχει αποτύχει από την αιτία  $i \in D$  μέχρι και την μελλοντική χρονική στιγμή  $t$  ή την πιθανότητα η μονάδα μας να βρίσκεται στην αρχική της κατάσταση μέχρι μια μελλοντική χρονική στιγμή  $s$ .

Στα προαναφερθέντα μοντέλα, η μονάδα ενδιαφέροντος μας παραμένει για κάποιο τυχαίο χρόνο στην θέση 0 -αρχική κατάσταση (initial state)- προτού τελικά μεταβεί σε κάποια από τις  $K$  αμοιβαίως αποκλειόμενες απορροφητικές καταστάσεις ή σταματήσει η μελέτη της (δεξιά αποκοπή). Η θέση της μονάδας ενδιαφέροντος μας θα μπορούσε να αναπαρασταθεί ως μια δεξιά συνεχής στοχαστική ανέλιξη  $X(t) : t \geq 0$ , ώστε  $\lim_{h \rightarrow 0} P(X(t+h) = k | X(t) = 0) = \lambda_{0,k}(t)$ , με πεπερασμένο χώρο καταστάσεων  $\{0, 1, 2, \dots, K\}$  και αρχική κατανομή  $q = (1, 0, 0, \dots, 0)$ . Έτσι, μπορούμε να δούμε έναν πιο αυστηρό ορισμό για τις απορροφητικές καταστάσεις - έστω  $i$ - ως αυτές για τις οποίες αν υπάρχει  $s > 0$  ώστε  $X(s) = i$  θα είναι  $X(t) = i$  για κάθε μεταγενέστερη χρονική στιγμή  $t > s$  που μελετούμε την αλυσίδα μας. Μπορούμε να ορίσουμε την ανά αιτία συνάρτηση διακινδύνευσης και σωρευτική συνάρτηση διακινδύνευσης για την αιτία αποτυχίας  $k \in D$  και με τον εξής τρόπο:

$$\lambda_k(t) = \lambda_{0,k}(t) = \lim_{dt \rightarrow 0} P(X(t+dt) = k | X(t) = 0)$$

Η παραπάνω συνάρτηση διακινδύνευσης περιγράφει την πιθανότητα μετάβασης από την αρχική κατάσταση μιας μονάδας ενδιαφέροντος στην κατάσταση  $k \in D$  σε ένα απειροστό χρονικό διάστημα  $(t, t+dt)$  προς το μήκος του διαστήματος αυτού, ή πιο απλά, εκφράζει το στιγμιαίο ρυθμό μετάβασης από την αρχική κατάσταση στην κατάσταση  $k \in D$  παρουσία όλων των υπόλοιπων πιθανών εναλλακτικών μεταβάσεων.

Όσον αφορά την σωρευτική συνάρτηση διακινδύνευσης για κάθε αιτία ξεχωριστά, έχουμε:

$$\Lambda_k(t) = \Lambda_{0,k}(t) = \int_0^t \lambda_{0,k}(u) du$$

Επίσης, συμβολίζουμε με:

$$P_{0,k}(s, t) = P(X(t) = k | X(s) = 0)$$

την πιθανότητα η μονάδα ενδιαφέροντος μας να μεταβεί από την αρχική της κατάσταση στην οποία βρισκόταν έως και την χρονική στιγμή  $s$ , στην απορροφητική κατάσταση  $k \in D$  μέχρι και την χρονική στιγμή  $t \geq T = \inf(t \geq s : X(t) \neq 0)$ .

Είναι φανερό από τον παραπάνω ορισμό ότι:

$$I_k(t) = P_{0,k}(0, t) = P(X(t) = k | X(0) = 0) = P(T \leq t, D = k)$$

αλλά και:

$$\begin{aligned} S(t) = P_{0,0}(0, t) &= 1 - \sum_{k=1}^K P_{0k}(0, t) = \exp\left(-\sum_{k=1}^K \int_0^t \lambda_{0,k}(u) du\right) \\ &= \exp\left(-\sum_{k=1}^K \Lambda_{0,k}(t)\right) \end{aligned}$$

Γενικεύοντας, με βάση αυτές τις δύο υπο-περιπτώσεις οι οποίες μας δίνουν τις πολύ σημαντικές συναρτήσεις επιβίωσης και αθροιστικής επίπτωσης (CIF), έχουμε τα εξής:

$$P_{0,0}(s, t) = \exp\left(-\sum_{k=1}^K \int_s^t \lambda_{0,k}(u) du\right) = \frac{S(t)}{S(s)}$$

$$P_{0,k}(s, t) = \int_s^t P_{0,0}(s, u) \lambda_{0,k}(u) du = \int_s^t \frac{S(u)}{S(s)} \lambda_{0,k}(u) du$$

Η τελευταία σχέση εξηγείται ως εξής: Μας ενδιαφέρει να μελετήσουμε την πιθανότητα μετάβασης της μονάδας ενδιαφέροντος μας από την αρχική της κατάσταση, στην κατάσταση  $k \in D$ , στο χρονικό διάστημα  $[s, t]$ . Θεωρούμε ότι η μονάδα ενδιαφέροντος παραμένει στην αρχική της θέση μέχρι μια τυχαία χρονική στιγμή  $u \in (s, t)$  με πιθανότητα  $P_{0,0}(s, u)$  κατά την οποία στιγμιαία - λόγω δεξιάς συνέχειας της στοχαστικής μας ανέλιξης-μεταφέρεται με πιθανότητα  $\lambda_{0,k}(u)$  στην κατάσταση  $k$ .

### Εκτιμήσεις των πιθανοτήτων μετάβασης:

Έστω  $F_{0,k} \forall k \in D$  οι χρονικές στιγμές κατά τις οποίες παρατηρείται, στα δεδομένα μας, μετάβαση από την κατάσταση 0 στην κατάσταση  $k$  (jump times).

Τότε:

$$\hat{S}(u) = \prod_{r \in F_{0,k}: r \leq u} \left(1 - \sum_{k=1}^K \hat{\lambda}_{0,k}(r)\right)$$

η οποία αποτελεί την διακριτή προσέγγιση της  $S(u) = \exp\left(-\sum_{k=1}^K \Lambda_{0,k}(u)\right)$ .

Επίσης, έχουμε:

$$\hat{P}_{0,k}(s, t) = \sum_{r \geq s: r \in F_{0,k}} \hat{\lambda}_{0,k}(r) \frac{\hat{S}(r)}{\hat{S}(s)}$$

Η παραπάνω αποτελεί την διακριτή εκτιμήτρια της  $P_{0,k}(s, t) = \int_s^t \frac{S(u)}{S(s)} \lambda_{0,k}(u) du$ .

Για  $s = 0$  παίρνουμε την εκτιμήτρια της συνάρτησης αθροιστικής επίπτωσης:

$$\hat{I}_k(t) = \sum_{r \in F_{0,k}}^t \hat{\lambda}_{0,k}(r) \frac{\hat{S}(r)}{\hat{S}(s)}$$

Οι παραπάνω εκτιμήσεις αποτελούν, όπως θα δούμε στο Κεφάλαιο 3, ειδικές περιπτώσεις των εκτιμητριών Aalen-Johansen οι οποίες μπορούν να χρησιμοποιηθούν σε πολύ πιο περίπλοκα προβλήματα, παρουσία και ενδιάμεσων καταστάσεων.

## 2.6 Ημι-παραμετρικό μοντέλο των Fine και Gray

Στη single-event ανάλυση επιβίωσης που μελετήσαμε στο Κεφάλαιο 1, η συνάρτηση αθροιστικής επίπτωσης έχει μια αρκετά απλή,  $1 - 1$ , σχέση με την συνάρτηση διακινδύνευσης:

$$I(t) = 1 - S(t) = 1 - \exp\left(-\int_0^t \lambda(u) du\right)$$

Σύμφωνα με την μοντελοποίηση των ανά αιτία συναρτήσεων διακινδύνευσης, προκύπτει ότι:

$$I_k(t) = \int_0^t \lambda_k(u) \exp\left(-\sum_{k=1}^K \int_0^u \lambda_k(v) dv\right) du$$

Η παραπάνω σχέση είναι σημαντικά πιο περίπλοκη από αυτή της ανάλυσης επιβίωσης με μοναδική αιτία αποτυχίας. Μάλιστα, ενώ στην προαναφερθείσα περίπτωση η συνάρτηση αθροιστικής επίπτωσης αποτελεί αύξουσα συνάρτηση της συνάρτησης διακινδύνευσης, στην περίπτωση του παραπάνω μοντέλου, κάτι τέτοιο εν γένει δεν συμβαίνει. Παρ' όλα αυτά, έχειδειχθεί ότι η επίδραση, η ποσότητα  $\exp(b_i)$  της  $i$ -οστής συμμεταβλητής στην  $k$ -οστή ανά αιτία συνάρτηση διακινδύνευσης δηλαδή, δεδομένου ότι οι υπόλοιπες συμμεταβλητές παραμένουν σταθερές, βρίσκεται σε συμφωνία με την επίδραση της, στην συνάρτηση αθροιστικής επίπτωσης  $I_k(t)$  στην περίπτωση που η συμμεταβλητή αυτή είναι στατιστικά μη-σημαντική για τα μοντέλα Cox των υπόλοιπων  $K - 1$  συναρτήσεων διακινδύνευσης.

Εν γένει, ωστόσο, η συνάρτηση αθροιστικής επίπτωσης για το αίτιο  $k$  εξαρτάται τόσο από την συνάρτηση διακινδύνευσης του αιτίου  $k$  όσο και από τις συναρτήσεις διακινδύνευσης όλων των υπολοίπων. Δεν μπορούμε, δηλαδή, με την παραπάνω μοντελοποίηση να κάνουμε χρήση των εκτιμημένων συναρτήσεων διακινδύνευσης για να βγάλουμε άμεσα συμπεράσματα για την συμπεριφορά της συνάρτησης αθροιστικής επίπτωσης (CIF) κάθε αιτίου.



Για την επίλυση του παραπάνω προβλήματος και την εξαγωγή μιας πιο απλής σχέσης μεταξύ της συνάρτησης αθροιστικής επίπτωσης και της συνάρτησης διακινδύνευσης των  $k \in D$  αιτιών αποτυχίας, ορίζουμε την subdistribution συνάρτηση διακινδύνευσης (Fine and Gray , 1999):

$$\begin{aligned}\bar{\lambda}_k(t) &= \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt, D = k | T \geq t \cup (T \leq t \cap D \neq k))}{dt} \\ &= - \frac{d \log(1 - I_k(t))}{dt}\end{aligned}$$

Η συνάρτηση αυτή αποτελεί τον στιγμιαίο ρυθμό αποτυχίας λόγω της αιτίας  $k$  ανάμεσα στις μονάδες ενδιαφέροντος που είτε δεν έχουν αποτύχει ή έχουν αποτύχει από αιτία διαφορετική της  $k$  μέχρι και την χρονική στιγμή  $t$ . Είναι φανερό, πως σε αυτή την περίπτωση, αντίθετα με την ανά αιτία συνάρτηση διακινδύνευσης, το σύνολο των μονάδων που βρίσκονται σε κίνδυνο την χρονική στιγμή  $t$  είναι αφύσικο/unnatural. Αυτό συμβαίνει διότι το σύνολο μονάδων που βρίσκονται υπό τον κίνδυνο αποτυχίας λόγω της αιτίας  $k$  περιλαμβάνει άτομα τα οποία έχουν αποτύχει από αιτία διαφορετική και αμοιβαίως αποκλειόμενη με την  $k$  (Fine and Gray, 1999).

Η συνάρτηση αθροιστικής επίπτωσης χρησιμοποιώντας τις subdistribution συναρτήσεις διακινδύνευσης γράφεται ως:

$$I_k(t) = 1 - \exp\left(- \int_0^t \bar{\lambda}_k(u) du\right)$$

η οποία είναι μια  $1 - 1$  και αύξουσα συνάρτηση, όπως η συνάρτηση που συνδέει την συνάρτηση αθροιστικής επίπτωσης με την συνάρτηση διακινδύνευσης στην περίπτωση της ανάλυσης επιβίωσης με μοναδική αιτία αποτυχίας.

Επιπλέον, οι Fine και Gray πρότειναν την κατασκευή ενός μοντέλου παλινδρόμησης Cox πάνω στην νέα αυτή συνάρτηση διακινδύνευσης, ως εξής:

$$\bar{\lambda}_k(t|\vec{z}) = \bar{\lambda}_{k,0}(t) \exp(\vec{b}_k^T \vec{z})$$

όπου  $\vec{z}$  το διάνυσμα των συμμεταβλητών,  $\vec{b}$  το διάνυσμα των παραμέτρων τους και  $\bar{\lambda}_{k,0}(t)$  η μη-παραμετρική, ευέλικτη, βασική (baseline) συνάρτηση διακινδύνευσης.

Τέλος, συνδυάζοντας τις παραπάνω σχέσεις όπως προέκυψαν από τις ανά αιτία (cause-specific) και subdistribution συναρτήσεις διακινδύνευσης (Latouche et al. , 2007), στην περίπτωση των 2 ανταγωνιζόμενων αιτιών αποτυχίας, προκύπτει ότι:

$$\lambda_1(t|\vec{z}) = \left(1 + \frac{I_2(t|\vec{z})}{S(t|\vec{z})}\right) \bar{\lambda}_1(t|\vec{z})$$

### Μέθοδος IPCW:

Μάλιστα, οι Fine και Gray έδειξαν ότι για την περίπτωση των δεξιά αποκομμένων παρατηρήσεων, οι οποίες εμφανίζονται συνεχώς σε προβλήματα ανάλυσης επιβίωσης, η εκτίμηση των παραμέτρων των παραπάνω μοντέλων μεγιστοποιώντας την προαναφερθείσα συνάρτηση Μερικής Πιθανοφάνειας αποτελεί μια έγκυρη μεθοδολογία. Στην περίπτωση δεξιά αποκομμένων παρατηρήσεων αναπτύχθηκε η μέθοδος IPCW (Inverse Probability of Censoring Weighting) η οποία βασίζεται στην συνάρτηση πιθανοφάνειας που θα προέκυπτε αν είχαμε μη-αποκομμένα δεδομένα.

Η παραπάνω συνάρτηση Μερικής Πιθανοφάνειας IPCW ορίζεται ως εξής:

Έστω,  $N$  αποτυχίες, ανεξαρτήτως αιτίας, στις χρονικές στιγμές  $t_1 < t_2 < \dots < t_N$ , τότε προκύπτει:

$$L(\vec{b}_j) = \prod_{i=1}^N \frac{e^{\vec{b}_j^T \vec{z}_i}}{\sum_{l \in R_i} w_{il} e^{\vec{b}_l^T \vec{z}_i}}$$

Το σύνολο των μονάδων που βρίσκονται σε κίνδυνο αποτυχίας από την αιτία  $j$  την χρονική στιγμή  $t_i$  είναι το  $R_i = \{l : (t_l \geq t_i) \cup (t_l \leq t_i \cap D \neq j)\}$  όπου οι μονάδες ενδιαφέροντος που έχουν αποτύχει από αιτία διαφορετική της  $j$  παραμένουν στο σύνολο διακινδύνευσης (risk set). Τα βάρη της μεθόδου IPCW είναι  $w_{il} = \frac{G(t_i)}{G(\min(t_i, t_l))}$  ενώ αν  $t_l \geq t_i$ ,  $w_{il} = 1$ , όπου  $G(t)$  είναι η εκτιμήτρια Kaplan-Meier της συνάρτησης επιβίωσης της κατανομής αποκοπών (censoring distribution) η οποία είναι ουσιαστικά η συνολική πιθανότητα μια μονάδα ενδιαφέροντος να παρακολουθείται μέχρι και την χρονική στιγμή  $t$ . Δηλαδή, οι μονάδες που δεν αποτυγχάνουν από κάποια αιτία πριν την  $t_i$  τους δίνεται βάρος  $w_{il} = 1$  ενώ σε αυτές που έχουν αποτύχει από αιτία διαφορετική της  $j$  δίνεται βάρος  $w_{il}$  το οποίο είναι μια φθίνουσα συνάρτηση του χρόνου.

Συμπερασματικά, στην παραπάνω μέθοδο οι μονάδες που έχουν αποτύχει από μια άλλη εκτός της  $j$  αιτίας πριν την  $t_j$  παραμένουν στο σύνολο διακινδύνευσης - ορισμός της subdistribution συνάρτηση διακινδύνευσης- με ένα βάρος, ωστόσο, το οποίο εξαρτάται τόσο από την χρονική στιγμή αποτυχίας όσο και από την κατανομή του μηχανισμού αποκοπών.

### Εκτίμηση της Συνάρτησης Αθροιστικής Επίπτωσης (CIF):

Η συνάρτηση αθροιστικής επίπτωσης παίρνει πλέον την μορφή:

$$I_k(t|\vec{z}) = 1 - \exp\left(-\int_0^t \bar{\lambda}_{k,0}(u) \exp(\vec{b}_k^T \vec{z}) du\right) = 1 - \exp(-\bar{\Lambda}_{k,0}(t) \exp(\vec{b}_k^T \vec{z}))$$

Εφαρμόζουμε την μέθοδο μέγιστης πιθανοφάνειας για να βρούμε τις εκτιμήτριες των συντελεστών των συμμεταβλητών. Επίσης, χρησιμοποιούμε τροποποιήσεις των εκτιμητριών Breslow για τις βασικές συναρτήσεις σωρευτικής διακινδύνευσης.

Συνδυάζοντας τα παραπάνω (Beyersmann et al. , 2012), παίρνουμε:

$$\hat{I}_k(t|\vec{z}) = 1 - \exp(-\hat{\Lambda}_{k,0}(t) \exp(\hat{\vec{b}}_k^T \vec{z})) \quad \forall k \in D$$

και:

$$\hat{\Lambda}_{k,0} = \sum_{j=1}^N \frac{d_{kj}}{n_j}$$

όπου, όπως αναφέραμε και στην μη-παραμετρική μας ανάλυση,  $d_{kj}$  είναι το σύνολο των αποτυχιών λόγω της αιτίας  $k$  την χρονική στιγμή  $t_j$  και  $n_j$  ο συνολικός αριθμός των μονάδων που βρίσκονται σε κίνδυνο αποτυχίας από οποιαδήποτε αιτία πριν την  $t_j$  και  $t_1 < \dots < t_N$  οι διατεταγμένοι χρόνοι αποτυχίας των μη-αποκομμένων παρατηρήσεων.

### Περιορισμοί:

**1.** Παρατηρήσεις οι οποίες είναι αριστερά κολοβές ή και συμμεταβλητές  $\vec{z}$  οι οποίες εξαρτώνται από τον χρόνο δεν έχουν ξεκάθαρη μοντελοποίηση χρησιμοποιώντας την παλινδρόμηση Fine and Gray. (Haller et al. , 2013)

**2.** Επίσης, το παραπάνω μοντέλο δεν μπορεί να χρησιμοποιηθεί σε πιο πολύπλοκα μοντέλα στα οποία υπάρχουν και μεταβατικές/transient καταστάσεις, τα οποία θα δούμε στο επόμενο κεφάλαιο. Με άλλα λόγια, το μοντέλο Fine and Gray είναι λιγότερο ευέλικτο από το μοντέλο παλινδρόμησης που βασίζεται στις ανά αιτία συναρτήσεις διακινδύνευσης.

### **Διαγνωστικός Έλεγχος:**

Ένας άμεσος τρόπος να ελέγξουμε την προσαρμογή των δύο ημι-παραμετρικών μοντέλων αναλογικής διακινδύνευσης Cox που κατασκευάσαμε - το ένα βασισμένο στην ανά αιτία συνάρτηση διακινδύνευσης και το άλλο στην subdistribution συνάρτηση διακινδύνευσης- είναι να συγκρίνουμε τις εκτιμημένες συναρτήσεις αθροιστικής επίπτωσης τους με αυτές που προκύπτουν από τις μη-παραμετρικές μεθόδους που αναφέραμε στην αρχή του κεφαλαίου.

### **Γενικό Συμπέρασμα:**

Αν ενδιαφερόμαστε να κατανοήσουμε πλήρως την στοχαστική συμπεριφορά ενός προβλήματος ανταγωνιστικών κινδύνων, τότε η ανάλυση των ανά αιτία συναρτήσεων διακινδύνευσης είναι απαραίτητη και αναπόφευκτη. Αν, από την άλλη, το ενδιαφέρον μας βρίσκεται κυρίως στην μελέτη της συμπεριφοράς της συνάρτησης αθροιστικής επίπτωσης, τότε ένα μοντέλο βασισμένο στις subdistribution συναρτήσεις διακινδύνευσης αποτελεί μια πιο εύκολη στην ερμηνεία εναλλακτική (Beyersmann et al. , 2007).

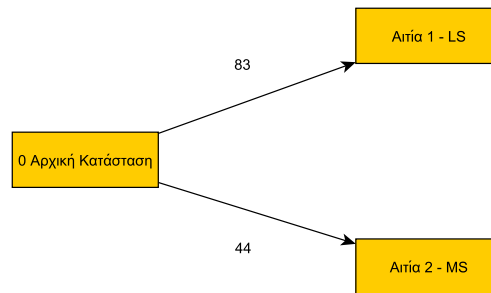
## 2.7 Εφαρμογή προβλήματος Ανταγωνιστικών Κινδύνων

### 2.7.1 Περιγραφή του προβλήματος

Θα μελετήσουμε δεδομένα από 194 ασθενείς με μαλπιγγιακό καρκίνωμα (κακοήθης όγκος) στην διάρκεια μιας κλινικής μελέτης της Eastern Cooperative Oncology Group για τον καρκίνο του πνεύμονα (Lagakos, 1978). Μελετήθηκαν δύο αίτια θανάτου ( $K = 2$ ), η τοπική εξάπλωση της ασθένειας (cause1-LS) και η μεταστατική εξάπλωση της ασθένειας (cause2-MS). Επιπλέον, θεωρούμε ότι υπάρχουν συνολικά τρεις παράγοντες ( $p = 3$ ) που θα μπορούσαν να επηρεάσουν τον χρόνο ζωής των ως προς μελέτη ασθενών.

Οι συμμεταβλητές μας είναι οι εξής:

- 1)  $z_1$ =Performance Status: Μια δίτιμη μεταβλητή που παίρνει την τιμή 0 αν η παρακολούθηση του ασθενή έγινε με περιπατητική εικοσιτετράωρη καταγραφή των στοιχείων του εκτός ιατρείου και 1 αν η παρακολούθησή του έγινε εντός του ιατρείου.
- 2)  $z_2$ =treatment: Μια, επίσης, δίτιμη τυχαία μεταβλητή που λαμβάνει την τιμή 0 αν χρησιμοποιήθηκε στον ασθενή η θεραπεία A και 1 αν χρησιμοποιήθηκε η B.
- 3)  $z_3$ =age: Χρησιμοποιήσαμε την διάμεσο της ηλικίας των ασθενών, σε χρόνια, και τελικά κάναμε τον διαχωρισμό  $z_3=0$  αν η ηλικία του ασθενή είναι μικρότερη ή ίση της διαμέσου (60 ετών) και  $z_3=1$  αν η ηλικία τους είναι μεγαλύτερη των 60.



Σχήμα 7: Θάνατοι λόγω καρκίνου του πνεύμονα στα δεδομένα του Lagakos

Όπως φαίνεται και από το Σχήμα 7, σε 83 (42.3%) από τους ασθενείς η αιτία θανάτου τους ήταν η τοπική εξάπλωση της ασθένειας ( $k = 1$ ), σε 44 (22.7%) η μεταστατική εξάπλωση ( $k = 2$ ) και για τις υπόλοιπες 67 (35%) δεν εκδηλώθηκε καμία εκ των δύο αιτιών αποτυχίας στην διάρκεια παρακολούθησής τους (δεξιά αποκομμένες). Θεωρούμε ότι οι χρόνοι αποκοπής είναι γνωστοί, καμία παρατήρηση δεν είναι αριστερά κολοβή και ο χρόνος ζωής των ασθενών είναι ανεξάρτητος του μηχανισμού αποκοπής.

Ας δούμε τις πρώτες έξι παρατηρήσεις των δεδομένων που θα χρησιμοποιήσουμε (Πίνακας 7):

Πίνακας 7: Δεδομένα από την μελέτη του Lagakos.

patnr	time	status	cause	performance	treatment	age
1	11	1	LS	1	0	1
2	11	2	MS	1	1	0
3	5	2	MS	1	0	0
4	24	2	MS	0	1	0
5	7	1	LS	0	1	0
6	17	2	MS	0	0	1

Οι μεταβλητές status και cause (Πίνακας 7) μας δείχνουν την αιτία αποτυχίας ή την δεξιά αποκοπή των μονάδων ενδιαφέροντος μας, αντίστοιχα. Εκτός από τον χρόνο, time, σε εβδομάδες, που οι ασθενείς μας βρίσκονται στην κλινική μελέτη, έχουμε και τις τιμές των τριών δίτιμων συμμεταβλητών που περιγράψαμε παραπάνω.

Επίσης, στον Πίνακα 8 παρουσιάζονται οι δυνατές μεταβάσεις του παραπάνω μοντέλου.

Πίνακας 8: Δυνατές Μεταβάσεις.

from\to	event-free	LS	MS
event-free	NA	1	2
LS	NA	NA	NA
MS	NA	NA	NA

Στην συνέχεια, για να προβούμε στην οποιαδήποτε ανάλυση των δεδομένων αυτών με την R (Putter, 2016) θα πρέπει να μετατρέψουμε τα δεδομένα από την μελέτη του Lagakos, σε μια επαυξημένη μορφή, όπου σε κάθε ασθενή αντιστοιχούν πλέον δύο γραμμές.

Ας δούμε στον Πίνακα 9 το νέο σύνολο δεδομένων για τους πρώτους τρεις ασθενείς:

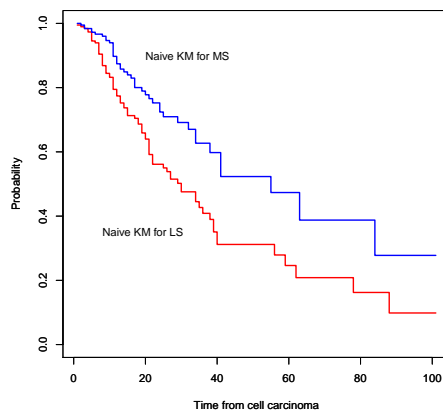
Πίνακας 9: Επαυξημένο Σύνολο Δεδομένων της μελέτης του Lagakos.

id	from	to	trans	Tstart	Tstop	time	status	performance
1	1	2	1	0	11	11	1	1
1	1	3	2	0	11	11	0	1
2	1	2	1	0	11	11	0	1
2	1	3	2	0	11	11	1	1
3	1	2	1	0	5	5	0	1
3	1	3	2	0	5	5	1	1
treatment	age	performance.1	performance.2	treatment.1	treatment.2	age.1	age.2	
0	1	1	0	0	0	1	0	
0	1	0	1	0	0	0	1	
1	0	1	0	1	0	0	0	
1	0	0	1	0	1	0	0	
0	0	1	0	0	0	0	0	
0	0	0	1	0	0	0	0	

## 2.7.2 Μη-παραμετρική ανάλυση

### Εκτιμήτριες Naive Kaplan-Meier:

Ξεκινάμε την μη-παραμετρική ανάλυση μας προσδιορίζοντας τις μη-παραμετρικές εκτιμήτριες Naive Kaplan-Meier, μία για κάθε αιτία θανάτου των ασθενών. Όπως είπαμε και παραπάνω, για κάθε αιτία  $k = 1, 2$ , η Naive Kaplan-Meier προσδιορίζεται, ουσιαστικά, θεωρώντας ένα πρόβλημα ανάλυσης επιβίωσης με μοναδική αιτία αποτυχίας, όπου οι χρόνοι αποτυχίας της άλλης αιτίας θεωρούνται ως δεξιά αποκομμένες τιμές (Σχήμα 8).



Σχήμα 8: Naive εκτιμήτριες Kaplan-Meier για τις δύο αιτίες αποτυχίας.

Παρατηρούμε ότι η πιθανότητα ένας ασθενής να παραμένει ζωντανός με το πέρας του χρόνου όταν η μοναδική αιτία αποτυχίας είναι η MS, θεωρώντας τους

θανάτους από LS ως δεξιά αποκομμένες παρατηρήσεις, είναι συστηματικά μεγαλύτερη από αυτή που προκύπτει θεωρώντας την LS ως την μοναδική αιτία αποτυχίας.

Το αποτέλεσμα αυτό ήταν αναμενόμενο λόγω των σχεδόν διπλάσιων (83 έναντι 44) θανάτων ασθενών που οφείλονται στην LS σε σχέση με την MS.

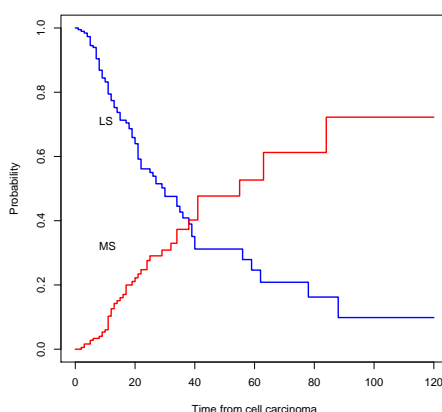
Για μερικές από τις χρονικές στιγμές καταγραφής θανάτου (είτε λόγω LS είτε λόγω MS) καταγράφονται παρακάτω οι τιμές των εκτιμητριών Naive Kaplan-Meier για τις δύο αιτίες θανάτου (Πίνακας 10):

Πίνακας 10: Naive Kaplan-Meier εκτιμήτριες για τις αποτυχίες λόγω LS και MS, αντίστοιχα.

time	KM for LS	KM for MS
1	0.99485862	1
2	0.98958087	0.9946950
3	0.98418807	0.9838832
84	0.16230205	0.2777129
88	0.09844117	0.2777129
101	0.09844117	0.2777129

Θα δείξουμε επίσης, ότι οι παραπάνω εκτιμήτριες για κάθε αιτία αποτυχίας, είναι μεροληπτικές. Αυτό συμβαίνει γιατί, όπως προείπαμε, βασίζονται στην μη-ρεαλιστική υπόθεση ότι ένας ασθενής που έχει αποτύχει από την άλλη, αμοιβαίως αποκλειόμενη, αιτία από αυτή που μελετούμε βρίσκεται σε κίνδυνο αποτυχίας από την αιτία μελέτης μας. Συνεπώς, για παράδειγμα ένας ασθενής που αποτυγχάνει λόγω MS, ενώ σίγουρα δεν θα αποτύχει από LS, αντιμετωπίζεται σαν να μπορούσε να αποτύχει από LS (δεξιά αποκομμένη τιμή) και έτσι η πιθανότητα αποτυχίας από LS υπερεκτιμάται (over-estimated).

Πράγματι, έχουμε (Σχήμα 9):



Σχήμα 9: Naive Kaplan-Meier για την LS και η συμπληρωματική αυτής για την αιτία MS.



Στο Σχήμα 9 σχεδιάσαμε την καμπύλη Naive Kaplan-Meier για την πρώτη αιτία αποτυχίας -LS- και την συμπληρωματική της Naive Kaplan-Meier για την δεύτερη αιτία αποτυχίας -MS-.

Στην περίπτωση της ανάλυσης επιβίωσης με μοναδική αιτία αποτυχίας, η εκτιμήτρια της πιθανότητας αποτυχίας μιας μονάδας ενδιαφέροντος δίνεται ως η συμπληρωματική πιθανότητα της αμερόληπτης Kaplan-Meier. Χρησιμοποιώντας τις εκτιμήτριες Naive Kaplan-Meier για τον υπολογισμό των πιθανοτήτων αποτυχίας από τις αιτίες LS και MS, αντίστοιχα, παρατηρούμε τα εξής:

Μετά από  $t = 37$  εβδομάδες οι καμπύλες των LS και MS τέμνονται, γεγονός που σημαίνει ότι οι εκτιμημένες πιθανότητες εμφάνισης των αιτιών αποτυχίας LS και MS έχουν άθροισμα πάνω από 1, το οποίο μας οδηγεί σε άτοπο.

Πράγματι, για  $t = 61$  οι εκτιμήσεις των πιθανοτήτων αποτυχίας από LS και MS είναι 0.754 και 0.527, αντίστοιχα, με άθροισμα  $1.281 > 1$ .

Ωστόσο, από το Θεώρημα Ολικής Πιθανότητας γνωρίζουμε ότι:

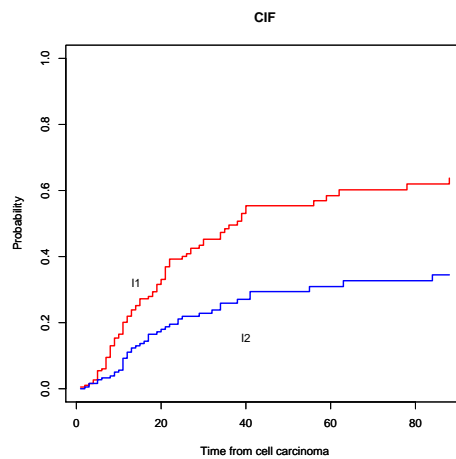
$$\hat{P}(T \leq 61 \cap D = 1) + \hat{P}(T \leq 61 \cap D = 2) = \hat{P}(T \leq 61) < 1$$

καθώς οι παραπάνω αιτίες είναι αμοιβαία αποκλειόμενες.

Είναι φανερό, δηλαδή, ότι οι εκτιμήτριες Naive Kaplan-Meier εμπεριέχουν σφάλματα στον υπολογισμό τους, τα οποία οφείλονται κατά κύριο λόγο στις προαναφερθείσες υποθέσεις εργασίας.

### Μη-παραμετρικές εκτιμήτριες των Συναρτήσεων Αθροιστικής Επίπτωσης:

Όσον αφορά τις εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης (CIF):  $\hat{I}_1(t)$  και  $\hat{I}_2(t)$ , έχουμε (Σχήμα 10):



Σχήμα 10: Εκτιμήσεις των CIF για τις αιτίες LS και MS, αντίστοιχα.

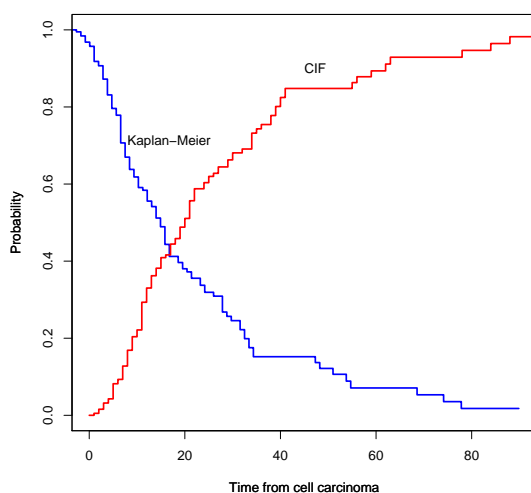
Από το Σχήμα 10 παρατηρούμε ότι η πιθανότητα αποτυχίας από την LS είναι συνεχώς μεγαλύτερη από αυτή της MS, γεγονός το οποίο διαφαίνεται και από τα δεδομένα μας, με 83 αποτυχίες λόγω LS και 44 λόγω MS.

Μάλιστα, σε συμφωνία και με τα παραπάνω δεδομένα, η εκτίμηση της συνάρτησης αθροιστικής επίπτωσης για την αποτυχία λόγω LS είναι συστηματικά περίπου διπλάσια της αντίστοιχης εκτιμήτριας για την αιτία MS.

### 2.7.3 Σύγκριση του προβλήματος ανταγωνιστικών κινδύνων με το αντίστοιχο single-event

Ας δούμε τώρα τις εκτιμήσεις των συναρτήσεων επιβίωσης και αθροιστικής επίπτωσης, στην περίπτωση μετατροπής του προβλήματος σε single-event, με μοναδική αιτία αποτυχίας (Σχήμα 11).

Εν προκειμένω, δηλαδή, δεχόμαστε ότι δεν υπάρχει διαχωρισμός μεταξύ των δύο αιτιών αποτυχίας.



Σχήμα 11: Kaplan-Meier και CIF, συγχωνεύοντας τις αιτίες θανάτου σε μία.

Στην συνέχεια, θα δείξουμε την ακρίβεια των παραπάνω μη-παραμετρικών μεθοδολογιών στα προβλήματα ανταγωνιστικών κινδύνων και την στενή σύνδεση τους με την μη-παραμετρική ανάλυση επιβίωσης με μοναδική αιτία αποτυχίας.

Θεωρητικά, ισχύει ότι:

$$P(T \leq t) = P(T \leq t \cap D = 1) + P(T \leq t \cap D = 2)$$

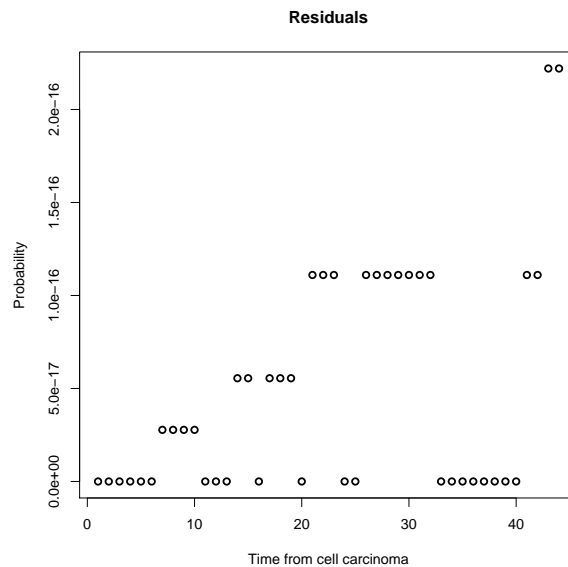
λόγω του Θεωρήματος Ολικής Πιθανότητας, όπως εξηγήσαμε παραπάνω.

Θα ελέγξουμε το κατά πόσο ισχύει η αντίστοιχη ισότητα αλλά για τα εκτιμημένα μεγέθη, χρησιμοποιώντας για το αριστερό μέλος την εκτιμήτρια ενός single-event προβλήματος και στο δεξί μέλος τις αντίστοιχες εκτιμήτριες ενός προβλήματος ανταγωνιστικών κινδύνων.

Σχεδιάζουμε τις απόλυτες τιμές των διαφορών:

$$|\hat{P}(T \leq t) - \hat{P}(T \leq t \cap D = 1) - \hat{P}(T \leq t \cap D = 2)|$$

και έχουμε (Σχήμα 12):



Σχήμα 12:  $|\hat{P}(T \leq t) - \hat{P}(T \leq t \cap D = 1) - \hat{P}(T \leq t \cap D = 2)|$ .

Από το Σχήμα 12 παρατηρούμε ότι οι διαφορές των ποσοτήτων  $\hat{P}(T \leq t)$  και  $\hat{P}(T \leq t \cap D = 1) + \hat{P}(T \leq t \cap D = 2)$  είναι εξαιρετικά μικρές, γεγονός που αποτελεί πολύ καλή ένδειξη για την ακρίβειά τους. Επίσης, το γεγονός ότι οι διαφορές των προαναφερθέντων εκτιμήσεων είναι πολύ μικρές, υποδεικνύει και τη λογική συνέχεια με την οποία οι μη-παραμετρικές εκτιμήτριες στα προβλήματα ανταγωνιστικών κινδύνων προέρχονται, κάνοντας τις προαναφερθείσες υποθέσεις, από τις αντίστοιχες εκτιμήτριες των προβλημάτων επιβίωσης με μοναδική αιτία αποτυχίας.

## 2.7.4 Ημι-παραμετρική ανάλυση

Στην συνέχεια, θα μελετήσουμε το παραπάνω πρόβλημα ανταγωνιστικών κινδύνων χρησιμοποιώντας τα προαναφερθέντα μοντέλα παλινδρόμησης Cox βασισμένα στις ανά αιτία (cause specific) συναρτήσεις διακινδύνευσης:

$$\lambda_1(t|\vec{z}) = \lambda_{1,0}(t)\exp(b_1^1 z_1 + b_2^1 z_2 + b_3^1 z_3)$$

$$\lambda_2(t|\vec{z}) = \lambda_{2,0}(t)\exp(b_1^2 z_1 + b_2^2 z_2 + b_3^2 z_3)$$

### Προσαρμογή μοντέλου Cox για την $\lambda_1(t|\vec{z})$ :

Θεωρούμε την αποτυχία των ασθενών λόγω LS ως το γεγονός που μας ενδιαφέρει να μελετήσουμε, ενώ οι θάνατοι λόγω MS θεωρούνται ως δεξιά αποκομμένες παρατηρήσεις. Στο τροποποιημένο αυτό σύνολο δεδομένων προσαρμόζουμε ένα μοντέλο Cox  $\lambda_1(t|\vec{z})$  (Πίνακας 11):

Πίνακας 11: Προσαρμογή μοντέλου Cox στην  $\lambda_1(t|\vec{z})$ :

	coef	exp(coef)	se(coef)	z-value	p-value
performance	0.5760	1.7788	0.2332	2.47	0.014
treatment	0.3327	1.3809	0.2743	1.1	0.239
age	-0.0247	0.9756	0.2999	-0.08	0.934

Likelihood ratio test=6.53 on 3 df, p=0.0886

Μετά την προσαρμογή του μοντέλου Cox στα τροποποιημένα μας δεδομένα (Πίνακας 11), παρουσιάζονται στο παραπάνω πίνακα οι εκτιμήσεις των παραμέτρων των τριών συμμεταβλητών που έχουμε θεωρήσει.

Έλεγχος Wald για την παράμετρο της  $i$ -οστής συμμεταβλητής:

$$H_0 : b_i^1 = 0$$

$$H_1 : b_i^1 \neq 0$$

Είναι  $p - value = P(|Z| > |z|)$ , όπου  $z = \frac{\hat{b}_i^1}{se(\hat{b}_i^1)}$  και  $Z \sim N(0, 1)$ .

Παρατηρούμε ότι μόνο η συμμεταβλητή performance ( $z_1$ ) είναι στατιστικά σημαντική για την συνάρτηση διακινδύνευσης  $\lambda_1(t|\vec{z})$ . Επίσης, για τιμή performance=1, με τις τιμές των υπόλοιπων συμμεταβλητών να παραμένουν σταθερές, η συνάρτηση διακινδύνευσης  $\lambda_1(t|\vec{z})$  θα αυξηθεί κατά 77.88%.

### Έλεγχος υπόθεσης αναλογικής διακινδύνευσης:

Επιπλέον, εφαρμόζουμε τον διαγνωστικό έλεγχο Schoenfeld για τους συντελεστές των συμμεταβλητών του παραπάνω μοντέλου (Πίνακας 12):

Πίνακας 12: Υπόλοιπα Schoenfeld για την προσαρμογή του  $\lambda_1(t|\vec{z})$ .

	rho	chisq	p-value
performance	-0.176	2.53	0.1118
treatment	-0.125	1.33	0.2489
age	0.205	3.62	0.0570
Global	NA	6.76	0.0799

Για τον έλεγχο:

$$H_0 : b_i \text{ Σταθερή ως προς τον χρόνο } \forall i = 1, 2, 3$$

$$H_1 : b_i \text{ Μη σταθερή ως προς τον χρόνο } \forall i = 1, 2, 3$$

Από τον Πίνακα 12 προκύπτει ότι σε επίπεδο σημαντικότητας του ελέγχου 95%, με οριακή εξαίρεση την συμμεταβλητή age, δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση αναλογικότητας διακινδύνευσης για το παραπάνω μοντέλο.

### Προσαρμογή εναλλακτικού μοντέλου για την $\lambda_1(t|\vec{z})$ :

Λόγω του όχι τόσο μικρού p-value στο Likelihood Ratio Test (0.0886) στην προσαρμογή του  $\lambda_1(t|\vec{z})$ , αποφασίζουμε να προσαρμόσουμε ένα μοντέλο Cox στα ίδια δεδομένα, αλλά χρησιμοποιώντας μόνο την συμμεταβλητή performance (Πίνακας 13):

Πίνακας 13: Προσαρμογή μοντέλου Cox στην  $\lambda_1(t|\vec{z})$ , με μόνη συμμεταβλητή την performance.

	coef	exp(coef)	se(coef)	z-value	p-value
performance	0.522	1.686	0.228	2.29	0.022

Likelihood ratio test=5.05 on 1 df, p=0.0246

Από τον Πίνακα 13 παρατηρούμε ότι για τον συνολικό έλεγχο του μοντέλου προκύπτει p-value=0.022 και άρα το παραπάνω μοντέλο Cox δείχνει να είναι ικανοποιητικό για τα δεδομένα μας. Στην συνέχεια θα χρησιμοποιήσουμε τα αποτελέσματα των Likelihood Ratio Test των δύο παραπάνω μοντέλων για να κάνουμε τον έλεγχο

$$H_0 : M_0 : b_2 = b_3 = 0$$

$$H_1 : M_1 : |b_2| + |b_3| \neq 0$$

Τότε,  $X^2 = -2(\hat{l}_0 - \hat{l}_1) = 1.48 \sim X_2^2$ , με p-value =  $P(X_2^2 > 1.48) = 0.4771139$ . Επομένως, σε επίπεδο σημαντικότητας του ελέγχου 95% δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση έναντι της εναλλακτικής, και άρα καταλήγουμε

στο συμπέρασμα ότι το μοντέλο με μόνη συμμεταβλητή την performance είναι καλύτερο από το αρχικό μοντέλο των τριών συμμεταβλητών.

**Έλεγχος υπόθεσης αναλογικής διακινδύνευσης του εναλλακτικού μοντέλου:**

Ακόμα, για τον διαγνωστικό έλεγχο Schoenfeld του παραπάνω μοντέλου μιας παραμέτρου, έχουμε (Πίνακας 14) :

Πίνακας 14: Υπόλοιπα Schoenfeld για την προσαρμογή του  $\lambda_1(t|\vec{z})$ , με μόνη συμμεταβλητή την performance.

	rh0	chisq	p-value
performance	-0.178	2.55	0.11

Δεν μπορούμε και πάλι, από τον Πίνακα 14, να απορρίψουμε την υπόθεση αναλογικότητας διακινδύνευσης της μεταβλητής performance, και άρα του παραπάνω μοντέλου, σε επίπεδο σημαντικότητας του ελέγχου 95%.

**Προσαρμογή μοντέλου Cox για την  $\lambda_2(t|\vec{z})$  :**

Ομοίως, θα προσαρμόσουμε ένα μοντέλο Cox  $\lambda_2(t|\vec{z})$  σε μια single-event ανάλυση με μοναδική αιτία αποτυχίας την MS, θεωρώντας τις αποτυχίες από LS, ως δεξιά αποκομμένες (Πίνακας 15):

Πίνακας 15: Προσαρμογή μοντέλου Cox στην  $\lambda_2(t|\vec{z})$ :

	coef	exp(coef)	se(coef)	p-value	z-value
performance	0.5909	1.8056	0.3227	1.83	0.067
treatment	0.4574	1.5799	0.3900	1.17	0.241
age	-0.0957	0.9088	0.4062	-0.24	0.814
Likelihood ratio test=4.06 on 3 df, p=0.255					

Από τον έλεγχο του Wald που περιγράψαμε παραπάνω αλλά και στο Κεφάλαιο 1, προκύπτει ότι μόνο η συμμεταβλητή performance ( $z_1$ ) είναι οριακά στατιστικά σημαντική για την συνάρτηση διακινδύνευσης  $\lambda_2(t|\vec{z})$ . Μάλιστα, προκύπτει ότι για performance=1, με τις τιμές των υπόλοιπων συμμεταβλητών να παραμένουν σταθερές, η συνάρτηση διακινδύνευσης αυξάνεται κατά 80.56% (Πίνακας 15).

### Έλεγχος υπόθεσης αναλογικής διακινδύνευσης:

Όσον αφορά τον διαγνωστικό έλεγχο Schoenfeld για το παραπάνω μοντέλο Cox, έχουμε (Πίνακας 16):

Πίνακας 16: Υπόλοιπα Schoenfeld για την προσαρμογή του  $\lambda_2(t|\vec{z})$ .

	rho	chisq	p-value
performance	0.00492	0.00112	0.973
treatment	0.07832	0.30435	0.581
age	0.17384	138.083	0.240
Global	NA	166.530	0.645

Τα p-value των ελέγχων του Πίνακα 16 είναι αρκετά μεγάλα, γεγονός που μας οδηγεί στο να αποδεχθούμε ότι η υπόθεση της αναλογικότητας διακινδύνευσης είναι ισχυρή για τα παραπάνω δεδομένα.

### Προσαρμογή εναλλακτικού μοντέλου για την $\lambda_2(t|\vec{z})$ :

Λόγω του αρκετά μεγάλου p-value στο Likelihood Ratio Test, αποφασίζουμε να προσαρμόσουμε ένα μοντέλο Cox στα ίδια δεδομένα αλλά χρησιμοποιώντας μόνο τη συμμεταβλητή performance (Πίνακας 17):

Πίνακας 17: Προσαρμογή μοντέλου Cox στην  $\lambda_2(t|\vec{z})$ , με μόνη συμμεταβλητή την performance.

	coef	exp(coef)	se(coef)	z-value	p-value
performance	0.507	1.661	0.315	1.61	0.11

Likelihood ratio test=2.5 on 1 df, p=0.114

Στην συνέχεια θα χρησιμοποιήσουμε τα αποτελέσματα των Likelihood Ratio Test των δύο παραπάνω μοντέλων για να κάνουμε τον έλεγχο

$$H_0 : M_0 : b_2 = b_3 = 0$$

$$H_1 : M_1 : |b_2| + |b_3| \neq 0$$

Τότε,  $X^2 = -2(\hat{l}_0 - \hat{l}_1) = 1 \sim X_2^2$ , με  $p - value = P(X_2^2 > 1.56) = 0.458406$  και άρα σε επίπεδο σημαντικότητας του ελέγχου 95% δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση έναντι της εναλλακτικής. Δεχόμαστε, δηλαδή, ότι το μοντέλο Cox με μόνη συμμεταβλητή την performance προσαρμόζεται καλύτερα στα δεδομένα μας από ότι το μοντέλο των τριών συμμεταβλητών για την συνάρτηση  $\lambda_2(t|\vec{z})$  (Πίνακας 17).

**Έλεγχος υπόθεσης αναλογικής διακινδύνευσης του εναλλακτικού μοντέλου:**

Επιπλέον, όσον αφορά το διαγνωστικό έλεγχο Schoenfeld της αναλογικότητας των κινδύνων για το παραπάνω μοντέλο Cox με την μια συμμεταβλητή, βλέπουμε τα εξής (Πίνακας 18):

Πίνακας 18: Υπόλοιπα Schoenfeld για την προσαρμογή του  $\lambda_2(t|\vec{z})$ , με μόνη συμμεταβλητή την performance.

	rho	chisq	p-value
performance	-0.0471	0.0913	0,763

Και στο παραπάνω μοντέλο, λόγω του μεγάλου p-value του ελέγχου αναλογικών κινδύνων δεν μπορούμε να απορρίψουμε την υπόθεση, σε επίπεδο σημαντικότητας του ελέγχου 95%, ότι ο συντελεστής της συμμεταβλητής performance στο παραπάνω μοντέλο είναι ανεξάρτητος του χρόνου (Πίνακας 18).

Συμπερασματικά, πρέπει να παρατηρήσουμε ότι η προσαρμογή  $\lambda_2(t|\vec{z})$  και με τρεις συμμεταβλητές και με μία είναι λιγότερο ικανοποιητική από την προσαρμογή των αντίστοιχων  $\lambda_1(t|\vec{z})$ . Κατά κύριο λόγο, αυτό συμβαίνει λόγω του μικρού αριθμού των μη-αποκομμένων παρατηρήσεων που λαμβάνονται υπόψιν κατά την προσαρμογή του  $\lambda_2(t|\vec{z})$ . Πράγματι, στην ανάλυση μας χρησιμοποιούμε 44 καταγεγραμμένους θανάτους λόγω MS, με τις υπόλοιπες 150, σχεδόν τις τετραπλάσιες δηλαδή, να είναι δεξιά αποκομμένες, ενώ στην προσαρμογή του  $\lambda_1(t|\vec{z})$  έχουμε συνολικά 83 παρατηρούμενους θανάτους με αιτία την LS, με τις υπόλοιπες 111 παρατηρήσεις να είναι δεξιά αποκομμένες.



### 2.7.5 Μέθοδοι μείωσης του συνολικού αριθμού παραμέτρων

Για τις δύο συναρτήσεις διακινδύνευσης, καθεμιά εκ των οποίων αντιστοιχεί και σε μία διαφορετική μετάβαση ( $0 \Rightarrow 1$  και  $0 \Rightarrow 2$ , αντίστοιχα -βλ. Σχήμα 7), συνοψίζουμε τα αποτελέσματα των στρωματοποιημένων (stratified) Cox μοντέλων που κατασκευάσαμε, στον Πίνακα 19, στον οποίο ο αριθμός μετά την εκάστοτε συμμεταβλητή υποδηλώνει την αιτία/μετάβαση  $k = 1, 2$ .

Πίνακας 19: Συγκεντρωτικά:

	coef	exp(coef)	se(coef)	z-value	p-value
performance.1	0.5760	1.7788	0.2332	2.47	0.014
performance.2	0.5909	1.8056	0.3227	1.83	0.067
treatment.1	0.3327	1.3809	0.2743	1.28	0.239
treatment.2	0.4574	1.5799	0.3900	1.17	0.241
age.1	-0.0247	0.9756	0.2999	-0.08	0.934
age.2	-0.0957	0.9088	0.4062	-0.24	0.814

Likelihood ratio test=10.6 on 6 df, p=0.102

Τώρα, είμαστε σε θέση να ελέγξουμε διάφορες υποθέσεις, με τελικό σκοπό να μειώσουμε, αν είναι εφικτό, το συνολικό αριθμό των παραμέτρων (έξι εν προκειμένω) που χρειαζόμαστε για να περιγράψουμε ικανοποιητικά τα παραπάνω μοντέλα.

#### Πρώτη υπόθεση:

Αρχικά, ας ελέγξουμε την υπόθεση ισότητας των παραμέτρων των συμμεταβλητών στα 2 στρώματα-μεταβάσεις μας (Πίνακας 20):

Πίνακας 20: Έλεγχοι  $b_1^1 = b_1^2$ ,  $b_2^1 = b_2^2$  και  $b_3^1 = b_3^2$ .

	coef	exp(coef)	se(coef)	z-value	p-value
performance	0.580	1.787	0.189	3.07	0.0021
treatment	0.368	1.445	0.224	1.64	0.1007
age	-0.049	0.952	0.241	-0.20	0.8389

Likelihood ratio test=10.5 on 3 df, p=0.0149

Παρατηρούμε από τον Πίνακα 20 ότι για τις συμμεταβλητές age και treatment δεν μπορούμε να απορρίψουμε την υπόθεση ισότητας των παραμέτρων των παραπάνω συμμεταβλητών στα 2 στρώματα,  $b_3^1 = b_3^2$  και  $b_2^1 = b_2^2$ , με  $\hat{b}_3^1 = \hat{b}_3^2 = \hat{b}_3 = -0.049$  και  $\hat{b}_2^1 = \hat{b}_2^2 = \hat{b}_2 = 0.368$ , αντίστοιχα.

### Δεύτερη υπόθεση:

Στην συνέχεια, υποθέτουμε ότι οι βασικές (baseline) συναρτήσεις των  $\lambda_1(t|\vec{z})$  και  $\lambda_2(t|\vec{z})$ ,  $\lambda_{1,0}(t)$  και  $\lambda_{2,0}(t)$ , είναι ανάλογες, με αντίστοιχα μοντέλα παλινδρόμησης:

$$\lambda_1(t|\vec{z}) = \lambda_{10}(t)\exp(b_1^1 z_1 + b_2^1 z_2 + b_3^1 z_3)$$

και:

$$\lambda_2(t|\vec{z}) = \lambda_{10}(t)\exp(b_1^2 z_1 + b_2^2 z_2 + b_3^2 z_3 + d)$$

όπου  $\exp(d)$  είναι η σταθερά αναλογίας των βασικών συναρτήσεων των παραπάνω συναρτήσεων διακινδύνευσης.

Πίνακας 21: Υποθέτουμε ανάλογες βασικές συναρτήσεις διακινδύνευσης μεταξύ των δύο μεταβάσεων.

	coef	exp(coef)	se(coef)	z-value	p-value
performance.1	0.5683	1.7652	0.2315	2.46	0.014
performance.2	0.5822	1.7900	0.3157	1.84	0.065
treatment.1	0.3389	1.4034	0.2739	1.24	0.216
treatment.2	0.4153	1.5149	0.3825	1.09	0.278
age.1	-0.0131	0.9870	0.297	-0.04	0.965
age.2	-0.1045	0.9008	0.3954	-0.26	0.792
trans	-0.6253	0.5351	0.6090	-1.03	0.305

Likelihood ratio test=22.5 on 7 df, p=0.00209

Για την ανάλυση του Πίνακα 21, χρησιμοποιήσαμε την μεταβλητή trans (μετάβαση 1 ή 2) ως μια επιπλέον συμμεταβλητή. Ο συντελεστής της είναι  $-0.6253$  και η επίδραση που θα είχε στον λόγο των συναρτήσεων διακινδύνευσης, δεδομένων ανάλογων βασικών συναρτήσεων, είναι  $0.5351$ . Επομένως, αν και η βασική ανά αιτία συνάρτηση του LS είναι μικρότερη από αυτή του MS, η διαφορά τους δεν είναι στατιστικά σημαντική (p-value= 0.305).

Τέλος, κάνοντας χρήση των υπολοίπων Schoenfeld έχουμε (Πίνακας 22):

Πίνακας 22: Υπόλοιπα Schoenfeld των συναρτήσεων διακινδύνευσης  $\lambda_1(t|\vec{z})$  και  $\lambda_2(t|\vec{z})$  όταν έχουν ανάλογες βασικές συναρτήσεις.

	rho	chisq	p-value
performance.1	-0.15350	3.002349	0.0831
performance.2	0.00147	0.000294	0.9863
treatment.1	-0.08614	0.967944	0.3252
treatment.2	0.04162	0.241398	0.6232
age.1	0.18899	4.687424	0.0304
age.2	0.08496	0.918297	0.3379
factor(trans)2	-0.03009	0.130958	0.7174
GLOBAL	NA	9.073109	0.2474

Από τα μεγάλα p-values των συντελεστών, του Πίνακα 22, όλων των παραμέτρων για τον έλεγχο αναλογικότητας διακινδύνευσης, καταλήγουμε στο συμπέρασμα ότι θα μπορούσαμε να θεωρήσουμε ανάλογες βασικές συναρτήσεις διακινδύνευσης για τις δύο αιτίες θανάτου.

### Τρίτη υπόθεση:

Τέλος, χρησιμοποιούμε την αλγοριθμική μέθοδο reduced rank ώστε να μειώσουμε τον αριθμό των παραμέτρων των αρχικών στρωματοποιημένων μοντέλων του Cox πάνω στις ανά αιτία συναρτήσεις διακινδύνευσης, από έξι σε πέντε, ώστε (Πίνακας 23):

Πίνακας 23: Συντελεστές μοντέλου Reduced Rank.

$a_1$	$a_2$	$a_3$	$c_1$	$c_2$
0.36412	0.23257	-0.0318	1.54585	1.68066

Έχουμε, δηλαδή:

$$\hat{\lambda}_1(t|\vec{z}) = \hat{\lambda}_{1,0}(t) \exp(c_1[a_1z_1 + a_2z_2 + a_3z_3])$$

$$\hat{\lambda}_2(t|\vec{z}) = \hat{\lambda}_{2,0}(t) \exp(c_2[a_1z_1 + a_2z_2 + a_3z_3])$$

## 2.7.6 Ημι-παραμετρική εκτίμηση των πιθανοτήτων μετάβασης

Σε αυτό το σημείο μας ενδιαφέρει να εκτιμήσουμε στην παραπάνω ανάλυση ανταγωνιστικών κινδύνων τις πιθανότητες μετάβασης  $P_{0,k}(s, t)$ , όπου:

$$s = 0 \Rightarrow P_{0,k}(0, t) = I_k(t)$$

και

$$k = s = 0 \Rightarrow P_{0,0}(0, t) = S(t)$$

### Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης των δύο αιτιών:

Θα εφαρμόσουμε τις προαναφερθείσες μεθοδολογίες εκτίμησης των παραπάνω πιθανοτήτων στα δεδομένα μας, αρχικά για τις συναρτήσεις αθροιστικής επίπτωσης (CIF) των δύο αιτιών.

Συνοψίζουμε άλλη μια φορά τις παραμέτρους των συμμεταβλητών μας για τις δύο ανά αιτία συναρτήσεις που μοντελοποιήσαμε εφαρμόζοντας δύο μοντέλα παλινδρόμησης κατά Cox (Πίνακας 24).

Πίνακας 24: Συγκεντρωτικά:

	coef	exp(coef)	se(coef)	z-value	p-value
performance.1	0.5760	1.7788	0.2332	2.47	0.014
performance.2	0.5909	1.8056	0.3227	1.83	0.067
treatment.1	0.3327	1.3809	0.2743	1.28	0.239
treatment.2	0.4574	1.5799	0.3900	1.17	0.241
age.1	-0.0247	0.9756	0.2999	-0.08	0.934
age.2	-0.0957	0.9088	0.4062	-0.24	0.814

Επίσης, από την θεωρία γνωρίζουμε ότι:

$$I_1(t|\vec{z}) = \int_0^t \lambda_1(u|\vec{z}) \exp\left(-\int_0^u \lambda_1(v|\vec{z})dv - \int_0^u \lambda_2(v|\vec{z})dv\right) du$$

και:

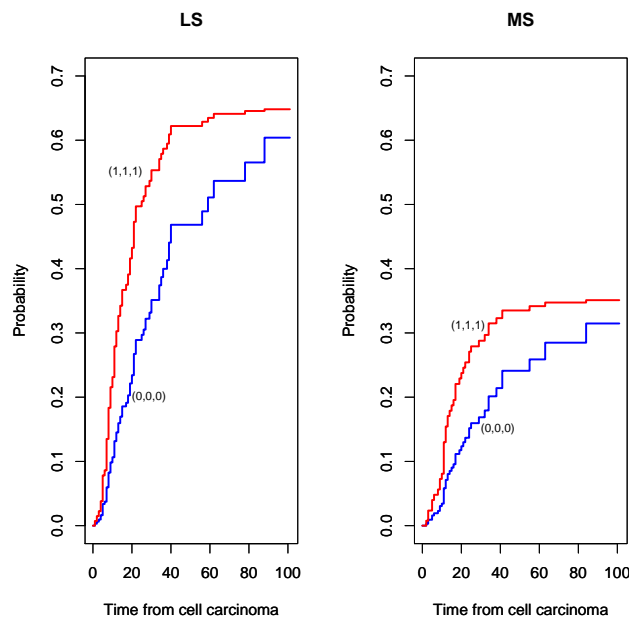
$$I_2(t|\vec{z}) = \int_0^t \lambda_2(u|\vec{z}) \exp\left(-\int_0^u \lambda_1(v|\vec{z})dv - \int_0^u \lambda_2(v|\vec{z})dv\right) du$$

Θα δούμε στην συνέχεια τον τρόπο με τον οποίο οι μεταβολές στις εκτιμημένες συναρτήσεις διακινδύνευσης που οφείλονται στην χρήση διαφορετικών διανυσμάτων συμμεταβλητών, επηρεάζει τις εκτιμήσεις των παραπάνω συναρτήσεων.

Αρχικά, ας μελετήσουμε δύο τύπους ασθενών, αυτούς με διάνυσμα συμμεταβλητών  $(0, 0, 0)$  (ασθενείς αναφοράς) και αυτούς με συμμεταβλητές  $(1, 1, 1)$ .

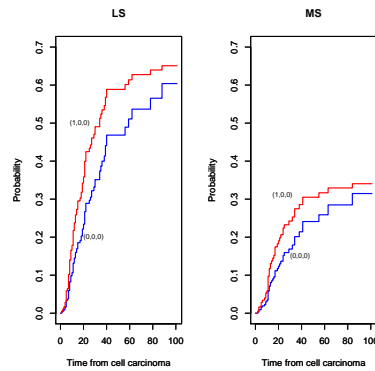
Υπολογίζουμε και σχεδιάζουμε τις  $\hat{P}_{0,1}(0, t) = \hat{I}_1(t)$  και  $\hat{P}_{0,2}(0, t) = \hat{I}_2(t)$  (Σχήμα 13):

Σχήμα 13: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης κάθε αιτίας για τους ασθενείς με διανύσματα συμμεταβλητών  $(0, 0, 0)$  και  $(1, 1, 1)$ , αντίστοιχα.



Από το Σχήμα 13, παρατηρούμε ότι η συνάρτηση αθροιστικής επίπτωσης (CIF) των ασθενών με συμμεταβλητές  $(1, 1, 1)$  είναι συστηματικά μεγαλύτερη από αυτή των  $(0, 0, 0)$  και για τις δύο αιτίες αποτυχίας. Η παρατήρηση αυτή συμφωνεί και με το ότι οι δύο εκ των τριών συμμεταβλητών έχουν αυξητική επίδραση (από 38% ως 80%) ενώ η τρίτη (age) έχει οριακά μειωτική (2.5% και 10%), στις συναρτήσεις διακινδύνευσης  $\lambda_1(t|\vec{z})$  και  $\lambda_2(t|\vec{z})$ , αντίστοιχα, δεδομένου ότι οι υπόλοιπες συμμεταβλητές παραμένουν σταθερές (Πίνακας 24). Οι ασθενείς αναφοράς, λοιπόν, βρίσκονται σε μικρότερο κίνδυνο θανάτου από οποιαδήποτε αιτία, σε σχέση με αυτούς με διάνυσμα συμμεταβλητών  $(1, 1, 1)$ .

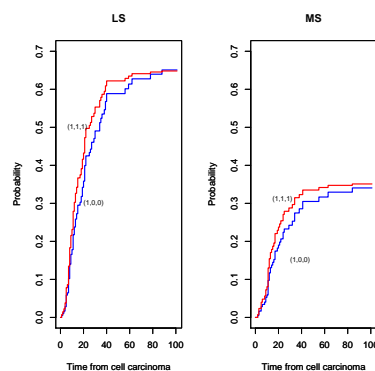
Υπολογίζουμε επίσης τις εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης για δύο τύπους ασθενών με διανύσματα συμμεταβλητών  $(0, 0, 0)$  και  $(1, 0, 0)$  (Σχήμα 14).



Σχήμα 14: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης κάθε αιτίας για τους ασθενείς με διανύσματα συμμεταβλητών  $(0, 0, 0)$  και  $(1, 0, 0)$ , αντίστοιχα.

Το παραπάνω αποτέλεσμα, του Σχήματος 14, συμφωνεί με το γεγονός ότι η πιο στατιστικά σημαντική συμμεταβλητή και των δύο μοντέλων Cox των συναρτήσεων διακινδύνευσης, performance, έχει αυξητική επίδραση στις συναρτήσεις διακινδύνευσης όταν αυξάνεται από 0 σε 1 δεδομένου ότι οι υπόλοιπες συμμεταβλητές παραμένουν σταθερές-0 εν προκειμένω-.

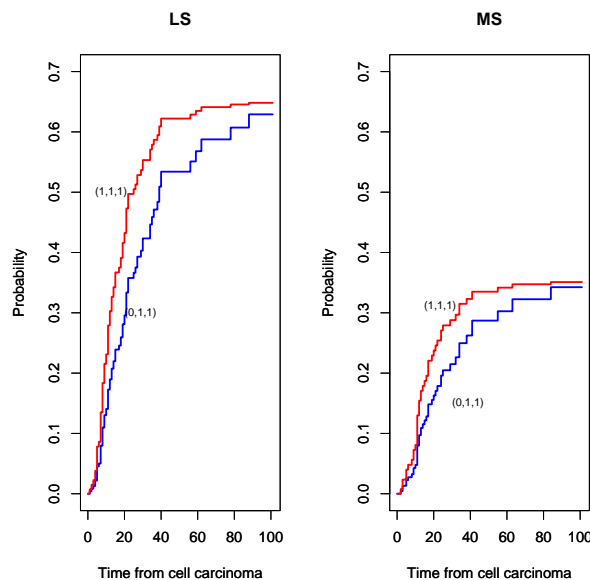
Επίσης, βλέπουμε τις εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης για δύο τύπους ασθενών με διανύσματα συμμεταβλητών  $(1, 0, 0)$  και  $(1, 1, 1)$  αντίστοιχα (Σχήμα 15).



Σχήμα 15: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης κάθε αιτίας για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 0, 0)$  και  $(1, 1, 1)$ , αντίστοιχα.

Μπορούμε να παρατηρήσουμε, από το Σχήμα 15, ότι οι συναρτήσεις αθροιστικής επίπτωσης των ασθενών με συμμεταβλητές  $(1, 1, 1)$  είναι συστηματικά μεγαλύτερες και για τις δύο αιτίες θανάτου. Το παραπάνω συμβαδίζει με το γεγονός ότι με την αύξηση της, από 0 σε 1, η μεταβλητή treatment έχει αυξητική επίδραση στις συναρτήσεις διακινδύνευσης ενώ η age, η οποία είναι λιγότερο σημαντική για τα μοντέλα Cox που κατασκευάσαμε, οριακά μειωτική.

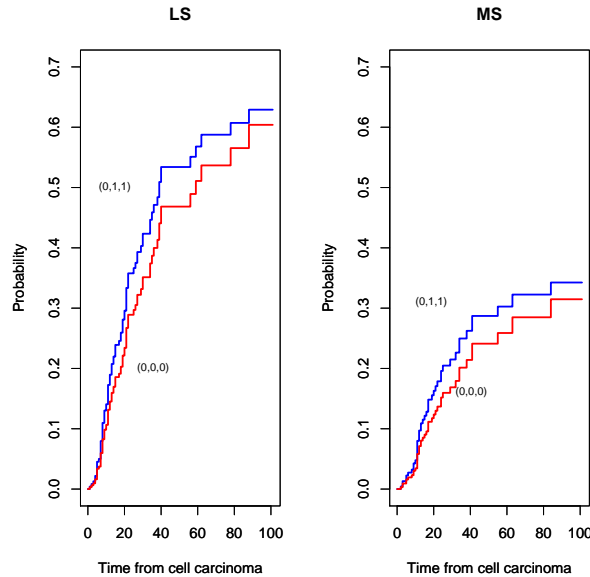
Επιπλέον, μπορούμε να δούμε και τις εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης για δύο τύπους ασθενών με διανύσματα συμμεταβλητών  $(0, 1, 1)$  και  $(1, 1, 1)$  (Σχήμα 16).



Σχήμα 16: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης κάθε αιτίας για τους ασθενείς με διανύσματα συμμεταβλητών  $(0, 1, 1)$  και  $(1, 1, 1)$ , αντίστοιχα.

Το συμπέρασμα που μπορούμε να εξάγουμε είναι ότι και πάλι (Σχήμα 16), η μειωτική επίδραση που έχει στις συναρτήσεις διακινδύνευσης η μεταβολή της performance από 1 σε 0 έρχεται σε συμφωνία και με τις αντίστοιχες συναρτήσεις αθροιστικής επίπτωσης, οι οποίες είναι συνεχώς μικρότερες για τους ασθενείς με συμμεταβλητές  $(0, 1, 1)$  και για τις δύο αιτίες θανάτου.

Για ασθενείς με συμμεταβλητές  $(0, 1, 1)$  και  $(0, 0, 0)$  (Σχήμα 17):



Σχήμα 17: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης κάθε αιτίας για τους ασθενείς με διάνυσματα συμμεταβλητών  $(0, 1, 1)$  και  $(0, 0, 0)$ , αντίστοιχα.

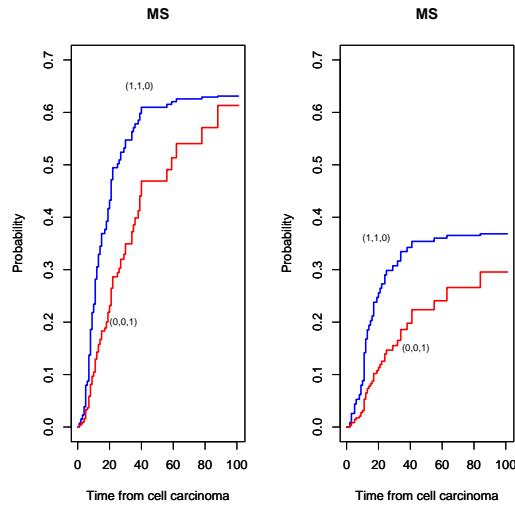
Οι συναρτήσεις αθροιστικής επίπτωσης (CIF) και για τις δύο αιτίες θανάτου είναι μεγαλύτερες για τους ασθενείς με διάνυσμα συμμεταβλητών  $(0, 1, 1)$  γεγονός που συμφωνεί με το ότι η συμμεταβλητή treatment τείνει να αυξήσει τις  $\lambda_1(t|\vec{z})$  και  $\lambda_2(t|\vec{z})$  καθώς αυξάνεται, ενώ η λιγότερο στατιστικά σημαντική age, να τις μειώσει οριακά (Σχήμα 17).

Συμπερασματικά, λοιπόν, είναι φανερό από την παραπάνω μας ανάλυση, για τις περιπτώσεις ασθενών που μελετήσαμε, ότι καθώς μεταβάλλονται οι τιμές των συμμεταβλητών, η συμπεριφορά των συναρτήσεων αθροιστικής επίπτωσης είναι σε συμφωνία με αυτή των συναρτήσεων διακινδύνευσης.

Οπότε, οι ασθενείς με συμμεταβλητές  $(1, 1, 0)$  θα έχουν τον μεγαλύτερο κίνδυνο αποτυχίας και από τις δύο αιτίες με την πάροδο του χρόνου σε σχέση με τους ασθενείς άλλων συμμεταβλητών. Αντίθετα, οι ασθενείς με διάνυσμα συμμεταβλητών  $(0, 0, 1)$ , συγκριτικά με όλους τους υπόλοιπους ασθενείς άλλων συμμεταβλητών, έχουν τις μεγαλύτερες πιθανότητες επιβίωσης.

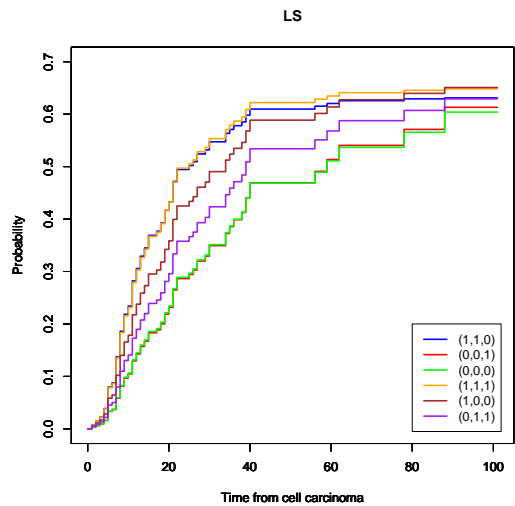


Για τους ασθενείς αυτούς, έχουμε (Σχήμα 18):



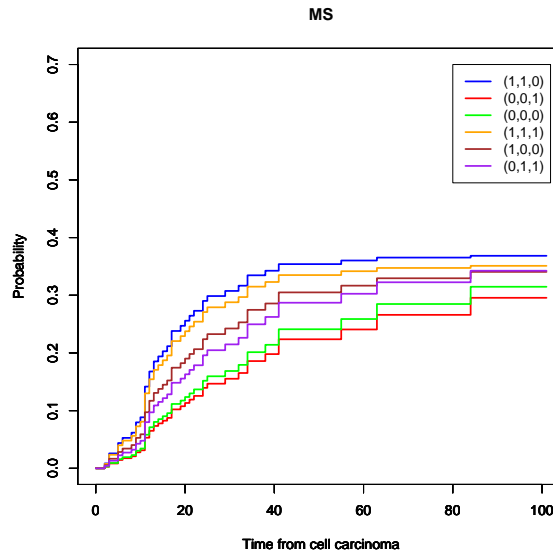
Σχήμα 18: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης κάθε αιτίας για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 1, 0)$  και  $(0, 0, 1)$ , αντίστοιχα.

Για τις εκτιμήσεις της  $I_1(t|\vec{z})$  για όλα τα διανύσματα συμμεταβλητών που μελετήσαμε παραπάνω, έχουμε (Σχήμα 19):



Σχήμα 19: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης της αιτίας LS για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 1, 0)$ ,  $(0, 0, 1)$ ,  $(0, 0, 0)$ ,  $(1, 1, 1)$ ,  $(1, 0, 0)$  και  $(0, 1, 1)$ , αντίστοιχα.

Για τις εκτιμήσεις της  $I_2(t|\vec{z})$  για όλα τα διανύσματα συμμεταβλητών που μελετήσαμε παραπάνω, έχουμε (Σχήμα 20):



Σχήμα 20: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης της αιτίας LS για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 1, 0)$ ,  $(0, 0, 1)$ ,  $(0, 0, 0)$ ,  $(1, 1, 1)$ ,  $(1, 0, 0)$  και  $(0, 1, 1)$ , αντίστοιχα.

Από τα Σχήματα 19 και 20 παρατηρούμε τα εξής:

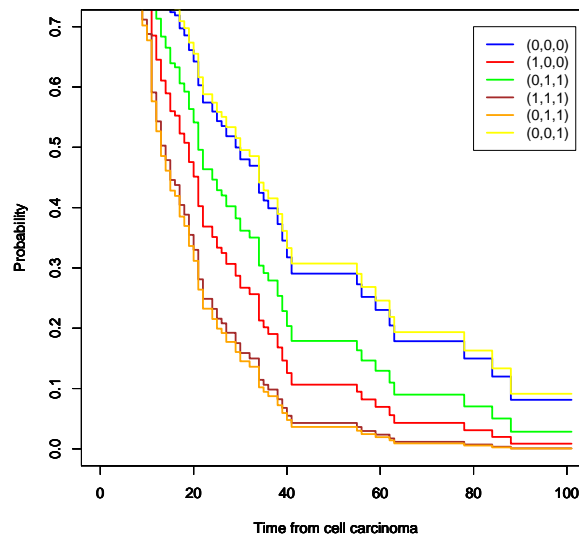
Όσον αφορά τις εκτιμήσεις της συνάρτησης αθροιστικής επίπτωσης της αιτίας LS, με εξαίρεση τους ασθενείς με συμμεταβλητές  $(1, 1, 1)$  και  $(1, 1, 0)$ , ακολουθούν τις προαναφερθείσες παρατηρήσεις μας. Οι αντίστοιχες εκτιμήσεις για την αιτία MS, ωστόσο, είναι σε πλήρη συμφωνία με όσο περιγράψαμε παραπάνω.

Συμπερασματικά, πρέπει να αναφέρουμε ότι όπως είδαμε στην θεωρία η συμφωνία που υπάρχει μεταξύ των μεταβολών των συναρτήσεων διακινδύνευσης με την συμπεριφορά των συναρτήσεων αθροιστικής επίπτωσης για τα διαφορετικά διανύσματα συμμεταβλητών, δεν επεκτείνεται σε όλα τα προβλήματα ανταγωνιστικών κινδύνων.

### Ημι-παραμετρικές εκτιμήσεις των συναρτήσεων επιβίωσης:

Ακόμα, θα μας ενδιέφερε να μελετήσουμε την πιθανότητα επιβίωσης των ασθενών σε σχέση με τον χρόνο για κάθε έναν από τους έξι τύπους ασθενών (Σχήμα 21):

Σχήμα 21: Εκτιμήσεις των συναρτήσεων επιβίωσης των έξι διαφορετικών τύπων ασθενών που μελετήσαμε.

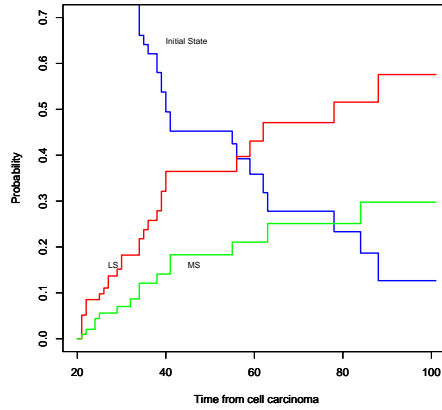


Το αποτέλεσμα του Σχήματος 21 συμφωνεί με τις παρατηρήσεις που κάναμε στη μελέτη των συναρτήσεων αθροιστικής επίπτωσης για τους έξι τύπους ασθενών. Πράγματι, για τους ασθενείς με συμμεταβλητές  $(0, 0, 1)$ , των οποίων η συνάρτηση αθροιστικής επίπτωσης έπαιρνε τις μικρότερες τιμές, έχουν την μεγαλύτερη πιθανότητα επιβίωσης. Αντίστροφα, για τους ασθενείς με διάνυσμα συμμεταβλητών  $(1, 1, 0)$ .

### Εκτιμήσεις των $P_{0j}(20, t)$ , $\forall j = 0, 1, 2$ :

Στην συνέχεια, θα μας ενδιέφερε να μελετήσουμε τις εκτιμημένες πιθανότητες  $\hat{P}_{0,j}(s, t)$  για  $j = 0, 1, 2$  και  $s \neq 0$ . Αν δηλαδή δεχθούμε ότι ένας ασθενής βρίσκεται στην αρχική κατάσταση 0 της μαρκοβιανής αλυσίδας την χρονική στιγμή  $s = 20$ , για παράδειγμα, θα θέλαμε να προσδιορίσουμε τις πιθανότητες παραμονής ή μετάβασης στις υπόλοιπες δύο καταστάσεις της αλυσίδας σε σχέση με τον χρόνο.

Για τους ασθενείς με συμμεταβλητές  $(0, 0, 0)$  (Σχήμα 22):

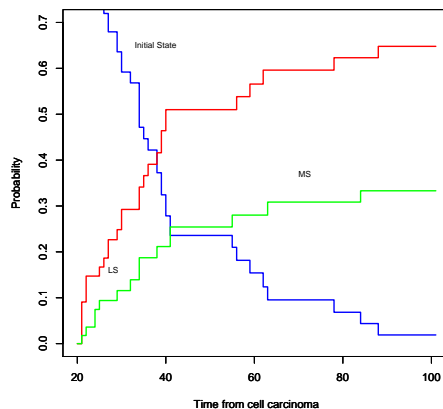


Σχήμα 22: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(0, 0, 0)$ , δεδομένου ότι την χρονική στιγμή  $s = 20$  είναι υγιής.

Από το Σχήμα 22, παρατηρούμε ότι, αρχικά, η πιθανότητα παραμονής στην αρχική κατάσταση 0 είναι αρκετά μεγαλύτερη των πιθανοτήτων μετάβασης στις καταστάσεις 1 και 2, αντίστοιχα. Ωστόσο, για χρονικές στιγμές  $t \geq 56$  η πιθανότητα αποτυχίας λόγω LS ξεπερνάει την πιθανότητα παραμονής στην 0 και παίρνει μέγιστη τιμή 0.576 για  $t \geq 88$ .

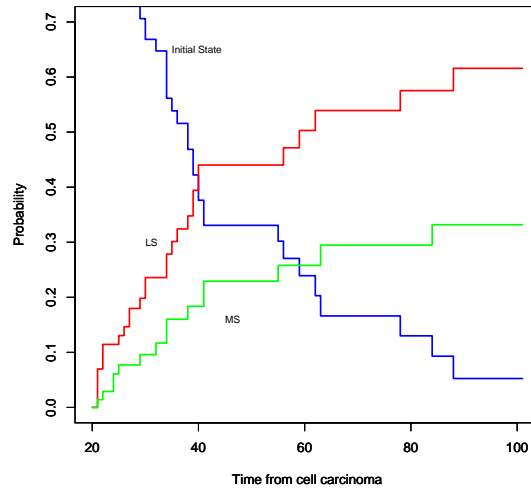
Για τους ασθενείς με συμμεταβλητές  $(1, 0, 0)$  (Σχήμα 23):

Σχήμα 23: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(1, 0, 0)$ , δεδομένου ότι την χρονική στιγμή  $s = 20$  είναι υγιής.

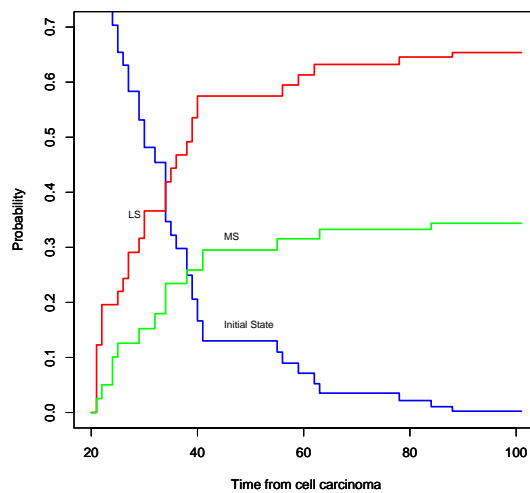


Για τους ασθενείς με συμμεταβλητές  $(0, 1, 1)$  (Σχήμα 24):

Σχήμα 24: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(0, 1, 1)$ , δεδομένου ότι την χρονική στιγμή  $s = 20$  είναι υγιής.



Για τους ασθενείς με συμμεταβλητές  $(1, 1, 1)$  (Σχήμα 25):



Σχήμα 25: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(1, 1, 1)$ , δεδομένου ότι την χρονική στιγμή  $s = 20$  είναι υγιής.

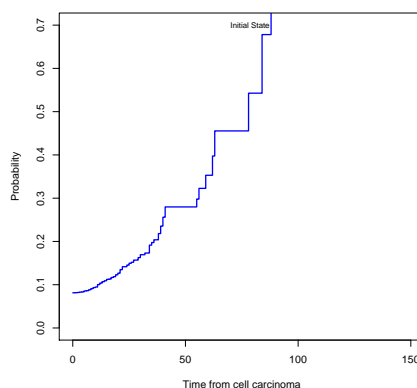
Από τις παραπάνω γραφικές παραστάσεις (Σχήματα 22 – 25), μπορούμε να εξάγουμε τα εξής:

Για χρόνους κοντά στο  $s = 20$ , είναι πιο πιθανό ο ασθενής να παραμείνει στην αρχική του θέση και άρα να παραμείνει ασθενής. Ωστόσο, σε βάθος χρόνου η πιθανότητα αποτυχίας λόγω LS είναι αυτή που υπερτερεί. Το αποτέλεσμα αυτό συμφωνεί και με τα δεδομένα μας στα οποία καταγράφονται 83 θάνατοι λόγω LS, 44 λόγω MS και 67 δεξιά αποκομμένες παρατηρήσεις. Μάλιστα, η πιθανότητα αυτή φαίνεται να είναι μεγαλύτερη για τους ασθενείς με performance=1 στο διάνυσμα συμμεταβλητών τους. Συγκεκριμένα, στους ασθενείς με συμμεταβλητές (1, 1, 1) παρατηρείται η πιο ταχεία αύξηση της πιθανότητας αποτυχίας λόγω LS, συνοδευόμενη και από την πιο ταχεία μείωση της πιθανότητας παραμονής στην 0, με μέγιστη τιμή 0.654 για  $t \geq 88$ .

### Εκτιμήσεις των $P_{00}(s, 150)$ :

Επίσης, κάνοντας χρήση της R, μπορούμε να σταθεροποιήσουμε τον ορίζοντα της πρόβλεψης μας και να εκτιμήσουμε την πιθανότητα  $\hat{P}(s, t)$  όπου το  $t$  θα είναι σταθερό. Επομένως, για  $t = 150$ , μπορούμε να εκτιμήσουμε την πιθανότητα ένας ασθενής να βρίσκεται σε κάθε μια από τις τρεις δυνατές θέσεις της Μαρκοβιανής αλυσίδας. Κάνουμε τις προβλέψεις μας θεωρώντας ότι την  $t = 150$  η αλυσίδα βρίσκεται στην αρχική της κατάσταση και μελετούμε την πιθανότητα παραμονής του για τις προηγούμενες χρονικές στιγμές  $[0, 150]$ .

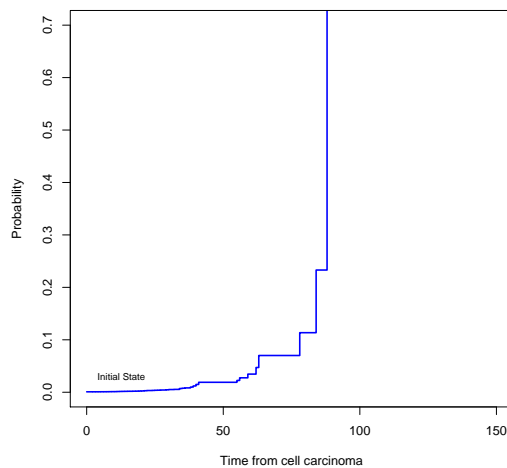
Ας δούμε για τον τύπο ασθενών με διάνυσμα συμμεταβλητών (0, 0, 0) (Σχήμα 26):



Σχήμα 26: Πιθανότητα παραμονής στην αρχική κατάσταση ενός ασθενή με συμμεταβλητές (0, 0, 0), στο χρονικό διάστημα  $[0, 150]$ , δεδομένου ότι την χρονική στιγμή  $t = 150$  είναι υγιής.

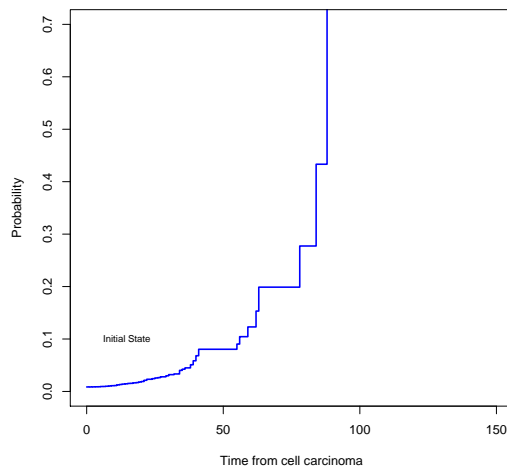
Για τους ασθενείς με με διάνυσμα συμμεταβλητών  $(1, 1, 1)$  (Σχήμα 27):

Σχήμα 27: Πιθανότητα παραμονής στην αρχική κατάσταση ενός ασθενή με συμμεταβλητές  $(1, 1, 1)$ , στο χρονικό διάστημα  $[0, 150]$ , δεδομένου ότι την χρονική στιγμή  $t = 150$  είναι υγιής.



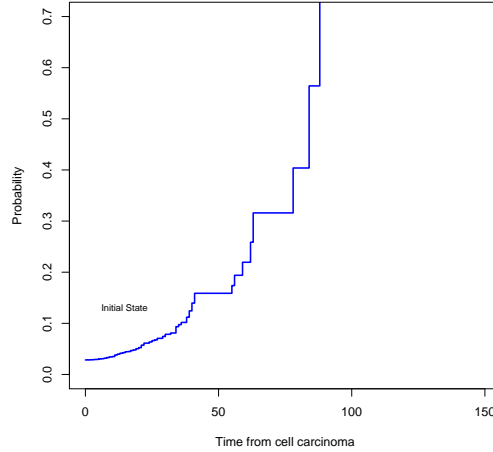
Επίσης, για τους ασθενείς με διάνυσμα συμμεταβλητών  $(1, 0, 0)$  (Σχήμα 28):

Σχήμα 28: Πιθανότητα παραμονής στην αρχική κατάσταση ενός ασθενή με συμμεταβλητές  $(1, 0, 0)$ , στο χρονικό διάστημα  $[0, 150]$ , δεδομένου ότι την χρονική στιγμή  $t = 150$  είναι υγιής.



Τέλος, για τους ασθενείς με διάγνωση συμμεταβλητών  $(0, 1, 1)$  (Σχήμα 29):

Σχήμα 29: Πιθανότητα παραμονής στην αρχική κατάσταση ενός ασθενή με συμμεταβλητές  $(0, 1, 1)$ , στο χρονικό διάστημα  $[0, 150]$ , δεδομένου ότι την χρονική στιγμή  $t = 150$  είναι υγιής.



Από τα Σχήματα 26 – 29 παρατηρούμε ότι η πιο απότομη αύξηση της πιθανότητας παραμονής στην 0, παρουσιάζεται στους ασθενείς με συμμεταβλητές  $(0, 0, 0)$ . Αντίθετα, η πιο αργή αύξηση της πιθανότητας παραμονής στην αρχική κατάσταση, παρουσιάζεται στους ασθενείς με συμμεταβλητές  $(1, 1, 1)$ . Μάλιστα, για όλους τους ασθενείς με τις συμμεταβλητές που μελετήσαμε παραπάνω, προκύπτει ότι η  $\hat{P}_{00}(s, 150)$  παίρνει την τιμή 1 για  $s \geq 88$ .

### 2.7.7 Προσαρμογή μοντέλων παλινδρόμησης Fine και Gray στις συναρτήσεις διακινδύνευσης

Επίσης, θα προσαρμόσουμε και μοντέλα παλινδρόμησης Cox στις δύο συναρτήσεις διακινδύνευσης του προβλήματος μας βασισμένα στις subdistribution συναρτήσεις. Η παραπάνω παλινδρόμηση λέγεται Fine and Gray και όπως είδαμε παραπάνω η εκτίμηση των παραμέτρων των μοντέλων Cox στηρίζεται σε μια τροποποιημένη συνάρτηση Μερικής Πιθανοφάνειας (IPCW), η κατασκευή της οποίας βασίζεται στο αφύσικο (Fine and Gray) σύνολο διακινδύνευσης που χρησιμοποιείται. Τα μοντέλα παλινδρόμησης είναι τα εξής:

$$\bar{\lambda}_1(t|\vec{z}) = \bar{\lambda}_{1,0}(t) \exp(b_1^1 z_1 + b_2^1 z_2 + b_3^1 z_3)$$

$$\bar{\lambda}_2(t|\vec{z}) = \bar{\lambda}_{2,0}(t) \exp(b_1^2 z_1 + b_2^2 z_2 + b_3^2 z_3)$$



### Προσαρμογή του $\bar{\lambda}_1(t|\vec{z})$ :

Αρχικά, θεωρούμε την αιτία LS ως την μοναδική αιτία αποτυχίας/θανάτου των ασθενών και προσαρμόζουμε ένα μοντέλο Cox σε μια single-event (με μοναδική αιτία αποτυχίας) ανάλυση επιβίωσης στα τροποποιημένα μας, ωστόσο, δεδομένα. Η τροποποίηση βρίσκεται στο σύνολο διακινδύνευσης  $R_j$  για την χρονική στιγμή  $t_j$ , το οποίο εμφανίζεται στην μερική συνάρτηση πιθανοφάνειας, στο οποίο θεωρούμε ότι ανήκουν και άτομα που έχουν αποτύχει από την MS μέχρι και την  $t_j$ . Έχουμε, λοιπόν (Πίνακας 25):

Πίνακας 25: Προσαρμογή μοντέλου Cox στην  $\bar{\lambda}_1(t|\vec{z})$ .

	coef	exp(coef)	se(coef)	z-value	p-value
performance	0.3148	1.370	0.223	1.410	0.16
treatment	0.1515	1.164	0.244	0.622	0.53
age	-0.0981	0.907	0.298	-0.329	0.74

### Προσαρμογή του $\bar{\lambda}_2(t|\vec{z})$ :

Με αντίστοιχη λογική, αλλά αυτή τη φορά για την αιτία MS, θεωρούμε μια single-event ανάλυση επιβίωσης, στα τροποποιημένα δεδομένα της οποίας -λαμβάνοντας άτομα που έχουν αποτύχει από την LS ως ασθενείς που κινδυνεύουν από αποτυχία λόγω της MS-προσαρμόζουμε ένα μοντέλο Cox (Πίνακας 26):

Πίνακας 26: Προσαρμογή μοντέλου Cox στην  $\bar{\lambda}_2(t|\vec{z})$ .

	coef	exp(coef)	se(coef)	z-value	p-value
performance	0.196	1.217	0.304	0.646	0.52
treatment	0.152	1.164	0.379	0.401	0.69
age	-0.144	0.866	0.395	-0.364	0.72

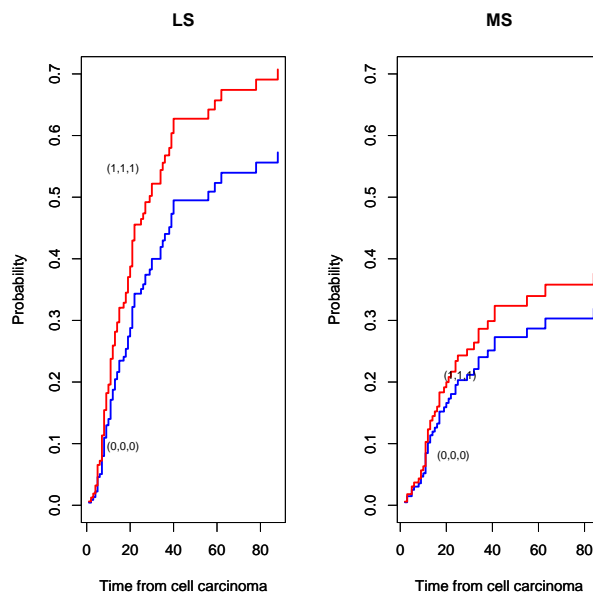
Και στα δύο παραπάνω μοντέλα Fine and Gray που προσαρμόσαμε (Πίνακες 25 και 26) για τις δύο δυνατές μεταβάσεις του προβλήματος μας, από τον έλεγχο Wald προκύπτει ότι η προσαρμογή τους στα δεδομένα μας δεν είναι ιδιαίτερα ικανοποιητική. Ωστόσο, σε αντίθεση με την ανά αιτία (cause-specific) ανάλυση παλινδρόμησης που περιγράψαμε παραπάνω, σε αυτή την περίπτωση υπάρχει εξ ορισμού άμεση σύνδεση μεταξύ της επίδρασης των συμμεταβλητών στις subdistribution συναρτήσεις διακινδύνευσης και των συναρτήσεων αθροιστικής επίπτωσης. Επίσης, οι συντελεστές επίδρασης (exp(coef)) των συμμεταβλητών μας ακολουθούν το ίδιο μοτίβο με τα ανά αιτία μοντέλα.

Πράγματι, οι performance και treatment έχουν θετική επίδραση στην συνάρτηση διακινδύνευσης για μεταβολή των τιμών τους από 0 σε 1, ενώ η age για την ίδια μεταβολή παρουσιάζει οριακή μείωση.

Αν και παραπάνω συντελεστές είναι μικρότεροι από τους αντίστοιχους της ανά αιτία/cause-specific παλινδρόμησης, αναμένουμε από τη θεωρία οι συναρτήσεις αθροιστικής επίπτωσης να έχουν την ίδια συμπεριφορά με τις ανά αιτία συναρτήσεις για τα διάφορα διανύσματα συμμεταβλητών.

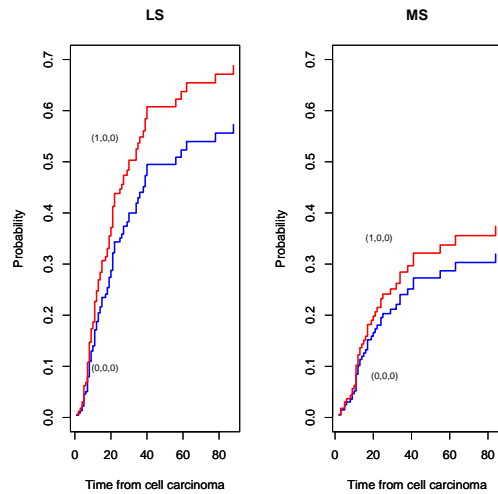
### Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης των δύο αιτίων:

Στην συνέχεια, σχεδιάζουμε τις συναρτήσεις αθροιστικής επίπτωσης για τις δύο αιτίες θανάτου LS και MS, για δύο ασθενείς με συμμεταβλητές  $\vec{z} = (0, 0, 0)$  και  $\vec{z} = (1, 1, 1)$ , αντίστοιχα (Σχήμα 30):



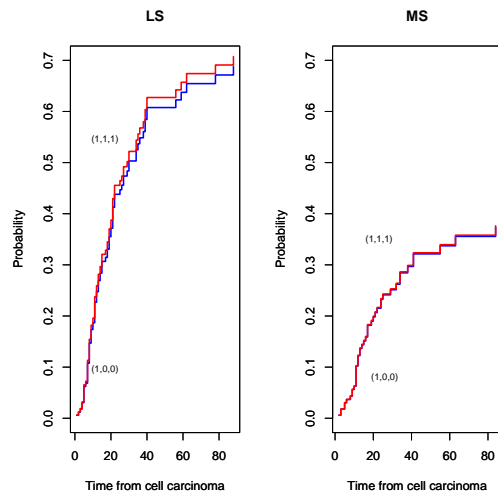
Σχήμα 30: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης των ασθενών με συμμεταβλητές  $(0, 0, 0)$  και  $(1, 1, 1)$ , αντίστοιχα, χρησιμοποιώντας το μοντέλο των Fine και Gray.

Ομοίως, για δύο τύπους ασθενών με διανύσματα συμμεταβλητών  $(0, 0, 0)$  και  $(1, 0, 0)$  (Σχήμα 31):



Σχήμα 31: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης των ασθενών με συμμεταβλητές  $(0, 0, 0)$  και  $(1, 0, 0)$ , αντίστοιχα, χρησιμοποιώντας το μοντέλο των Fine και Gray.

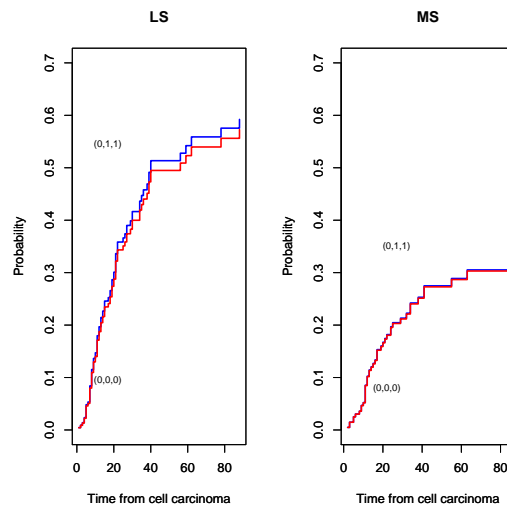
Αλλά και για τους ασθενείς με συμμεταβλητές  $(1, 0, 0)$  και  $(1, 1, 1)$  (Σχήμα 32):



Σχήμα 32: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης των ασθενών με συμμεταβλητές  $(1, 0, 0)$  και  $(1, 1, 1)$ , αντίστοιχα, χρησιμοποιώντας το μοντέλο των Fine και Gray.

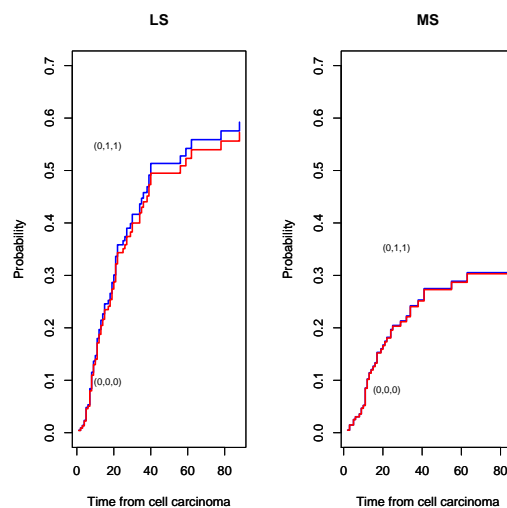
Για ασθενείς με αντίστοιχα διανύσματα  $(0, 1, 1)$  και  $(0, 0, 0)$  (Σχήμα 33):

Σχήμα 33: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης των ασθενών με συμμεταβλητές  $(0, 1, 1)$  και  $(0, 0, 0)$ , αντίστοιχα, χρησιμοποιώντας το μοντέλο των Fine και Gray.



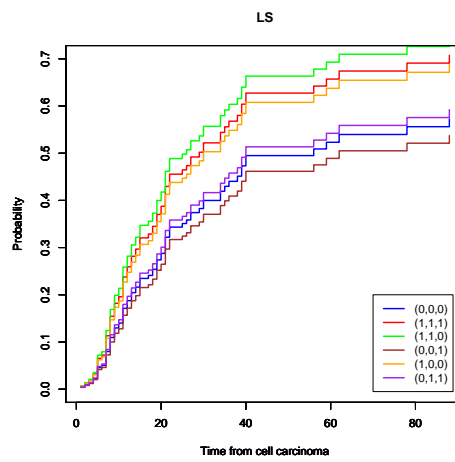
Για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 1, 0)$  και  $(0, 0, 1)$  (Σχήμα 34):

Σχήμα 34: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης των ασθενών με συμμεταβλητές  $(1, 1, 0)$  και  $(0, 0, 1)$ , αντίστοιχα, χρησιμοποιώντας το μοντέλο των Fine και Gray.

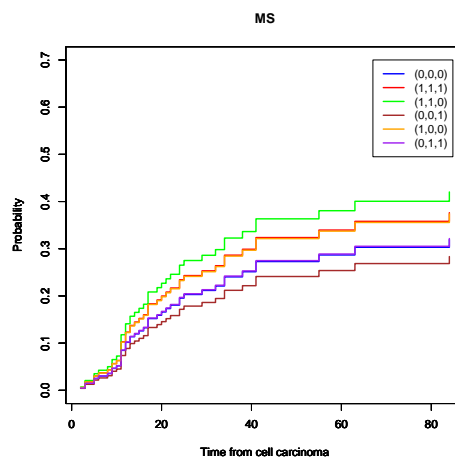


Πράγματι, από τα Σχήματα 30 – 34, παρατηρούμε μεγαλύτερες συναρτήσεις αθροιστικής επίπτωσης για τους ασθενείς τιμές performance και treatment ίσες με 1 και age ίση με 0.

Τέλος, θα μελετήσουμε τις εκτιμήτριες των συναρτήσεων αθροιστικής επίπτωσης για όλα τα διανύσματα συμμεταβλητών που αναφέραμε παραπάνω (Σχήματα 35 και 36).



Σχήμα 35: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης της αιτίας LS για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 1, 0)$ ,  $(0, 0, 1)$ ,  $(0, 0, 0)$ ,  $(1, 1, 1)$ ,  $(1, 0, 0)$  και  $(0, 1, 1)$ , αντίστοιχα, χρησιμοποιώντας το μοντέλο των Fine και Gray.



Σχήμα 36: Εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης της αιτίας MS για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 1, 0)$ ,  $(0, 0, 1)$ ,  $(0, 0, 0)$ ,  $(1, 1, 1)$ ,  $(1, 0, 0)$  και  $(0, 1, 1)$ , αντίστοιχα, χρησιμοποιώντας το μοντέλο των Fine και Gray.

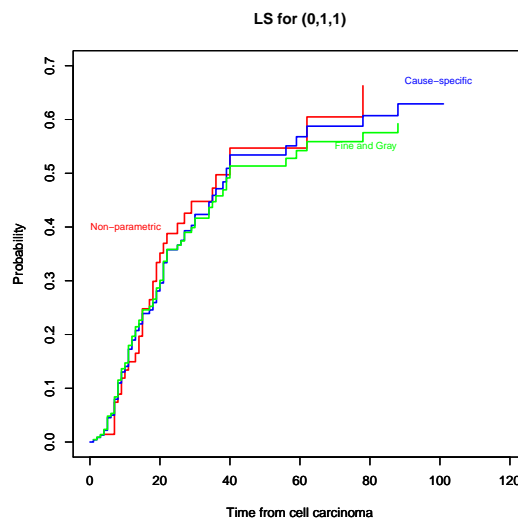
Τα συμπεράσματα που προκύπτουν από τα Σχήματα 30–36, είναι όμοια με την ανά αιτία/cause-specific ανάλυση των συναρτήσεων διακινδύνευσης και, όπως εξηγήσαμε και παραπάνω είναι θεωρητικά αναμενόμενα λόγω των συντελεστών των συμμεταβλητών στα προσαρμοσμένα μας μοντέλα.

### 2.7.8 Σύγκριση των ανά αιτία και Fine και Gray μοντέλων παλινδρόμησης

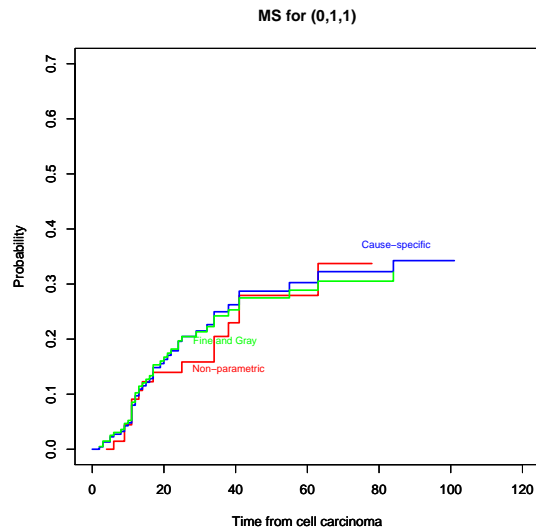
Όπως είδαμε και στην θεωρία ένας από τους πιο ενδεδειγμένους τρόπους να ελέγξουμε την προσαρμογή των ανά αιτία/cause-specific μοντέλων παλινδρόμησης και Fine and Gray στις συναρτήσεις διακινδύνευσης είναι να συγκρίνουμε τις συναρτήσεις αθροιστικής επίπτωσης κάθε αιτίας όπως προκύπτουν από τα προαναφερθέντα μοντέλα με τις μη-παραμετρικές εκτιμήτριες τους. Θα δούμε τα παραπάνω για τους ασθενείς με διανύσματα συμμεταβλητών  $(0, 1, 1)$  και  $(1, 1, 1)$  καθώς από όλα τα διανύσματα συμμεταβλητών που μελετήσαμε, μόνο για τους ασθενείς με αυτές τις συμμεταβλητές έχουμε επαρκή δεδομένα για να προβούμε σε μη-παραμετρική ανάλυση.

Για τους ασθενείς με διάνυσμα συμμεταβλητών  $(0, 1, 1)$  (Σχήματα 37 και 38):

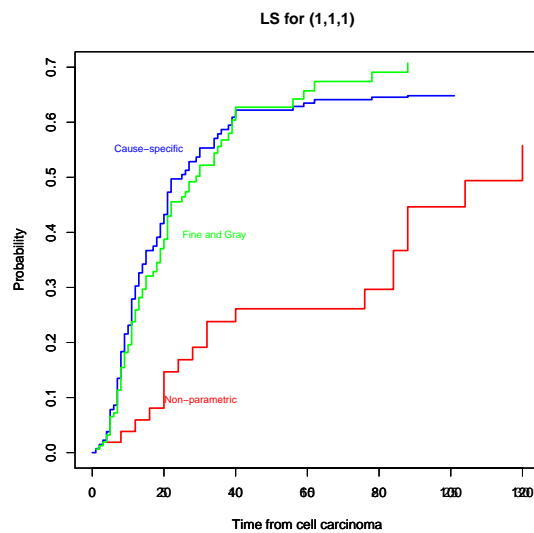
Σχήμα 37: Σύγκριση της μη-παραμετρικής εκτίμησης της  $I_1(t|\vec{z})$  με την ανά αιτία (cause-specific) και Fine and Gray ανάλυση παλινδρόμησης για τους ασθενείς με συμμεταβλητές  $(0, 1, 1)$ .



Σχήμα 38: Σύγκριση της μη-παραμετρικής εκτίμησης της  $I_2(t|\vec{z})$  με την ανά αιτία (cause-specific) και Fine and Gray ανάλυση παλινδρόμησης για τους ασθενείς με συμμεταβλητές (0, 1, 1).

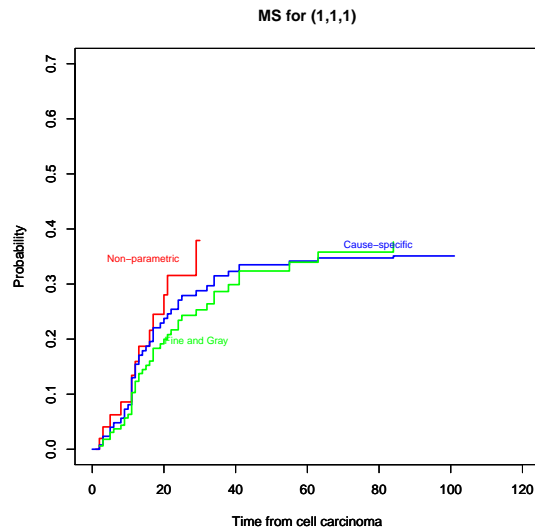


Ενώ, για αυτούς με διανύσματα συμμεταβλητών (1, 1, 1) έχουμε (Σχήματα 39 και 40):



Σχήμα 39: Σύγκριση της μη-παραμετρικής εκτίμησης της  $I_1(t|\vec{z})$  με την ανά αιτία (cause-specific) και Fine and Gray ανάλυση παλινδρόμησης για τους ασθενείς με συμμεταβλητές (1, 1, 1).

Σχήμα 40: Σύγκριση της μη-παραμετρικής εκτίμησης της  $I_2(t|\bar{z})$  με την ανά αιτία (cause-specific) και Fine and Gray ανάλυση παλινδρόμησης για τους ασθενείς με συμμεταβλητές (1, 1, 1).



Συμπερασματικά, από τα Σχήματα 37 – 40, με εξαίρεση τις συναρτήσεις αθροιστικής επίπτωσης της LS που προκύπτουν με βάση την ανά αιτία (cause-specific) και την Fine and Gray παλινδρόμηση για τους ασθενείς με συμμεταβλητές (1, 1, 1), τα δύο αυτά μοντέλα παλινδρόμησης προσαρμόζονται αρκετά ικανοποιητικά στα δεδομένα μας, δίνοντας εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης οι οποίες έχουν μικρές διαφορές μεταξύ τους.



### 3 Μοντέλα Multi-state

#### 3.1 Μοντελοποίηση

Στο προηγούμενο κεφάλαιο, μελετήσαμε μοντέλα τα οποία είχαν μια αρχική κατάσταση και  $k$  απορροφητικές/absorbing. Όπως προείπαμε, απορροφητικές είναι οι καταστάσεις  $j$  για τις οποίες αν  $\exists s > 0 : X(s) = j \Rightarrow X(t) = j \forall t > s$ .

Υπάρχουν, ωστόσο, διάφορα μοντέλα ανάλυσης επιβίωσης τα οποία περιέχουν και μεταβατικές-παροδικές/transient καταστάσεις.

Θα δώσουμε δύο ορισμούς για τις παροδικές/μεταβατικές καταστάσεις μιας μαρκοβιανής αλυσίδας (Norris, 1997) :

1. Έστω  $X(t)$  μια δεξιά συνεχής στοχαστική ανέλιξη και  $i$  μια κατάσταση της. Η κατάσταση  $i$  λέγεται παροδική/μεταβατική αν η πιθανότητα το σύνολο  $\{t \geq 0 : X(t) = i\}$  να μην είναι φραγμένο είναι μηδέν.

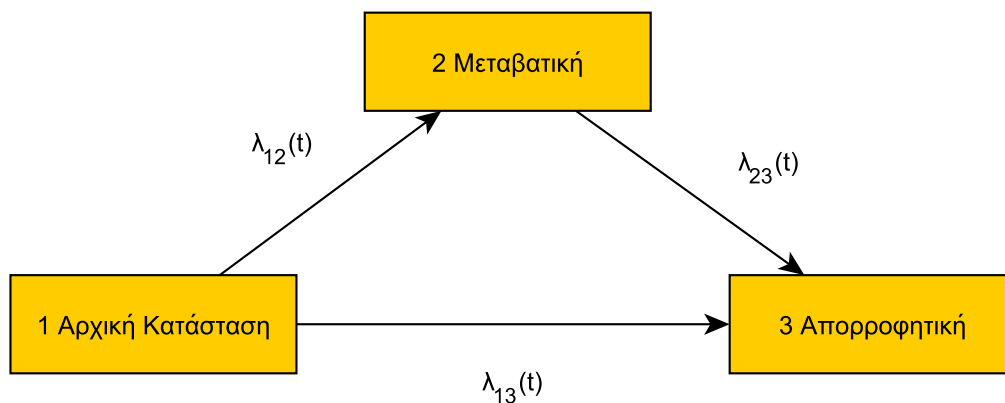
2. Έστω  $p_{ii}(t) = P(X(t) = i | X(0) = i)$ , η πιθανότητα επιστροφής της αλυσίδας στην κατάσταση  $i$  μετά από χρονικό διάστημα  $t$  και

$T_i^+ = \inf(t > 0 : X(t) = i | X(0) = i)$ , ο λεγόμενος χρόνος πρώτης επανόδου της αλυσίδας στην κατάσταση  $i$ .

Αν  $P(T_i^+ < \infty | X(0) = i) < 1$  τότε η κατάσταση  $i$  λέγεται παροδική/μεταβατική, και μάλιστα,  $\int_0^\infty p_{ii}(t) dt < \infty$ .

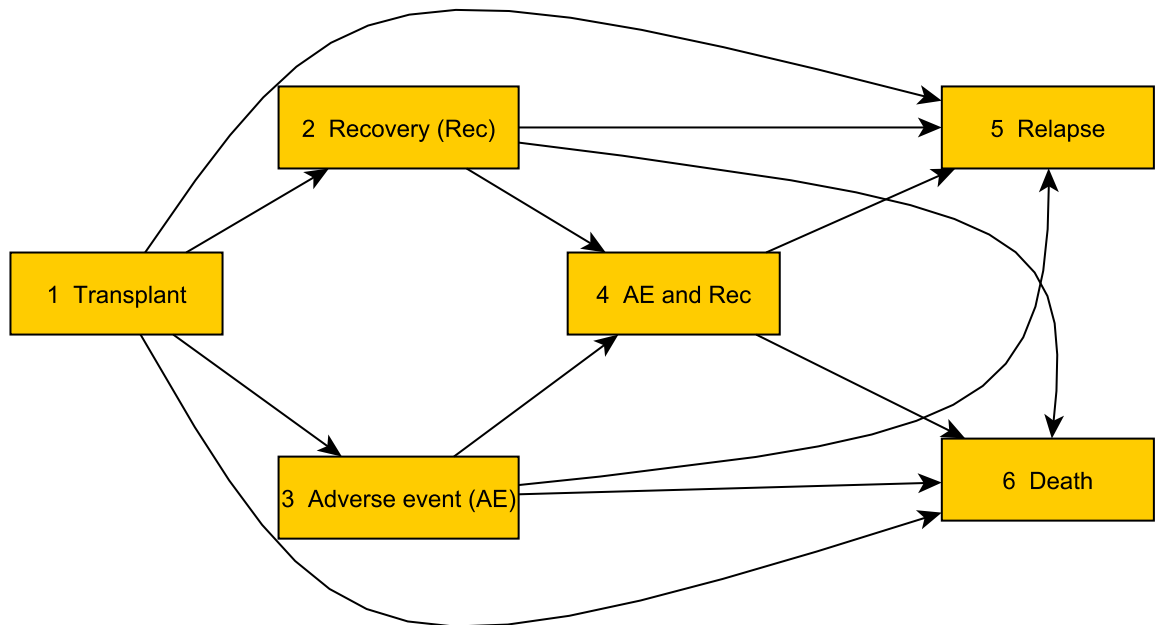
Παρακάτω, βλέπουμε την απλούστερη μορφή ενός μοντέλου multi-state (Σχήμα 41).

Σχήμα 41: Uni-directional Μοντέλο Illness-Death:



Επίσης παραθέτουμε και ένα πιο σύνθετο μοντέλο που αφορά την διάρκεια ζωής ασθενών με καρκίνο του αίματος που έχουν υποστεί κάποια θεραπεία. Τα δεδομένα αυτά τα παρείχε η EBMT (European Group for Blood and Marrow Transplantation) και βρίσκονται στο σύνολο δεδομένων της R , ebmt4 (De Wreede et al. , 2011) στην βιβλιοθήκη mstate (Putter et al. , 2016).

Σχήμα 42: Σύνθετο μοντέλο multi-state:



Οι ενδιάμεσες/μεταβατικές καταστάσεις είναι οι Recovery (Rec), Adverse Event (AE) και ένας συνδυασμός των υπολοίπων και οι Relapse καθώς και Death αποτελούν τις απορροφητικές καταστάσεις του παραπάνω μοντέλου.

#### Υποθέσεις Εργασίας:

1. Θα αναφερθούμε μόνο σε uni-directional μοντέλα, δηλαδή για κάθε μεταβατική κατάσταση  $j$  στην οποία μπορεί να βρεθεί μια μονάδα ενδιαφέροντος από μια επίσης μεταβατική  $i$ , η πορεία  $j \Rightarrow i$  είναι αδύνατη.
2. Θεωρούμε τους μηχανισμούς δεξιά αποκομμένων παρατηρήσεων ανεξάρτητους με τον χρόνο επιβίωσης.
3. Η παραπάνω διαδικασία εξελίσσεται σε συνεχή χρόνο και η ακριβής χρονική στιγμή κατά την οποία γίνεται η μετάβαση είναι πάντα γνωστή και την θεωρούμε απειροελάχιστη. Η στοχαστική διαδικασία, δηλαδή, είναι δεξιά συνεχής.
4. Τα μοντέλα ανταγωνιστικών κινδύνων/competing risks αποτελούν ειδική περίπτωση των μοντέλων multi-state χωρίς καμία μεταβατική/παροδική κατάσταση.

Σύμφωνα με τις παραπάνω υποθέσεις, είμαστε σε θέση να ορίσουμε κάποιες συναρτήσεις που θα παίξουν καίριο λόγο στην ανάλυση των μοντέλων multi-state.

Αρχικά, ορίζουμε την πιθανότητα μια μονάδας ενδιαφέροντος μας να βρίσκεται στην κατάσταση  $j$  την χρονική στιγμή  $t$ , δεδομένου ότι βρισκόταν στην κατάσταση  $i$  την χρονική στιγμή  $s$ . Η παραπάνω πιθανότητα ονομάζεται πιθανότητα μετάβασης και ορίζεται ως εξής:

$$P_{ij}(s, t) = P(X(t) = j | X(s) = i, F_s)$$

όπου  $F_s = \sigma(X(u) : 0 \leq u \leq s)$ , η μικρότερη  $\sigma$ -άλγεβρα που παράγεται από τις καταστάσεις της στοχαστικής διαδικασίας πριν και κατά την χρονική στιγμή  $s$ . Με άλλα λόγια, εκφράζει το παρελθόν της διαδικασίας στο χρονικό διάστημα  $[0, s]$ .

Επίσης, ορίζουμε (Geskus, 2016):

$$\lambda_{ij}(t) = \lim_{dt \rightarrow 0} \frac{P_{ij}(t, t + dt) - P_{ij}(t, t)}{dt} = \lim_{dt \rightarrow 0} \frac{P(X(t + dt) = j | X(t) = i, F_t) - P(X(t) = j | X(t) = i, F_t)}{dt}$$

Η παραπάνω συνάρτηση διακινδύνευσης εκφράζει την πιθανότητα μετάβασης, από τη  $i \Rightarrow j$  σε ένα στοιχειώδη χρόνο, προς τον χρόνο αυτό ή αλλιώς την τάση/ρυθμό των μονάδων ενδιαφέροντος να μεταβούν από την  $i$  στην  $j$ , στο χρονικό διάστημα  $(t, t + dt)$ . Η συνάρτηση αυτή λέγεται επίσης, για λόγους που θα δούμε στην συνέχεια, και transition intensity (τάση μετάβασης).

Στην συνέχεια, μπορούμε να ορίσουμε και την σωρευτική συνάρτηση διακινδύνευσης για την μετάβαση  $i \Rightarrow j$ , ως εξής:

$$\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(u) du$$

Η παραπάνω συνάρτηση εκφράζει τον συνολικό αναμενόμενο αριθμό μεταβάσεων  $i \Rightarrow j$  μέχρι την χρονική στιγμή  $t$  αν η μετάβαση αυτή μπορεί να επαναλαμβάνεται.

Είναι φανερό ότι η ποσότητα  $F_s$  αυξάνει σημαντικά την πολυπλοκότητα των υπολογισμών μας για τα παραπάνω μοντέλα με αποτέλεσμα να δυσχεραίνεται η ανάλυσή τους. Για τον σκοπό αυτό, θα αναφέρουμε δύο απλοποιημένα μοντέλα multi-state, το δεύτερο εκ των οποίων θα μας απασχολήσει σημαντικά στην συνέχεια.

## 3.2 Δύο απλοποιημένα μοντέλα multi-state

### Semi-Markov Models:

Τα παραπάνω μοντέλα, τα οποία ονομάζονται και embedded (ενσωματωμένα) Markov, είναι τα μοντέλα στα οποία η πιθανότητα μετάβασης  $P_{ij}(s, t)$  εξαρτάται μόνο από τον χρόνο  $T_{0i} = \inf(t > 0 : X(t) = i)$  μετάβασης της στοχαστική διαδικασίας στην κατάσταση  $i$ .

Στην περίπτωση αυτή, η νέα συνάρτηση διακινδύνευσης γίνεται (Andersen and Keiding, 2011):

$$\lambda_{ij}(t, t_{01}) = \lim_{dt \rightarrow 0} \frac{P(X(t+dt) = j | X(t) = i, T_{01} = t_{01})}{dt}$$

### Μη-ομοιογενείς Μαρκοβιανές Αλυσίδες Συνεχούς Χρόνου:

Όπως προείπαμε, στα μοντέλα multi-state που θα μελετήσουμε, αναφερόμαστε σε στοχαστικές διαδικασίες συνεχούς χρόνου με πεπερασμένο πλήθος καταστάσεων. Υποθέτουμε, τώρα ότι για τις πιθανότητες μετάβασης μεταξύ των διάφορων καταστάσεων ισχύει η ισχυρή μαρκοβιανή ιδιότητα. Δηλαδή, θεωρούμε ότι η πιθανότητα η μονάδα ενδιαφέροντος μας να βρίσκεται στην κατάσταση  $j$  την χρονική στιγμή  $t$  δεδομένου ότι βρισκόταν στην  $i$  την στιγμή  $s$ , εξαρτάται μόνο από την θέση της διαδικασίας την στιγμή  $s$ . Επίσης, θεωρούμε ότι, προσεγγιστικά, ο χρόνος αναμονής μιας μονάδας ενδιαφέροντος σε μια μη-απορροφητική κατάσταση της στοχαστικής διαδικασίας ακολουθεί την εκθετική κατανομή. Μάλιστα, θεωρούμε ότι οι εκθετικές αυτές κατανομές είναι ανεξάρτητες μεταξύ τους. Η παραπάνω μαρκοβιανή στοχαστική διαδικασία συνεχούς χρόνου και πεπερασμένων καταστάσεων λέγεται και μαρκοβιανή αλυσίδα συνεχούς χρόνου και οι πιθανότητες μετάβασης των διαφόρων καταστάσεων της, δίνονται από τον εξής απλουστευμένο τύπο (Andersen et al. , 2002):

$$P_{ij}(s, t) = P(X(t) = j | X(s) = i)$$

Η συνάρτηση διακινδύνευσης μεταξύ των καταστάσεων ή αλλιώς η ανά μετάβαση συνάρτηση διακινδύνευσης (transition hazard), ομοίως απλοποιείται αρκετά:

$$\lambda_{ij}(t) = \lim_{dt \rightarrow 0} \frac{P(X(t+dt) = j | X(t) = i)}{dt}$$

και αντίστοιχα η συνάρτηση σωρευτικής διακινδύνευσης:

$$\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(u) du$$

Είναι αξιοσημείωτο, επίσης, ότι όλα τα μοντέλα ανταγωνιστικών κινδύνων είναι μαρκοβιανά λόγω της απουσίας της παρελθούσας ιστορίας τους (event history). Με άλλα λόγια, στα παραπάνω μοντέλα είναι δυνατή μόνο μια μετάβαση και άρα η σ-άλγεβρα  $F_s$  στην ουσία εκφυλίζεται.

### 3.3 Μαρκοβιανές Αλυσίδες Συνεχούς Χρόνου

Στο πλαίσιο των μαρκοβιανών αλυσίδων συνεχούς χρόνου, μας ενδιαφέρει ο προσδιορισμός του πίνακα πιθανοτήτων μετάβασης (transition matrix)  $P(s, t) = (P_{ij}(s, t))$ , όπου  $P_{ij}(s, t) = P(X(t) = j | X(s) = i)$ . Από την θεωρία των Μαρκοβιανών Αλυσίδων Συνεχούς Χρόνου αποδεικνύεται ότι για την εύρεση των παραπάνω πιθανοτήτων είναι αναγκαία η αρχική κατανομή της αλυσίδας και ο πίνακας  $Q$  των τάσεων μετάβασης (transition intensities).

Ο πίνακας αυτός σε μια αλυσίδα  $n$  καταστάσεων, ορίζεται ως εξής:

$$q_{ij} = \lim_{dt \rightarrow 0} \frac{P(X(t+dt) = j | X(t) = i)}{dt}, \quad i \neq j$$

και έχει τις εξής ιδιότητες:

- 1)  $0 \leq -q_{ii} < \infty, \forall i = 1, 2, \dots, n$
- 2)  $q_{ij} \geq 0 \quad \forall i \neq j : i, j = 1, 2, \dots, n$
- 3) Το άθροισμα των στοιχείων κάθε γραμμής του πίνακα  $Q$  είναι ίσο με 0.

Δηλαδή:

$$q_{ii} = - \sum_{j \neq i}^n q_{ij}$$

Αν ο παραπάνω πίνακας είναι ανεξάρτητος του χρόνου, τότε από τις εξισώσεις Kolmogorov (Norris, 1997) έχουμε:

$$\frac{dP(t)}{dt} = P(t)Q \quad (\text{Forward Kolmogorov Equation})$$

$$\frac{dP(t)}{dt} = QP(t) \quad (\text{Backward Kolmogorov Equation})$$

Από την επίλυση των παραπάνω εξισώσεων, προκύπτει ο πίνακας  $P(t) = (P_{ij}(0, t))$ , όπου το κάθε  $(i, j)$  στοιχείο αναπαριστά την πιθανότητα  $P(X(t) = j | X(0) = i)$ . Μάλιστα, επιλύοντας το παραπάνω σύστημα Συνήθων Διαφορικών Εξισώσεων (Σ.Δ.Ε.) προκύπτει ότι:

$$P(t) = e^{tQ} = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}$$

Επίσης, αποδεικνύεται ότι για την μη-απορροφητική κατάσταση  $i$  μιας μαρκοβιανής αλυσίδας συνεχούς χρόνου, ισχύει ότι:

$$q_{ij} = v_i P_{ij}$$

όπου τα  $q_{ij}$  και  $P_{ij}$  είναι τα στοιχεία των προαναφερθέντων πινάκων και  $v_i$  είναι η παράμετρος της εκθετικής κατανομής που περιγράφει τον χρόνο αναμονής της αλυσίδας στην κατάσταση αυτή.

Τέλος, προκύπτουν οι εξής σχέσεις (Sigman, 2016):

$$\lim_{h \rightarrow 0} \frac{1 - P_{ij}(h)}{h} = v_i$$

$$\lim_{h \rightarrow 0} \frac{P_{ij}(h)}{h} = q_{ij}$$

Οι παραπάνω, ιδιότητες ωστόσο, αφορούν ομοιογενείς μαρκοβιανές αλυσίδες συνεχούς χρόνου, όπου τα στοιχεία του πίνακα  $Q$  είναι ανεξάρτητα του χρόνου.

### 3.3.1 Εκτιμήτριες Aalen-Johansen

Στα μοντέλα multi-state τα οποία έχουμε μοντελοποιήσει χρησιμοποιώντας μαρκοβιανές αλυσίδες συνεχούς χρόνου, τα στοιχεία του πίνακα  $Q$  (τάσεις μεταβάσης/transition intensities) είναι τέτοια ώστε (Andersen et al. , 2002):

$$q_{ij}(t) = \lambda_{ij}(t) = \lim_{dt \rightarrow 0} \frac{P(X(t+dt) = j | X(t) = i)}{dt}, \quad i \neq j$$

Η αλυσίδα μας λοιπόν είναι μη-ομοιογενής μαρκοβιανή συνεχούς χρόνου ή ICTMC (Inhomogenous Continuous Time Markov Chain). Τα στοιχεία του πίνακα  $Q(t)$  έχουν παρόμοιες ιδιότητες με αυτές του παραπάνω πίνακα  $Q$  για την περίπτωση ομοιογενούς αλυσίδας  $n$  καταστάσεων.

1)  $q_{ij}(t) > 0 \quad i \neq j$

2)  $q_{ii}(t) < 0$

3) Το άθροισμα των στοιχείων κάθε γραμμής του πίνακα  $Q(t)$  είναι ίσο με 0.

Δηλαδή:

$$q_{ii}(t) = - \sum_{j \neq i}^n q_{ij}(t)$$

4) Τα στοιχεία του πίνακα  $Q(t)$  ως συναρτήσεις του χρόνου είναι δεξιά συνεχής:

$$\lim_{h \rightarrow 0} q_{ij}(t+h) = q_{ij}(t), \forall t \geq 0$$

και:

$$\int_0^1 q_{ij}(t) dt < \infty$$

Και στις ICTMC με δεδομένο τον πίνακα  $Q(t)$  και μια αρχική συνθήκη  $P(s, s) = I$ , είμαστε σε θέση να προσδιορίσουμε τον πίνακα πιθανοτήτων μετάβασης

$P(s, t) = (P_{ij}(s, t))$ , όπου το  $(i, j)$  στοιχείο του πίνακα είναι η πιθανότητα

$P(X(t) = j | X(s) = i)$ .

Προκύπτει, λοιπόν, το εξής σύστημα μερικών διαφορικών εξισώσεων, Charman-Kolmogorov (Aalen and Johansen, 1978):

$$\frac{\partial P(s, t)}{\partial s} = -Q(s)P(s, t)$$

$$\frac{\partial P(s, t)}{\partial t} = P(s, t)Q(t)$$

Από την επίλυση του παραπάνω συστήματος Μερικών Διαφορικών Εξισώσεων (Μ.Δ.Ε.) προκύπτουν τα στοιχεία του πίνακα  $P(s, t)$ , τα οποία είναι απολύτως συνεχή ως συναρτήσεις των  $s$  και  $t$ , ως εξής:

$$P(s, t) = \prod_{[s, t]} (I + Q(u)du) \quad 0 \leq s \leq t \leq 1$$

Επίσης, αν  $B_{ij}(t) = \int_0^t q_{ij}(s)ds$ , η παραπάνω σχέση γίνεται:

$$P(s, t) = \prod_{[s, t]} (I + dB(u)) \quad 0 \leq s \leq t \leq 1$$

Πρέπει να σημειώσουμε, επίσης, ότι τα στοιχεία του παραπάνω πίνακα  $B$ , όπως και του  $Q$ , αθροίζονται στο 0.

Το παραπάνω σύμβολο λέγεται product integral και ορίζεται ως εξής:  
Εστω διαμέριση  $t_0, t_1, \dots, t_K$  του συνόλου  $[s, t]$ .

Τότε:

$$\prod_{[s, t]} (I + dB(u)) = \lim_{M \rightarrow 0} \prod_{k=1}^K (I + B(t_k) - B(t_{k-1}))$$

όπου  $M = \max_{k=1,2,\dots,K} |t_k - t_{k-1}|$  είναι το μήκος του μεγαλύτερου υποδιαστήματος  $t_k - t_{k-1}$ ,  $\forall k = 1, 2, \dots, K$ .

Πλέον, λοιπόν, είμαστε σε θέση να μεταφέρουμε την θεωρία των ICTMC στα multi-state προβλήματα που μελετούμε (Klein and Shu, 2002), με  $q_{ij}(t) = \lambda_{ij}(t)$  και  $B_{ij}(t) = \Lambda_{ij}(t)$ , ώστε να εκτιμήσουμε τις παραπάνω πιθανότητες μετάβασης  $i \Rightarrow j$ , χρησιμοποιώντας τις εκτιμήτριες Aalen-Johansen:

$$\hat{P}_{ij}(s, t) = \prod_{s < t_k \leq t} (I + \Delta \hat{\Lambda}_{ij}(t_k)) = \prod_{k=1}^K (I + (\hat{\Lambda}_{ij}(t_k) - \hat{\Lambda}_{ij}(t_{k-1})))$$

όπου  $t_0, t_1, \dots, t_k, \dots, t_K$  μια διαμέριση του  $[s, t]$  και  $\hat{\Lambda}_{ii}(t) = -\sum_{j \neq i} \hat{\Lambda}_{ij}(t)$ . Προφανώς, όσο πιο λεπτή η διαμέριση του  $[s, t]$ , τόσο πιο ακριβής θα είναι η εκτιμήτρια Aalen-Johansen.

### Μη-παραμετρική εκτίμηση των $\Lambda_{ij}(t)$ :

Αρχικά, μπορούμε να εκτιμήσουμε μη-παραμετρικά τις παραπάνω πιθανότητες, χρησιμοποιώντας τις εκτιμήτριες Nelson-Aalen για τις σωρευτικές συναρτήσεις διακινδύνευσης κάθε μετάβασης. (Ferguson et al. , 2012).

Οι μη-παραμετρικές εκτιμήτριες Nelson-Aalen για κάθε μετάβαση προκύπτουν ως εξής:

$$\hat{\Lambda}_{ij}^{NA}(t) = \sum_{t_k < t} \frac{dN_{ij}(t_k)}{Y_i(t_k)}$$

όπου  $t_k$  είναι οι χρονικές στιγμές κατά τις οποίες καταγράφηκαν μεταβάσεις μεταξύ των καταστάσεων  $i$  και  $j$  της ICTMC,  $dN_{ij}(t_k)$  είναι ο παρατηρούμενος αριθμός μεταβάσεων  $i \Rightarrow j$  κατά την χρονική στιγμή  $t_k$  και  $Y_i(t_k)$  είναι ο αριθμός των μονάδων ενδιαφέροντος που βρίσκονται σε κίνδυνο μετάβασης από την κατάσταση  $i$  στην  $j$ .

Επίσης, ισχύει ότι:

$$\hat{\Lambda}_{ii}^{NA}(t) = - \sum_{j \neq i} \hat{\Lambda}_{ij}^{NA}(t)$$

### Εκτίμηση των $\Lambda_{ij}(t)$ χρησιμοποιώντας το ημι-παραμετρικό μοντέλο του Cox:

Επίσης, σε κάθε μετάβαση της παραπάνω μαρκοβιανής αλυσίδας συνεχούς χρόνου, μπορούμε να προσαρμόσουμε ένα ημι-παραμετρικό μοντέλο Cox το οποίο να μοντελοποιεί την αντίστοιχη συνάρτηση διακινδύνευσης  $\lambda_{ij}(t|\vec{z})$ , παρουσία διανύσματος συμμεταβλητών  $\vec{z}$  για κάθε μονάδα  $i = 1, 2, \dots, n$  και αντίστοιχων παραμέτρων  $\vec{b}_{ij}$ , όπου  $\vec{z}$  και  $\vec{b}$  είναι διανύσματα  $p \times 1$ .

Τα παραπάνω μοντέλα παλινδρόμησης θα έχουν την εξής μορφή, για τις ανά μετάβαση συναρτήσεις διακινδύνευσης:

$$\lambda_{ij}(t|\vec{z}) = \lambda_{ij,0}(t) \exp(\vec{b}_{ij}^T \vec{z})$$

με τις ανά μετάβαση συναρτήσεις σωρευτικής διακινδύνευσης να είναι:

$$\Lambda_{ij}(t|\vec{z}) = \Lambda_{ij,0}(t) \exp(\vec{b}_{ij}^T \vec{z})$$

Επομένως οι εκτιμήσεις θα είναι:

$$\hat{\Lambda}_{ij}(t|\vec{z}) = \hat{\Lambda}_{ij,0}(t) \exp(\vec{b}_{ij}^T \vec{z})$$

όπου οι συντελεστές  $\vec{b}_{ij}^T$  προσδιορίζονται από το αντίστοιχο μοντέλο παλινδρόμησης κατά Cox για την μετάβαση  $i \Rightarrow j$ .



Για την βασική (baseline) συνάρτηση της παραπάνω σωρευτικής συνάρτησης κινδύνου (εκτιμήτρια Breslow):

$$\hat{\Lambda}_{ij,0}(t) = \sum_{k=1}^K \frac{N_{ij}(t_k) - N_{ij}(t_{k-1})}{S_{ij}^{(0)}(\hat{\vec{b}}_{ij}, s)}$$

όπου  $0 = t_0 < t_1 < \dots < t_K = t$  μια διαμέριση του  $[0, t]$ , όπου όσο λεπτότερη είναι τόσο καλύτερη θα είναι η εκτίμηση. Επίσης,  $N_{ij}(s)$  είναι ο αριθμός των μεταβάσεων  $i \Rightarrow j$  στο χρονικό διάστημα  $[0, s]$  και

$$S_{ij}^{(0)}(\vec{b}_{ij}, t) = \sum_{l=1}^n \exp(\vec{b}_{ij} \vec{z}_l) Y_{i;l}(t)$$

όπου  $n$  ο αριθμός των μονάδων ενδιαφέροντος μας και  $Y_{i;l}(t)$  είναι μια δείκτρια μεταβλητή που παίρνει την τιμή 1 αν η μονάδα ενδιαφέροντος  $l$  βρίσκεται σε κίνδυνο να μεταβεί σε κατάσταση εκτός της  $i$  την χρονική στιγμή  $t$  και αλλιώς 0.

### 3.4 Uni-Directional Μοντέλα Illness-Death

#### 3.4.1 Ανά μετάβαση συναρτήσεις διακινδύνευσης

Ιδιαίτερη έμφαση θα δώσουμε στα uni-directional μοντέλα illness-Death, τους αντίστοιχους πίνακες των οποίων, βλέπουμε παρακάτω (Jackson, 2011):

$$Q(t) = \lambda_{illness-death}(t) = \begin{bmatrix} -\lambda_{12}(t) - \lambda_{13}(t) & \lambda_{12}(t) & \lambda_{13}(t) \\ 0 & -\lambda_{23}(t) & \lambda_{23}(t) \\ 0 & 0 & 0 \end{bmatrix}$$

με  $B_{ij}(t) = \Lambda_{ij}(t) = \int_0^t \lambda_{ij}(u) du$

$$\Lambda_{illness-death}(t) = \begin{bmatrix} -\Lambda_{12}(t) - \Lambda_{13}(t) & \Lambda_{12}(t) & \Lambda_{13}(t) \\ 0 & -\Lambda_{23}(t) & \Lambda_{23}(t) \\ 0 & 0 & 0 \end{bmatrix}$$

Όπως προκύπτει από τα παραπάνω, οι ανά αιτία (cause specific) συναρτήσεις διακινδύνευσης των μοντέλων ανταγωνιστικών κινδύνων είναι στην ουσία οι ανά μετάβαση συναρτήσεις διακινδύνευσης (transition hazards)  $\lambda_{0k}(t)$  μεταξύ των καταστάσεων  $0 \Rightarrow k \forall k \in D$ . Γενικεύοντας, δηλαδή, από τα μοντέλα που μελετήσαμε στο Κεφάλαιο 2, είμαστε σε θέση να ελέγξουμε την υπόθεση ισότητας των συντελεστών κάποιου ή κάποιων συμμεταβλητών για διαφορετικές μεταβάσεις αλλά και το αν οι βασικές συναρτήσεις διακινδύνευσης μεταξύ κάποιων μεταβάσεων είναι ανάλογες ή όχι. Οι παραπάνω υποθέσεις αν ισχύουν, συμβάλλουν σημαντικά στην μείωση του συνολικού αριθμού παραμέτρων που απαιτούνται για τον προσδιορισμό των παραπάνω συναρτήσεων.

### Ανά μετάβαση συναρτήσεις διακινδύνευσης:

Οι συναρτήσεις διακινδύνευσης για τις τρεις διαφορετικές δυνατές μεταβάσεις, θεωρώντας τρία ξεχωριστά μοντέλα Cox, έχουν την εξής μορφή (Στρωματοποιημένη/Stratified Παλινδρόμηση κατά Cox):

$$\lambda_{13}(t|\vec{z}) = \lambda_{13,0}(t)\exp(\vec{b}_{13}^T\vec{z}) \quad [1 \Rightarrow 3]$$

$$\lambda_{12}(t|\vec{z}) = \lambda_{12,0}(t)\exp(\vec{b}_{12}^T\vec{z}) \quad [1 \Rightarrow 2]$$

$$\lambda_{23}(t|\vec{z}) = \lambda_{23,0}(t)\exp(\vec{b}_{23}^T\vec{z}) \quad [2 \Rightarrow 3]$$

### Υπόθεση αναλογικότητας:

Θα μπορούσαμε να υποθέσουμε ότι οι βασικές συναρτήσεις διακινδύνευσης για τις μεταβάσεις  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$  είναι ανάλογες. Αυτό είναι ισοδύναμο με το να ομαδοποιήσουμε τις παραπάνω μεταβάσεις σε μια ομάδα και να χρησιμοποιήσουμε την μεταβατική/ενδιάμεση κατάσταση ως μια συμμεταβλητή που εξαρτάται από τον χρόνο. Τα νέα μοντέλα Cox για τις συναρτήσεις διακινδύνευσης των μεταβάσεων  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$  θα γίνουν:

$$\lambda_{13}(t|\vec{z}) = \lambda_{3,0}(t)\exp(\vec{b}_{13}^T\vec{z})$$

$$\lambda_{23}(t|\vec{z}) = \lambda_{3,0}(t)\exp(\vec{b}_{23}^T\vec{z} + d)$$

όπου  $\exp(d)$  είναι η ενδεχόμενη σταθερά αναλογίας των βασικών συναρτήσεων.

## 3.4.2 Πιθανότητες Μετάβασης

### Πιθανότητες Μετάβασης από την κατάσταση 2 :

$$P_{22}(s, t) = \exp\left(-\int_s^t \lambda_{23}(v)dv\right)$$

Η παραπάνω σχέση είναι η πιθανότητα μια μονάδα ενδιαφέροντος να παραμείνει από την χρονική στιγμή  $s$  μέχρι και την χρονική στιγμή  $t$  στην κατάσταση 2.

$$P_{23}(s, t) = \int_s^t P_{22}(s, u)\lambda_{23}(u)du$$

Η παραπάνω πιθανότητα προκύπτει ως εξής: Από την χρονική στιγμή  $s$  έως μια  $u \in (s, t)$  η μαρκοβιανή αλυσίδα παραμένει στην κατάσταση 2 (Illness) από όπου μεταβαίνει με πιθανότητα  $\lambda_{23}(u)$  σε ένα απειροστό χρονικό διάστημα στην κατάσταση 3 στην οποία θα παραμείνει με πιθανότητα 1 μέχρι την χρονική στιγμή  $u$  καθώς η 3 είναι απορροφητική.

Επιπλέον:

$$P_{22}(s, t) + P_{23}(s, t) = 1$$

### Πιθανότητες Μετάβασης από την κατάσταση 1 :

$$P_{11}(s, t) = \exp\left(-\int_s^t \lambda_{12}(v)dv - \int_s^t \lambda_{13}(v)dv\right)$$

Η σχέση αυτή εκφράζει την πιθανότητα η μονάδα ενδιαφέροντος μας να παραμείνει στην αρχική της κατάσταση (κατάσταση 1) στο χρονικό διάστημα  $(s, t)$ .

$$P_{12}(s, t) = \int_s^t P_{11}(s, u)\lambda_{12}(u)P_{22}(u, t)du$$

Έχουμε, δηλαδή, ότι η μονάδα μας παραμένει στην κατάσταση 1 στο διάστημα  $(s, u)$  από όπου μεταβαίνει με πιθανότητα  $\lambda_{12}(u)$  σε απειροστό χρόνο στην κατάσταση 2, στην οποία παραμένει στο διάστημα  $(u, t)$ .

$$P_{13}^1(s, t) = \int_s^t P_{11}(s, u)\lambda_{13}(u)du$$

Η πιθανότητα μετάβασης μιας μονάδας ενδιαφέροντος απ' ευθείας από την 1  $\Rightarrow$  3 χωρίς να περάσει πρώτα από την κατάσταση 2. Η εξήγηση είναι όμοια με αυτή της πιθανότητας  $P_{23}(s, t)$ .

$$P_{13}^2(s, t) = \int_s^t P_{11}(s, u)\lambda_{12}(u)P_{23}(u, t)du$$

Τέλος, η παραπάνω σχέση εκφράζει την πιθανότητα μετάβασης μιας μονάδας ενδιαφέροντος από την 1  $\Rightarrow$  3 αφού περάσει, όμως, και από την κατάσταση 2. Η μονάδα μας παραμένει στην κατάσταση 1 από την  $s$  μέχρι κάποια  $u \in (s, t)$  κατά την οποία με πιθανότητα  $\lambda_{12}(u)$  μεταβαίνει στην κατάσταση 2. Από εκεί κάποια στιγμή  $r \in (u, t)$  με την ίδια λογική με παραπάνω, μεταβαίνει στην 3 στην οποία και παραμένει καθ' όλο το διάστημα  $(r, t)$  με πιθανότητα 1 καθώς η 3 είναι απορροφητική. Η τελευταία πιθανότητα που περιγράψαμε, με βάση και τα παραπάνω, δεν είναι κάτι άλλο παρά η  $P_{23}(u, t)$ .

Επίσης:

$$P_{13}(s, t) = P_{13}^1(s, t) + P_{13}^2(s, t)$$

αλλά και:

$$P_{11}(s, t) + P_{12}(s, t) + P_{13}(s, t) = 1$$

### 3.4.3 Εκτιμήσεις των πιθανοτήτων μετάβασης

Ανακεφαλαιώνοντας, έχουμε τα εξής:

Μη-παραμετρική εκτίμηση των  $\Lambda_{ij}(t)$  χρησιμοποιώντας τις Nelson-Aalen:

$$\hat{\Lambda}_{ij}^{NA}(t) = \sum_{t_k < t} \frac{dN_{ij}(t_k)}{Y_i(t_k)}$$

Προσαρμογή ημι-παραμετρικών μοντέλων Cox στις ανά μετάβαση συναρτήσεις διακινδύνευσης και σωρευτικής διακινδύνευσης:

$$\hat{\lambda}_{ij}(t|\vec{z}) = \hat{\lambda}_{ij,0}(t) \exp(\vec{b}_{ij}^T \vec{z})$$

$$\hat{\Lambda}_{ij}(t|\vec{z}) = \hat{\Lambda}_{ij,0}(t) \exp(\vec{b}_{ij}^T \vec{z})$$

Με τον πίνακα των εκτιμημένων ολοκληρωμάτων των τάσεων μετάβασης/transition intensities (Q-matrix), να έχει την μορφή:

$$\hat{\Lambda}(t) = \begin{bmatrix} -\hat{\Lambda}_{12}(t) - \hat{\Lambda}_{13}(t) & \hat{\Lambda}_{12}(t) & \hat{\Lambda}_{13}(t) \\ 0 & -\hat{\Lambda}_{23}(t) & \hat{\Lambda}_{23}(t) \\ 0 & 0 & 1 \end{bmatrix}$$

Ο γενικός τύπος των πιθανοτήτων μετάβασης  $i \Rightarrow j$  που αναζητούμε είναι ο εξής:

$$\hat{P}_{ij}(s, t) = \prod_{s < t_k \leq t} (I + \Delta \hat{\Lambda}_{ij}(t_k)) = \prod_{k=1}^K (I + (\hat{\Lambda}_{ij}(t_k) - \hat{\Lambda}_{ij}(t_{k-1})))$$

Με πίνακα πιθανοτήτων μετάβασης:

$$\hat{P}(s, t) = \begin{bmatrix} \hat{P}_{11}(s, t) & \hat{P}_{12}(s, t) & \hat{P}_{13}(s, t) \\ 0 & \hat{P}_{22}(s, t) & \hat{P}_{23}(s, t) \\ 0 & 0 & 1 \end{bmatrix}$$

και:

$$I + \Delta \hat{\Lambda}(t) = \begin{bmatrix} 1 - \Delta \hat{\Lambda}_{12}(t) - \Delta \hat{\Lambda}_{13}(t) & \Delta \hat{\Lambda}_{12}(t) & \Delta \hat{\Lambda}_{13}(t) \\ 0 & 1 - \Delta \hat{\Lambda}_{23}(t) & \Delta \hat{\Lambda}_{23}(t) \\ 0 & 0 & 1 \end{bmatrix}$$

Έστω διαμέριση  $s = t_0 < t_1 < \dots < t_K = t$  του συνόλου  $[s, t]$ . Προφανώς, όσο λεπτότερη η διαμέριση τόσο καλύτερες οι αντίστοιχες εκτιμήτριες. Συνδυάζοντας τις παραπάνω σχέσεις, τελικά θα προκύψουν τα εξής:

**Χρησιμοποιώντας τις εκτιμήτριες Nelson-Aalen  $\hat{\Lambda}_{ij}^{NA}(t)$  :**

$$\hat{P}_{11}(s, t) = \prod_{k=1}^K (1 - [\hat{\Lambda}_{12}^{NA}(t_k) - \hat{\Lambda}_{12}^{NA}(t_{k-1})] - [\hat{\Lambda}_{13}^{NA}(t_k) - \hat{\Lambda}_{13}^{NA}(t_{k-1})])$$

$$\hat{P}_{12}(s, t) = \prod_{k=1}^K (\hat{\Lambda}_{12}^{NA}(t_k) - \hat{\Lambda}_{12}^{NA}(t_{k-1}))$$

$$\hat{P}_{13}(s, t) = \prod_{k=1}^K (\hat{\Lambda}_{13}^{NA}(t_k) - \hat{\Lambda}_{13}^{NA}(t_{k-1}))$$

$$\hat{P}_{22}(s, t) = \prod_{k=1}^K (1 - [\hat{\Lambda}_{23}^{NA}(t_k) - \hat{\Lambda}_{23}^{NA}(t_{k-1})])$$

$$\hat{P}_{23}(s, t) = \prod_{k=1}^K (\hat{\Lambda}_{23}^{NA}(t_k) - \hat{\Lambda}_{23}^{NA}(t_{k-1}))$$

**Χρησιμοποιώντας τα μοντέλα αναλογικής διακινδύνευσης του Cox:**

$$\hat{\Lambda}_{ij}(t|\vec{z}) = \hat{\Lambda}_{ij,0}(t) \exp(\vec{b}_{ij}^T \vec{z}) :$$

$$\hat{P}_{11}(s, t) = \prod_{k=1}^K (1 - [\hat{\Lambda}_{12}(t_k|\vec{z}) - \hat{\Lambda}_{12}(t_{k-1}|\vec{z})] - [\hat{\Lambda}_{13}(t_k|\vec{z}) - \hat{\Lambda}_{13}(t_{k-1}|\vec{z})])$$

$$\hat{P}_{12}(s, t) = \prod_{k=1}^K (\hat{\Lambda}_{12}(t_k|\vec{z}) - \hat{\Lambda}_{12}(t_{k-1}|\vec{z}))$$

$$\hat{P}_{13}(s, t) = \prod_{k=1}^K (\hat{\Lambda}_{13}(t_k|\vec{z}) - \hat{\Lambda}_{13}(t_{k-1}|\vec{z}))$$

$$\hat{P}_{22}(s, t) = \prod_{k=1}^K (1 - [\hat{\Lambda}_{23}(t_k|\vec{z}) - \hat{\Lambda}_{23}(t_{k-1}|\vec{z})])$$

$$\hat{P}_{23}(s, t) = \prod_{k=1}^K (\hat{\Lambda}_{23}(t_k|\vec{z}) - \hat{\Lambda}_{23}(t_{k-1}|\vec{z}))$$

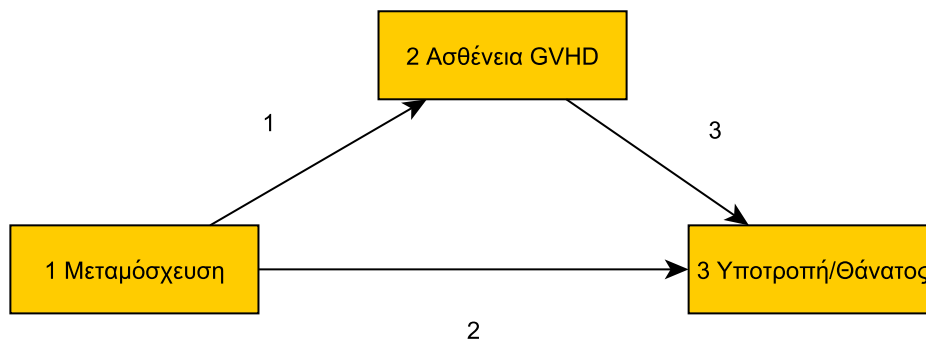
Είναι φανερό, λοιπόν, πως αν προσδιορίσουμε τις ποσότητες  $\hat{\Lambda}_{ij}^{NA}$ , αν θέλουμε να προβούμε σε μια μη-παραμετρική ανάλυση, ή  $\hat{\Lambda}_{ij}(t|\vec{z})$ , αν θέλουμε να συμπεριλάβουμε και συμμεταβλητές στην ανάλυση μας, είμαστε σε θέση να εκτιμήσουμε όλες τις παραπάνω πιθανότητες.

### 3.5 Εφαρμογή σε Uni-directional μοντέλο illness-death

#### 3.5.1 Περιγραφή του προβλήματος

Θα αναλύσουμε τα δεδομένα (Klein and Moeschberger, 1997 και Aalen et al. , 2008) 137 ασθενών με οξεία λευχαιμία που έλαβαν μεταμόσχευση μυελού των οστών από τον Μάρτιο του 1984 μέχρι και τον Ιούνιο του 1989. Καταγράφεται ο χρόνος, σε μέρες, μέχρι τον θάνατο/υποτροπή των ασθενών ή το τέλος της παρακολούθησής τους. Επίσης καταγράφεται η ενδεχόμενη ενδιάμεση ασθένεια acute (οξεία) ή και chronic (χρόνια) graft-versus-host (αντίδραση μοσχεύματος εναντίον ξενιστή) κάποιων ασθενών (Σχήμα 43).

Σχήμα 43: Μεταμόσχευση μυελού των οστών.



Θα κάνουμε χρήση τεσσάρων μεταβλητών που σχετίζονται άμεσα με την συμπεριφορά της μαρκοβιανής αλυσίδας που έχουμε προσαρμόσει στο παραπάνω μοντέλο.

Αυτές είναι:

1. **prtime**: Ο χρόνος που πήρε στον κάθε ασθενή να μεταβεί στην μεταβατική/transient κατάσταση 2 (GVHD) ή ο συνολικός χρόνος που μελετήθηκε ο ασθενής στην περίπτωση που παρέμεινε υγιής (κατάσταση 1) καθ' όλη την διάρκεια παρακολούθησής του.
2. **prstat**: Δίτιμη μεταβλητή: Παίρνει την τιμή 1 αν ο ασθενής μετέβη κατά τον χρόνο μελέτης του από την κατάσταση 2 (GVHD) και την τιμή 0, αλλιώς.
3. **rfstime**: Ο συνολικός χρόνος που πήρε στον ασθενή να αποτύχει (Υποτροπή/Θάνατος-Relapse/Death) ή ο χρόνος της δεξιάς του αποκοπής.
4. **rfstat**: Δίτιμη μεταβλητή που παίρνει την τιμή 1 αν η αποτυχία του ασθενή παρατηρήθηκε κατά την διάρκεια μελέτης του και 0 αν όχι (δεξιά αποκοπή).

Επίσης, θα προσθέσουμε στην ανάλυση μας κάποιες συμμεταβλητές οι οποίες θα χρησιμοποιηθούν για την προσαρμογή των μοντέλων Cox για τις συναρτήσεις διακινδύνευσης των μεταβάσεων του παραπάνω μοντέλου. Εν προκειμένω, θα κατασκευάσουμε συνολικά τρία μοντέλα Cox για τις μεταβάσεις  $1 \Rightarrow 2$ ,  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$ , αντίστοιχα, με τρεις συμμεταβλητές έκαστος.

Οι συμμεταβλητές που θα χρησιμοποιήσουμε είναι οι εξής:

**1. group:** Παίρνει τρεις τιμές (1-ALL, 2-AML low-risk και 3-AML high risk), ανάλογα με τον τύπο της ασθένειας. Για την παραπάνω συμμεταβλητή, στην ανάλυση μας θα χρησιμοποιήσουμε δύο δίτιμες ψευδομεταβλητές οι οποίες ορίζονται ως εξής:

$$group1 = \begin{cases} 1, & \text{ο ασθενής έχει τύπο ασθένειας ALL-1} \\ 0, & \text{ο ασθενής έχει άλλο τύπο ασθένειας} \end{cases}$$

$$group2 = \begin{cases} 1, & \text{ο ασθενής έχει τύπο ασθένειας AML low-risk-2} \\ 0, & \text{ο ασθενής έχει άλλο τύπο ασθένειας} \end{cases}$$

**2. fab:** Οι ασθενείς με λευχαιμία μπορούν να κατηγοριοποιηθούν και με βάση το κριτήριο FAB (French-American-British) το οποίο βασίζεται στα συνηθισμένα μορφολογικά χαρακτηριστικά τους. Η παραπάνω μεταβλητή είναι δίτιμη και παίρνει την τιμή 1 για τους ασθενείς με τιμή group:AML (low-risk ή high-risk) οι οποίοι έχουν κατάταξη FAB M4 ή M5, και την τιμή 0 σε κάθε άλλη περίπτωση.

**3. mtx:** Δίτιμη μεταβλητή που παίρνει την τιμή 1 αν χρησιμοποιήθηκε ο αντιμεταβολίτης MTX (methotrexate) στον ασθενή ως μέτρο προφύλαξης εναντίον του GVHD και 0 αν όχι.

Ας δούμε τις πρώτες έξι παρατηρήσεις του συνόλου δεδομένων μας:

Πίνακας 27: Δεδομένα Μεταμόσχευσης Μυελού των Οστών.

id	prtime	prstat	rfstime	rfsstat	group1	group2	fab	mtx
1	67	1	2081	0	1	0	0	0
2	139	1	1602	0	1	0	0	0
3	307	1	1496	0	1	0	0	0
4	70	1	1462	0	1	0	0	0
5	236	1	1433	0	1	0	0	0
6	123	1	1377	0	1	0	0	0

Επίσης, στον Πίνακα 28, σε συνδυασμό και με το Σχήμα 43, βλέπουμε τις πιθανές μεταβάσεις της ICTMC μας.

Πίνακας 28: Μεταβάσεις.

from\to	Tplant	GVHD	RelDeath
Tplant	NA	1	2
GVHD	NA	NA	3
RelDeath	NA	NA	NA

Επιπλέον, μπορούμε να δούμε και τον αριθμό των ασθενών στις διάφορες καταστάσεις μετά το πέρας της κλινικής μελέτης (Πίνακας 29):

Πίνακας 29: Συχνότητες.

from/to	Tplant	GVHD	RelDeath	no event	total entering
Tplant	0	71	46	20	137
GVHD	0	0	37	34	71
RelDeath	0	0	0	83	85

Από τον Πίνακα 29 παρατηρούμε, αρχικά, ότι μόνο 20 από τους 137 ασθενείς παρέμειναν υγιείς μέχρι το πέρας της μελέτης τους. Επίσης, 71 μεταβαίνουν στην κατάσταση 2, ασθενούν δηλαδή, και 46 αποτυγχάνουν απευθείας, χωρίς να μεταβούν στην ενδιάμεση κατάσταση. Από τους 71 που μεταβαίνουν στην παροδική 2, οι 37 αποτυγχάνουν και οι υπόλοιποι 34 παραμένουν εκεί, χωρίς να αποτύχουν, μέχρι το τέλος της παρακολούθησής τους.

Επομένως, συνολικά έχουμε (Πίνακας 30):

Πίνακας 30: Συνολικός αριθμός ασθενών ανά μετάβαση.

Συνολικά	Ασθενείς
1->1	20 (14.6 %)
1->2	34 (24.8 %)
1->3	46 (33.6 %)
1->2->3	37 (27 %)
Rel/Death	83 (60.1 %)

Από τον Πίνακα 30 βλέπουμε ότι η πλειοψηφία των ασθενών τελικά αποτυγχάνουν, με τους πιο πολλούς εξ αυτών (24% παραπάνω) να μεταβαίνουν στην 3, χωρίς να περάσουν πρώτα από την μεταβατική 2.



Προκειμένου να αρχίσουμε την ανάλυση των παραπάνω δεδομένων, θα πρέπει να τα φέρουμε σε μια επαυξημένη μορφή. Το νέο σύνολο δεδομένων για τους δύο πρώτους ασθενείς γίνεται (Πίνακας 31):

Πίνακας 31: Επαυξημένα Δεδομένα Μεταμόσχευσης Μυελού των Οστών.

id	from	to	trans	Tstart	Tstop	time	status	group1	group2	fab	mtx	prtime	group1.1	group1.2
1	1	2	1	0	67	67	1	1	0	0	0	67	1	0
1	1	3	2	0	67	67	0	1	0	0	0	67	0	1
1	2	3	3	67	2081	2014	0	1	0	0	0	67	0	0
2	1	2	1	0	139	139	1	1	0	0	0	139	1	0
2	1	3	2	0	139	139	0	1	0	0	0	139	0	1
2	2	3	3	139	1602	1463	0	1	0	0	0	139	0	0

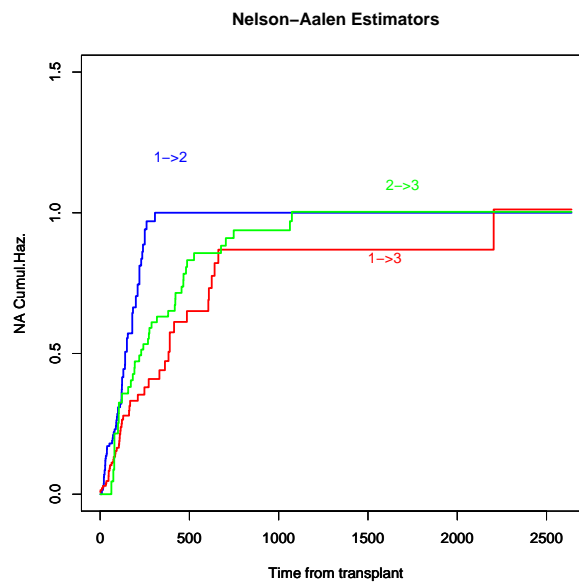
  

group1.3	group2.1	group2.2	group2.3	fab.1	fab.2	fab.3	mtx.1	mtx.2	mtx.3	prtime.1	prtime.2	prtime.3
0	0	0	0	0	0	0	0	0	0	67	0	0
0	0	0	0	0	0	0	0	0	0	0	67	0
1	0	0	0	0	0	0	0	0	0	0	0	67
0	0	0	0	0	0	0	0	0	0	139	0	0
0	0	0	0	0	0	0	0	0	0	0	139	0
1	0	0	0	0	0	0	0	0	0	0	0	139

### 3.5.2 Μη-παραμετρική εκτίμηση των πιθανοτήτων μετάβασης

Ξεκινάμε την ανάλυση μας, εκτιμώντας τις εκτιμήτριες Nelson-Aalen για τις ανά μετάβαση συναρτήσεις σωρευτικής διακινδύνευσης  $\hat{\Lambda}_{ij}^{NA}(t)$ , μέσω της R (De Wreede et al. , 2011):

Σχήμα 44: Μη-παραμετρικές εκτιμήσεις Nelson-Aalen για τις συναρτήσεις  $\hat{\Lambda}_{12}(t)$ ,  $\hat{\Lambda}_{13}(t)$  και  $\hat{\Lambda}_{23}(t)$ .



Από το Σχήμα 44 παρατηρούμε ότι μέχρι την  $s = 2204$  τις υψηλότερες τιμές παίρνει συστηματικά η  $\hat{\Lambda}_{12}^{NA}(t)$ , ενώ η  $\hat{\Lambda}_{13}^{NA}(t)$  την ξεπερνάει, με μέγιστη τιμή 1.012 για χρόνους  $t \geq 2204$ .

Από τα δεδομένα μας, παρατηρούμε ότι τα χρονικά διαστήματα πραγματοποίησης μετάβασης  $1 \Rightarrow 2$  είναι τα πιο μικρά, με αυτά των  $2 \Rightarrow 3$  και  $1 \Rightarrow 3$  να ακολουθούν, αντίστοιχα. Έτσι, ο συνολικός κίνδυνος να πραγματοποιηθεί η μετάβαση ή αλλιώς η ανά μετάβαση συνάρτηση σωρευτικής διακινδύνευσης για την  $1 \Rightarrow 2$  υπερτερεί των υπολοίπων, με τις συναρτήσεις των μεταβάσεων  $2 \Rightarrow 3$  και  $1 \Rightarrow 3$  να παίρνουν μικρότερες τιμές για την συντριπτική πλειοψηφία των χρονικών στιγμών παρακολούθησης των ασθενών.

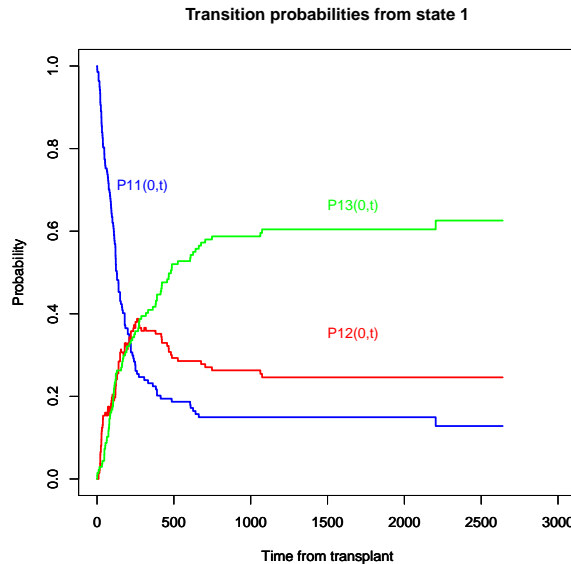
Χρησιμοποιώντας τις παραπάνω εκτιμήτριες, μπορούμε να βρούμε τις εκτιμήτριες Aalen-Johansen για τις πιθανότητες μετάβασης στο παραπάνω μοντέλο illness-death:

$$\hat{P}(X(t) = j | X(s) = i) = \hat{P}_{ij}(s, t) = \prod_{k=1}^K (I + (\hat{\Lambda}_{ij}^{NA}(t_k) - \hat{\Lambda}_{ij}^{NA}(t_{k-1})))$$

**Για τις  $P_{1j}(0, t)$ ,  $\forall j = 1, 2, 3$  :**

Έχουμε, λοιπόν, για τις  $\hat{P}_{11}(0, t)$ ,  $\hat{P}_{12}(0, t)$  και  $\hat{P}_{13}(0, t)$  (Σχήμα 45):

Σχήμα 45: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή που είναι υγιής την χρονική στιγμή  $t = 0$ .

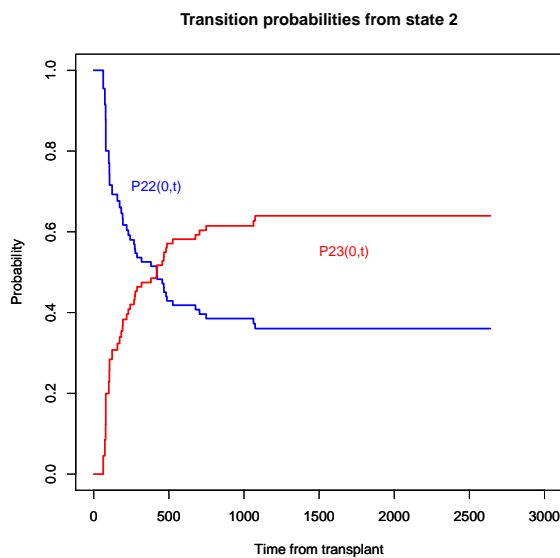


Παρατηρούμε, από το Σχήμα 45, ότι αν ένας ασθενής βρίσκεται την χρονική στιγμή  $s = 0$  στην αρχική του κατάσταση, είναι πιο πιθανό να παραμείνει εκεί μέχρι την  $t = 219$ . Για  $t \geq 219$  επικρατεί η πιθανότητα μετάβασης στην απορροφητική κατάσταση 3, με μέγιστη τιμή 0.626 για  $t \geq 2204$ . Το παραπάνω αποτέλεσμα είναι αναμενόμενο, καθώς 83 από τους 137 ασθενείς που μελετήσαμε, τελικά αποτυγχάνουν. Οπότε, σε βάθος χρόνου, περιμένουμε η πιθανότητα αποτυχίας να είναι σημαντικά μεγαλύτερη των υπολοίπων.

**Για τις  $P_{2j}(0, t)$ ,  $\forall j = 2, 3$  :**

Για τις  $\hat{P}_{22}(0, t)$  και  $\hat{P}_{23}(0, t)$  έχουμε (Σχήμα 46):

Σχήμα 46: Πιθανότητες παραμονής και μετάβασης στην κατάσταση 3 για έναν ασθενή που βρίσκεται στην 2 την χρονική στιγμή  $t = 0$ .



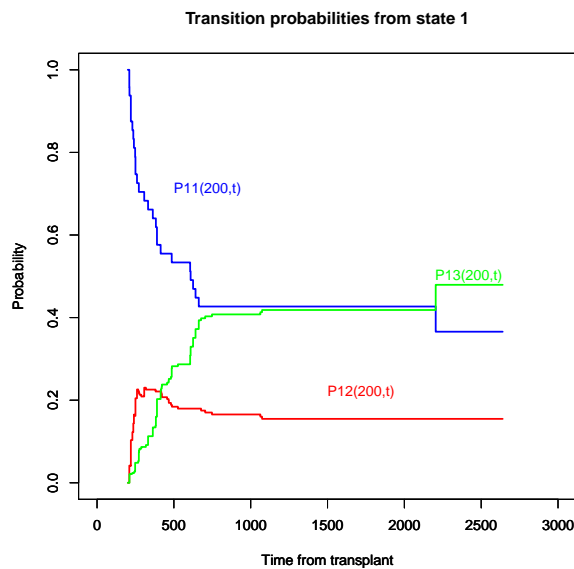
Από το Σχήμα 46, δεδομένου ότι ένας ασθενής βρίσκεται την  $s = 0$  στην μεταβατική κατάσταση 2, παρατηρούμε ότι είναι πιο πιθανό να παραμείνει εκεί μέχρι την  $t = 421$  από ότι να μεταβεί στην 3. Για  $t \geq 421$ , ωστόσο, επικρατεί η πιθανότητα αποτυχίας, με μέγιστη τιμή 0.640 που παρατηρείται από την χρονική στιγμή 1074 και έπειτα. Παρατηρούμε ότι η πιθανότητα αποτυχίας ξεπερνάει την πιθανότητα παραμονής για μεγάλη τιμή του χρόνου. Αυτό οφείλεται κυρίως στο ότι από την κατάσταση 2 παρατηρούνται συνολικά 34 δεξιά αποκομμένες παρατηρήσεις και μόλις 37 αποτυχίες, με τις περισσότερες αποτυχίες να παρατηρούνται για συγκριτικά μεγαλύτερες χρονικές στιγμές.

Για τις  $P_{1j}(200, t)$ ,  $\forall j = 1, 2, 3$

Επιπλέον, μπορούμε να κάνουμε προβλέψεις για τις πιθανότητες μετάβασης ενός ασθενή στις δυνατές καταστάσεις μετάβασης της παραπάνω ICTMC δεδομένου ότι την χρονική στιγμή  $t = 200$  βρίσκεται στην κατάσταση 1.

Για τις  $\hat{P}_{11}(200, t)$ ,  $\hat{P}_{12}(200, t)$  και  $\hat{P}_{13}(200, t)$  (Σχήμα 47):

Σχήμα 47: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή που είναι υγιής την χρονική στιγμή  $t = 200$ .



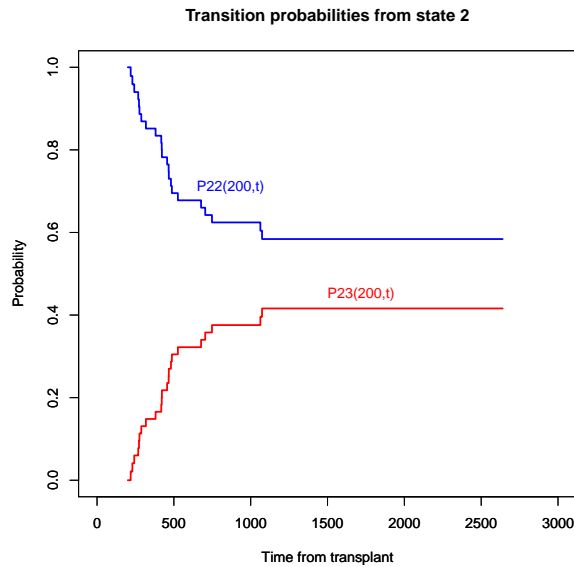
Παρατηρούμε, από το Σχήμα 47, ότι η πιθανότητα παραμονής στην κατάσταση 1 υπερτερεί των υπολοίπων μέχρι την χρονική στιγμή  $t = 2204$ , όπου η πιθανότητα μετάβασης στην απορροφητική 3, την ξεπερνάει, με μέγιστη τιμή 0.480. Επίσης, η πιθανότητα μετάβασης στην 2 κινείται σε πολύ χαμηλά επίπεδα, με μέγιστη τιμή 0.230. Προκύπτει, δηλαδή, από το παραπάνω σχήμα, ότι αν ο ασθενής δεν έχει αποτύχει μέχρι την  $t = 200$  είναι πιο πιθανό να παραμείνει υγιής μέχρι την  $t = 2204$  μετά την οποία το διάλυσμα πιθανοτήτων παραμονής και μετάβασης της αλυσίδας είναι το  $(0.366, 0.155, 0.480)$ .

**Για τις  $P_{2j}(200, t)$ ,  $\forall j = 2, 3$**

Ομοίως με παραπάνω, δεδομένου ότι την χρονική στιγμή  $t = 200$  βρίσκεται στην κατάσταση 2.

Για τις  $\hat{P}_{22}(200, t)$  και  $\hat{P}_{23}(200, t)$  (Σχήμα 48):

Σχήμα 48: Πιθανότητες παραμονής και μετάβασης στην κατάσταση 3 για έναν ασθενή που βρίσκεται στην 2 την χρονική στιγμή  $t = 200$ .



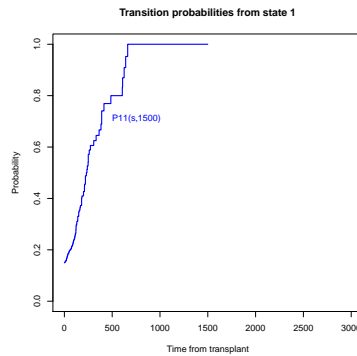
Παρατηρούμε, από το Σχήμα 48, ότι οι ασθενείς που την χρονική στιγμή  $t = 200$  βρίσκονταν στην μεταβατική κατάσταση 2, είναι συστηματικά πιο πιθανό να παραμείνουν εκεί παρά να αποτύχουν, με το πέρασμα του χρόνου. Μάλιστα, το διάστημα πιθανοτήτων μετάβασης της αλυσίδας για  $t \geq 1074$  είναι το  $(0.584, 0.420)$ .

**Για την  $P_{11}(0, 1500)$  :**

Επίσης, δεδομένου ότι την χρονική στιγμή  $t = 1500$  η μαρκοβιανή αλυσίδα βρίσκεται στην κατάσταση 1 μπορούμε να εκτιμήσουμε, κινούμενοι ανάποδα, την πιθανότητα με την οποία βρίσκονταν στην αρχική της κατάσταση τις προηγούμενες χρονικές στιγμές  $s \in [0, 1500]$ .

Για την  $P_{11}(s, 1500)$ , έχουμε: (Σχήμα 49):

Σχήμα 49: Πιθανότητα παραμονής ενός ασθενή στην αρχική κατάσταση στο χρονικό διάστημα  $[0, 1500]$ , δεδομένου ότι την χρονική στιγμή  $t = 1500$  είναι υγιής.



Από το Σχήμα 49 παρατηρούμε ότι η πιθανότητα  $\hat{P}_{11}(s, 1500)$  ξεκινάει την χρονική στιγμή  $s = 0$  με τιμή 0.150. Η αύξηση της είναι αρκετά απότομη και αποκτά την τιμή 1 για χρονικές στιγμές  $s \geq 662$ .

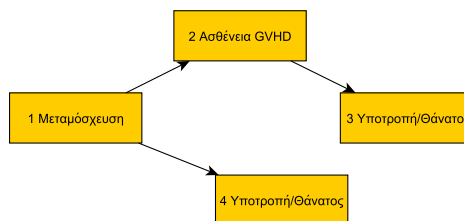
**Διάκριση μεταξύ των μεταβάσεων  $1 \Rightarrow 3$  και  $1 \Rightarrow 2 \Rightarrow 3$  :**

Ωστόσο, στην παραπάνω ανάλυση δεν γίνεται διαχωρισμός μεταξύ των απευθείας μεταβάσεων  $1 \Rightarrow 3$  και των μεταβάσεων, μέσω της μεταβατικής κατάστασης 2,  $1 \Rightarrow 2 \Rightarrow 3$ . Όπως είδαμε και στην θεωρία:

$$P_{13}(s, t) = P_{13}^1(s, t) + P_{13}^2(s, t)$$

Για να επιλύσουμε το παραπάνω πρόβλημα, θεωρούμε ένα διαφορετικό multi-state μοντέλο με δύο απορροφητικές καταστάσεις, όπου η πρώτη (κατάσταση 3) παριστάνει την αποτυχία μετά το πέρασμα της αλυσίδας από την GVHD και η δεύτερη (κατάσταση 4) παριστάνει την απευθείας μετάβαση από την αρχική κατάσταση στην Relapse/Death (Σχήμα 50).

Σχήμα 50: Εναλλακτικό Μοντέλο Multi-state στο οποίο γίνεται ο διαχωρισμός μεταξύ των μεταβάσεων  $1 \Rightarrow 3$  και  $1 \Rightarrow 2 \Rightarrow 3$ .



Ο πίνακας μετάβασης του παραπάνω μοντέλου multi-state (Πίνακας 32):

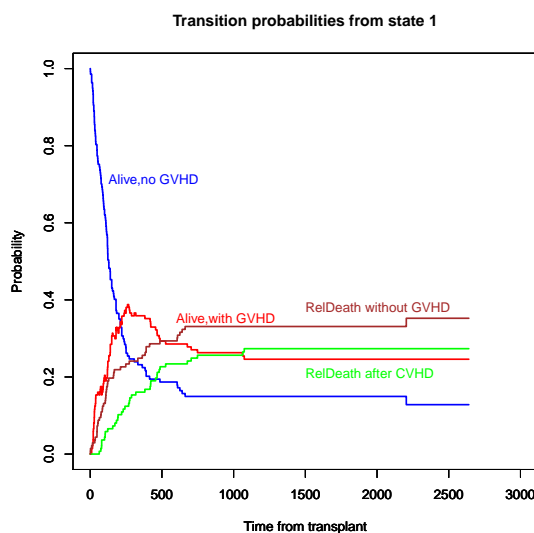
Πίνακας 32: Μεταβάσεις

from/to	State1	State2	State3	State4
State1	NA	1	NA	2
State2	NA	NA	3	NA
State3	NA	NA	NA	NA
State4	NA	NA	NA	NA

### Υπολογισμός των πιθανοτήτων παραμονής και μετάβασης από την 1 :

Για τις  $\hat{P}_{11}(0, t)$ ,  $\hat{P}_{12}(0, t)$ ,  $\hat{P}_{13}^2(0, t)$  και  $\hat{P}_{13}^1(0, t)$  (Σχήμα 51):

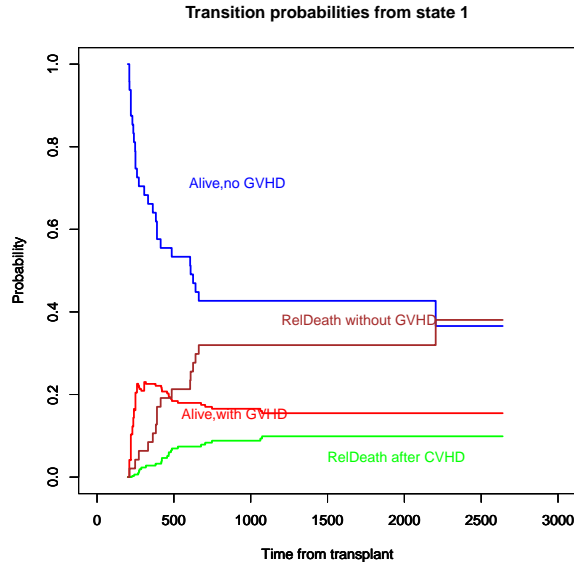
Σχήμα 51: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή που είναι υγιής την χρονική στιγμή  $t = 0$  στο νέο μοντέλο multi-state.



Από το Σχήμα 51 παρατηρούμε ότι για χρονικές στιγμές  $t \in (0, 210)$  η πιθανότητα παραμονής των ασθενών στην αρχική κατάσταση τους κατάσταση ήταν μεγαλύτερη από την πιθανότητα μετάβασης στις υπόλοιπες καταστάσεις της αλυσίδας. Για  $t \geq 210$  μεγαλύτερη είναι η πιθανότητα μετάβασης στην 2 και παραμονής εκεί. Ωστόσο, για  $t \geq 487$  η πιθανότητα μετάβασης  $1 \Rightarrow 3$ , επικρατεί των υπολοίπων, με μέγιστη τιμή 0.352 για χρονικές στιγμές 2204 και έπειτα. Επίσης, η πιθανότητα της μετάβασης  $1 \Rightarrow 2 \Rightarrow 3$  παραμένει συνεχώς μικρότερη της  $\hat{P}_{13}^1$ , γεγονός που το αναμέναμε λόγω των 37 αποτυχιών στις οποίες έγινε πρώτα μετάβαση στη GVHD, έναντι των 46 απευθείας μεταβάσεων στην Relapse/Death.

Για τις  $\hat{P}_{11}(200, t)$ ,  $\hat{P}_{12}(200, t)$ ,  $\hat{P}_{13}^2(200, t)$  και  $\hat{P}_{13}^1(200, t)$  (Σχήμα 52):

Σχήμα 52: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή που είναι υγιής την χρονική στιγμή  $t = 200$  στο νέο μοντέλο multi-state.



Με δεδομένο ότι ένας ασθενής είναι υγιής μέχρι την  $t = 200$ , παρατηρούμε από το Σχήμα 52 ότι είναι πιο πιθανό να παραμείνει εκεί για την συντριπτική πλειοψηφία των παρατηρούμενων χρόνων. Πράγματι, η πιθανότητα παραμονής στην 1 επικρατεί των υπολοίπων μέχρι την  $t = 2204$  μετά την οποία υπερτερεί οριακά η πιθανότητα απευθείας μετάβασης στην κατάσταση 3, με πιθανότητες 0.366 και 0.381, αντίστοιχα.

Οι μεταβάσεις από την κατάσταση 2 έχουν τις ίδιες ακριβώς εκτιμήτριες και στις δύο παραπάνω περιπτώσεις με αυτές που προκύπτουν θεωρώντας το αρχικό μοντέλο multi-state που περιγράψαμε παραπάνω.

### 3.5.3 Ημι-παραμετρική ανάλυση

Η μοντελοποίηση των συναρτήσεων διακινδύνευσης για τις τρεις δυνατές μεταβάσεις είναι η εξής:

$$\lambda_{12}(t|\vec{z}) = \lambda_{12,0}(t) \exp(b_{12}^1 \text{group1} + b_{12}^2 \text{group2} + b_{12}^3 \text{fab} + b_{12}^4 \text{mtx})$$

$$\lambda_{13}(t|\vec{z}) = \lambda_{13,0}(t) \exp(b_{13}^1 \text{group1} + b_{13}^2 \text{group2} + b_{13}^3 \text{fab} + b_{13}^4 \text{mtx})$$

$$\lambda_{23}(t|\vec{z}) = \lambda_{23,0}(t) \exp(b_{23}^1 \text{group1} + b_{23}^2 \text{group2} + b_{23}^3 \text{fab} + b_{23}^4 \text{mtx})$$



### Προσαρμογή μοντέλων Cox για τις $\lambda_{12}(t|\bar{z})$ , $\lambda_{13}(t|\bar{z})$ και $\lambda_{23}(t|\bar{z})$ :

Εφαρμόζοντας την μεθοδολογία των ανά αιτία (cause-specific) συναρτήσεων διακινδύνευσης, προσαρμόζουμε τρία ξεχωριστά μοντέλα παλινδρόμησης Cox για τις τρεις δυνατές μεταβάσεις του παραπάνω μοντέλου (Στρωματοποιημένη Παλινδρόμηση κατά Cox ) (Πίνακας 33):

Πίνακας 33: Προσαρμογή τριών Stratified Cox μοντέλων στις ανά μετάβαση συναρτήσεις διακινδύνευσης του προβλήματος των Klein και Moeschberger.

	coef	se(coef)	exp(coef)	z-value	p-value
group1.1	0.30562	135.746	0.33942	0.90	0.3679
group2.1	-0.38025	0.68369	0.29942	-1.27	0.2041
group1.2	0.07985	108.312	0.43202	0.18	0.8534
group2.2	-1.05936	0.34668	0.37229	-2.85	0.0044
group1.3	-0.11454	0.89178	0.49114	-0.23	0.8156
group2.3	-0.63613	0.52934	0.39886	-1.59	0.1107
fab.1	-0.00725	0.99278	0.29981	-0.02	0.9807
fab.2	0.53448	170.657	0.36086	1.48	0.1386
fab.3	0.95512	259.899	0.41267	2.31	0.0206
mtx.1	-0.18443	0.83158	0.28821	-0.64	0.5222
mtx.2	0.66257	193.976	0.30682	2.16	0.0308
mtx.3	-0.06621	0.93593	0.43394	-0.15	0.8787

Likelihood ratio test=34.9 on 12 df, p=0.000494

Στον Πίνακα 33 παρουσιάζονται οι εκτιμήσεις των συντελεστών (coef) των συμμεταβλητών των τριών στρωματοποιημένων (stratified) μοντέλων Cox. Επίσης, δίνονται τα τυπικά σφάλματα τους (se(coef)) αλλά και η επίδραση τους στην αντίστοιχη ανά μετάβαση συνάρτηση διακινδύνευσης (exp(coef)) δεδομένου ότι οι υπόλοιπες συμμεταβλητές του συγκεκριμένου μοντέλου-μετάβασης παραμένουν σταθερές. Στην συνέχεια, καταγράφεται η τιμή της ελεγχοσυνάρτησης του ελέγχου Wald και το p-value του ελέγχου  $H_0 : b = 0$  με  $H_1 : b \neq 0$  σε επίπεδο σημαντικότητας 95% για κάθε συμμεταβλητή και κάθε μετάβαση.

Τέλος, δίνεται και η τιμή του ελέγχου:

$$H_0 : b_{ij}^k = 0, \forall (i, j) = \{(1, 2), (1, 3), (2, 3)\}, k = 1, 2, 3, 4$$

$H_1$ : τουλάχιστον 1 από τους παραπάνω συντελεστές είναι διάφορος του μηδενός

Προκύπτει p-value=0.000494 και άρα δεν υπάρχει ισχυρός λόγος να απορρίψουμε σε επίπεδο σημαντικότητας του ελέγχου 95% την προσαρμογή τριών στρωματοποιημένων (stratified) Cox μοντέλων στις συναρτήσεις διακινδύνευσης του παραπάνω μοντέλου illness-death.

**Έλεγχος υπόθεσης ανάλογων βασικών συναρτήσεων των  $\lambda_{13}(t|\bar{z})$  και  $\lambda_{23}(t|\bar{z})$  :**

Στην συνέχεια, είμαστε σε θέση να διαπιστώσουμε εαν μπορούμε να θεωρήσουμε ανάλογες βασικές (baseline) συναρτήσεις διακινδύνευσης για τις μεταβάσεις  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$ , όπως είδαμε στην θεωρία, ή όχι. Θεωρούμε την χρονικά εξαρτημένη μεταβλητή *gvhd* η οποία παίρνει την τιμή 0 για τους ασθενείς που έκαναν την μετάβαση  $1 \Rightarrow 3$  και την τιμή 1 για αυτούς που έκαναν την  $2 \Rightarrow 3$ . Προσαρμόζουμε και πάλι τρία μοντέλα παλινδρόμησης Cox, όπως παραπάνω, λαμβάνοντας αυτή την φορά την *gvhd* ως μια ακόμα συμμεταβλητή, και έχουμε (Πίνακας 34):

Πίνακας 34: Προσαρμογή τριών μοντέλων Cox στις συναρτήσεις διακινδύνευσης του μοντέλου των Klein και Moeschberger θεωρώντας ότι τα  $\lambda_{13}(t|\bar{z})$  και  $\lambda_{23}(t|\bar{z})$  έχουν ανάλογες βασικές συναρτήσεις.

	coef	exp(coef)	se(coef)	z-value	p-value
group1.1	0.30562	135.746	0.33942	0.90	0.3679
group2.1	-0.38025	0.68369	0.29942	-1.27	0.2041
group1.2	0.04958	105.083	0.42979	0.12	0.9082
group2.2	-101.524	0.36232	0.36831	-2.76	0.0058
group1.3	-0.02920	0.97122	0.48643	-0.06	0.9521
group2.3	-0.62858	0.53335	0.39796	-1.58	0.1142
fab.1	-0.00725	0.99278	0.29981	-0.02	0.9807
fab.2	0.55133	173.556	0.35893	1.54	0.1245
fab.3	102.371	278.350	0.40952	2.50	0.0124
mtx.1	-0.18443	0.83158	0.28821	-0.64	0.5222
mtx.2	0.65291	192.111	0.30558	2.14	0.0326
mtx.3	-0.09368	0.91058	0.43150	-0.22	0.8281
gvhd	0.03462	103.522	0.54010	0.06	0.9489

Likelihood ratio test=34.8 on 13 df, p=0.00091

Αρχικά, παρατηρούμε από τον Πίνακα 34 ότι το p-value για τον συνολικό έλεγχο του μοντέλου (αν όλοι οι συντελεστές των μοντέλων μας είναι μηδέν) είναι 0.00091, αρκετά μικρό δηλαδή, γεγονός που σημαίνει ότι δεν μπορούμε να απορρίψουμε την προσαρμογή του. Επίσης, η μεταβλητή *gvhd* έχει πολύ μεγάλο p-value (0.949) γεγονός που αποτελεί ένδειξη υπέρ της μηδενικής μας υπόθεσης των αναλογικών βασικών συναρτήσεων διακινδύνευσης για τις δύο παραπάνω μεταβάσεις.

Παρακάτω παραθέτουμε και τα αποτελέσματα του ελέγχου Schoenfeld για την υπόθεση αναλογικότητας των βασικών συναρτήσεων για κάθε συμμεταβλητή (Πίνακας 35):

Πίνακας 35: Υπόλοιπα Schoenfeld της προσαρμογής των τριών μοντέλων Cox στις συναρτήσεις διακινδύνευσης, με την υπόθεση ότι τα  $\lambda_{13}(t|\vec{z})$  και  $\lambda_{23}(t|\vec{z})$  έχουν ανάλογες βασικές συναρτήσεις.

	rho	chisq	p-value
group1.1	-0.08164	9.98e-01	0.3177
group2.1	-0.13690	2.96e+00	0.0855
group1.2	0.11346	1.81e+00	0.1789
group2.2	0.04943	3.46e-01	0.5562
group1.3	0.12717	2.45e+00	0.1179
group2.3	0.07570	8.57e-01	0.3546
fab.1	-0.06407	6.08e-01	0.4355
fab.2	0.00128	2.39e-04	0.9877
fab.3	0.12104	2.22e+00	0.1360
mtx.1	-0.00391	2.42e-03	0.9608
mtx.2	-0.16112	3.70e+00	0.0544
mtx.3	-0.13282	2.91e+00	0.0878
gvhd	-0.12415	2.23e+00	0.1357
GLOBAL	NA	1.67e+01	0.2111

Τα p-value των ελέγχων του Πίνακα 35 για την συντριπτική πλειοψηφία των συντελεστών των συμμεταβλητών είναι αρκετά μεγάλα. Η υπόθεση, δηλαδή, ότι παραμένουν σταθεροί ως προς τον χρόνο, φαίνεται να ευσταθεί. Επομένως, θα μπορούσαμε, δυνητικά, να προσαρμόσουμε τις συναρτήσεις διακινδύνευσης των μεταβάσεων  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$  ώστε να έχουν ανάλογες βασικές συναρτήσεις.

### Ημι-παραμετρική ανάλυση θεωρώντας semi-Markov μοντέλο illness-death:

Στην συνέχεια, θα μας ενδιέφερε να μελετήσουμε την προσαρμογή των τριών παραπάνω μοντέλων παλινδρόμησης, αλλά, θεωρώντας αυτή τη φορά μια semi-Markov (state arrival extended Markov) διαδικασία για να περιγράψουμε το illness-death μοντέλο μας. Όπως είχαμε αναφέρει και στην θεωρία, semi-Markov είναι τα μοντέλα στα οποία η πιθανότητα μετάβασης  $i \Rightarrow j$ ,  $P_{ij}(s, t)$ , εξαρτάται μόνο από τον χρόνο που πήρε στην αλυσίδα να φτάσει στην κατάσταση  $i$ . Εν προκειμένω, μας ενδιαφέρει να επαναλάβουμε την προσαρμογή των στρωματοποιημένων μοντέλων Cox ώστε να διαπιστώσουμε εάν ο χρόνος που έκανε ένας ασθενής να φτάσει στην κατάσταση 2 (illness/gvhd) επηρεάζει την συνάρτηση διακινδύνευσης  $\lambda_{23}(t|\vec{z})$  ή όχι.

Προσαρμόζουμε λοιπόν τα μοντέλα μας όπως πριν, με την προσθήκη ωστόσο, μιας ακόμα συμμεταβλητής (Πίνακας 36).

Πίνακας 36: Προσαρμογή τριών μοντέλων Cox στις συναρτήσεις διακινδύνευση του μοντέλου των Klein και Moeschberger με την υπόθεση ότι η αλυσίδα μας είναι semi-Markov.

	coef	exp(coef)	se(coef)	z-value	p-value
group1.1	0.30562	135.746	0.33942	0.90	0.3679
group2.1	-0.38025	0.68369	0.29942	-1.27	0.2041
group1.2	0.05484	105.638	0.42993	0.13	0.8985
group2.2	-105.846	0.34699	0.36878	-2.87	0.0041
group1.3	-0.30564	0.73665	0.49574	-0.62	0.5375
group2.3	-0.87209	0.41808	0.41547	-2.10	0.0358
fab.1	-0.00725	0.99278	0.29981	-0.02	0.9807
fab.2	0.55149	173.585	0.35910	1.54	0.1246
fab.3	0.87023	238.746	0.41134	2.12	0.0344
mtx.1	-0.18443	0.83158	0.28821	-0.64	0.5222
mtx.2	0.64899	191.361	0.30566	2.12	0.0337
mtx.3	-0.11437	0.89193	0.43915	-0.26	0.7945
gvhd	0.75148	212.014	0.61244	1.23	0.2198
prtime.3	-0.00560	0.99442	0.00276	-2.03	0.0428

Likelihood ratio test=39.2 on 14 df, p=0.000343

Η συμμεταβλητή prtime.3, όπως φαίνεται στον Πίνακα 36, εκφράζει τον χρόνο που πήρε στους ασθενείς που μετέβησαν από την κατάσταση 2 στην 3, να φτάσουν στην 2. Από το p-value (0.0428) της συμμεταβλητής αυτής, δεν μπορούμε να απορρίψουμε σε επίπεδο σημαντικότητας του ελέγχου 95% την υπόθεση ότι είναι στατιστικά σημαντική. Ισοδύναμα, σε επίπεδο σημαντικότητας του ελέγχου 95% αποδεχόμαστε την υπόθεση ότι ένα semi-Markov μοντέλο έχει καλή προσαρμογή στα δεδομένα μας.

Αν και η προσαρμογή ενός τέτοιου μοντέλου είναι εφικτή στα δεδομένα του μοντέλου μας, δεν θα το αναπτύξουμε περαιτέρω καθώς η ισχυρή μαρκοβιανή ιδιότητα της παραπάνω αλυσίδας είναι κομβικής σημασίας για τους υπολογισμούς που θα κάνουμε στην συνέχεια.

### 3.5.4 Εκτιμήτριες Breslow

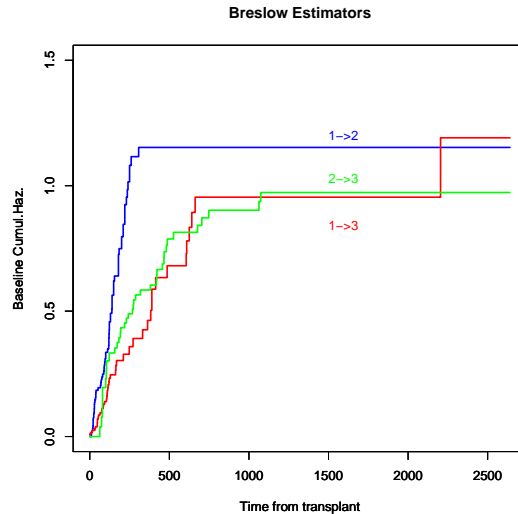
Με βάση και τα όσα αναφέραμε στις προηγούμενες παραγράφους, θα αναφερθούμε στις εκτιμήσεις των πιθανοτήτων μετάβασης Aalen-Johansen για την παραπάνω illness-death, Μη-ομοιογενή Συνεχούς Χρόνου Μαρκοβιανή Αλυσίδα (ICTMC).

#### Εκτιμήτριες Breslow για τη stratified περίπτωση:

Αρχικά, μπορούμε να υπολογίσουμε με την βοήθεια της R (De Wreede et al. , 2011) τις εκτιμήσεις των συναρτήσεων  $\Lambda_{12,0}(t)$ ,  $\Lambda_{13,0}(t)$  και  $\Lambda_{23,0}(t)$ , τις εκτιμήτριες Breslow, δηλαδή, των βασικών συναρτήσεων σωρευτικής διακινδύνευσης στα 3 στρωματοποιημένα μοντέλα Cox που έχουμε προσαρμόσει. Συμβολίζουμε με 1 την μετάβαση  $1 \Rightarrow 2$ , με 2 την  $1 \Rightarrow 3$  και με 3 την  $2 \Rightarrow 3$  (Σχήμα 43). Σχεδιάζουμε τις παραπάνω συναρτήσεις (Σχήμα 53):

$$\hat{\Lambda}_{12,0}(t), \hat{\Lambda}_{13,0}(t) \text{ και } \hat{\Lambda}_{23,0}(t)$$

Σχήμα 53: Εκτιμήσεις Breslow για τις τρεις βασικές συναρτήσεις σωρευτικής διακινδύνευσης των ανά μετάβαση στρωματοποιημένων μοντέλων Cox που προσαρμόσαμε.



Παρατηρούμε από το Σχήμα 53 ότι μέχρι την χρονική στιγμή  $s = 2204$  τις υψηλότερες τιμές παίρνει συστηματικά η  $\hat{\Lambda}_{12,0}(t)$ , ενώ η  $\hat{\Lambda}_{13,0}(t)$  την ξεπερνάει με τιμή 1.191 για χρονικές στιγμές  $t > 2204$ .

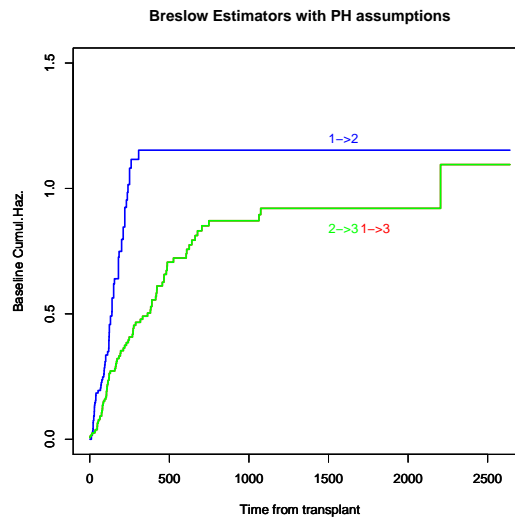
Όπως είδαμε και στην θεωρία, οι παραπάνω συναρτήσεις παίζουν κομβικό ρόλο στον προσδιορισμό των εκτιμήσεων των πιθανοτήτων μετάβασης  $\hat{P}_{ij}(s, t)$  κατά Aalen-Johansen.

### Εκτιμήσεις Breslow με την υπόθεση αναλογίας των $\lambda_{13,0}(t|\bar{z})$ και $\lambda_{23,0}(t|\bar{z})$ :

Επαναλαμβάνουμε την ίδια ανάλυση με παραπάνω, θεωρώντας όμως αυτή την φορά ότι τα μοντέλα Cox που προσαρμόζουμε συναρτήσεις διακινδύνευσης για τις μεταβάσεις  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$  έχουν ανάλογες βασικές συναρτήσεις (Σχήμα 54): Έχουμε:

$$\hat{\Lambda}_{12,0}(t), \hat{\Lambda}_{13,0}(t) \text{ και } \hat{\Lambda}_{23,0}(t)$$

Σχήμα 54: Εκτιμήτριες Breslow για τις τρεις βασικές συναρτήσεις σωρευτικής διακινδύνευσης των ανά μετάβαση Cox μοντέλων που προσαρμόσαμε, με την υπόθεση ότι οι  $\lambda_{13}(t|\bar{z})$  και  $\lambda_{23}(t|\bar{z})$  έχουν ανάλογες βασικές συναρτήσεις.



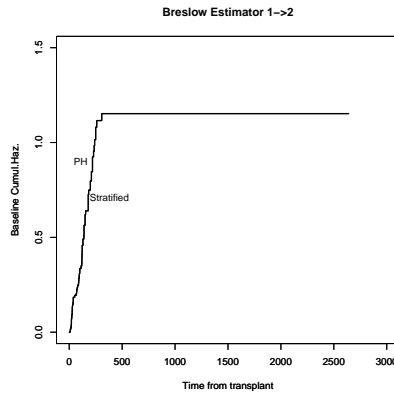
Παρατηρούμε, από το Σχήμα 54, ότι οι εκτιμήτριες  $\hat{\Lambda}_{13,0}(t)$  και  $\hat{\Lambda}_{23,0}(t)$  ταυτίζονται, προκύπτει δηλαδή σταθερά αναλογίας των βασικών συναρτήσεων ίση με 1.

### Σύγκριση:

Συγκρίνουμε τις παραπάνω εκτιμήσεις κατά Breslow για κάθε μετάβαση (1,2 και 3) όπως προκύπτουν από την στρωματοποιημένη και αναλογικής διακινδύνευσης, αντίστοιχα, παλινδρόμηση κατά Cox που περιγράψαμε παραπάνω. (Σχήματα 55, 56 και 57):

$$\hat{\Lambda}_{12,0}(t) :$$

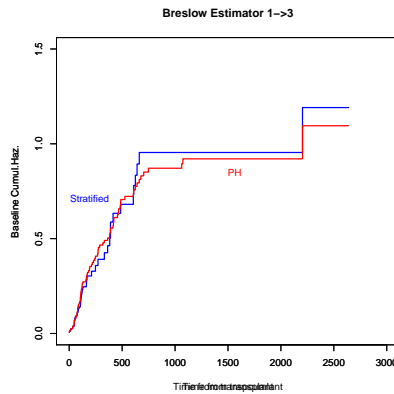
Σχήμα 55: Σύγκριση της εκτίμησης Breslow της  $\Lambda_{12,0}(t)$  όπως προέκυψε προσαρμόζοντας τα στρωματοποιημένα μοντέλα Cox και όπως προέκυψε κάνοντας την υπόθεση αναλογικότητας διακινδύνευσης μεταξύ των μεταβάσεων 2 και 3.



Από το Σχήμα 55 παρατηρούμε ότι, όπως αναμέναμε, και στις 2 περιπτώσεις, οι συναρτήσεις  $\hat{\Lambda}_{12,0}(t)$  ταυτίζονται. Αυτό συμβαίνει διότι κατά την προσαρμογή των μοντέλων Cox με αναλογικές βασικές συναρτήσεις διακινδύνευσης και άρα σωρευτικής διακινδύνευσης, για τις μεταβάσεις 2 και 3, αντίστοιχα, η προσαρμογή της συνάρτησης διακινδύνευσης και σωρευτικής διακινδύνευσης είναι η ίδια για την μετάβαση  $1 \Rightarrow 2$  και στις δύο περιπτώσεις.

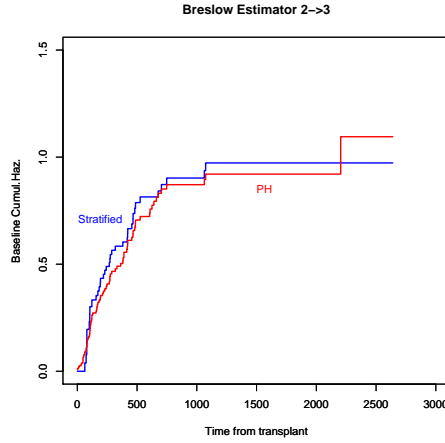
$$\hat{\Lambda}_{13,0}(t) :$$

Σχήμα 56: Σύγκριση της εκτίμησης Breslow της  $\Lambda_{13,0}(t)$  όπως προέκυψε προσαρμόζοντας τα στρωματοποιημένα μοντέλα Cox και όπως προέκυψε κάνοντας την υπόθεση αναλογικότητας διακινδύνευσης μεταξύ των μεταβάσεων 2 και 3.



$$\hat{\Lambda}_{23,0}(t) :$$

Σχήμα 57: Σύγκριση της εκτίμησης Breslow της  $\Lambda_{23,0}(t)$  όπως προέκυψε προσαρμόζοντας τα στρωματοποιημένα μοντέλα Cox και όπως προέκυψε κάνοντας την υπόθεση αναλογικότητας διακινδύνευσης μεταξύ των μεταβάσεων 2 και 3.



Οι συναρτήσεις των Σχημάτων 56 και 57 δεν συμπεριφέρονται με τον ίδιο τρόπο για τα 2 διαφορετικά μοντέλα παλινδρόμησης (στρωματοποιημένα και αναλογικής διακινδύνευσης) που προσαρμόσαμε παραπάνω. Για τις  $\hat{\Lambda}_{13,0}(t)$ , η εκτιμήτρια των στρωματοποιημένων (stratified) μοντέλων παίρνει αρχικά μικρότερες τιμές από τις εκτιμήτριες που προκύπτουν θεωρώντας ανάλογες τις βασικές συναρτήσεις διακινδύνευσης των  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$  ωστόσο για χρονικές στιγμές  $t \geq 609$ , και έπειτα, η κατάσταση αντιστρέφεται. Αντίθετα, στις  $\hat{\Lambda}_{23,0}(t)$  η εκτιμήτρια για τα στρωματοποιημένα μοντέλα είναι συστηματικά μεγαλύτερη αυτής που προκύπτει κάνοντας την προαναφερθείσα υπόθεση αναλογικότητας μέχρι την χρονική στιγμή  $t = 2204$  όπου το μοτίβο αντιστρέφεται και πάλι.

### 3.5.5 Εκτιμήτριες Aalen-Johansen

Από την θεωρία γνωρίζουμε για τις ICTMC ότι:

$$\hat{P}(X(t) = j | X(s) = i) = \hat{P}_{ij}(s, t) = \prod_{k=1}^K (I + (\hat{\Lambda}_{ij}(t_k | \vec{z}) - \hat{\Lambda}_{ij}(t_{k-1} | \vec{z})))$$

όπου  $s = t_0 < t_1 < \dots < t_K = t$  είναι μια διαμέριση του  $[s, t]$  και:

$$\hat{\Lambda}_{ij}(t | \vec{z}) = \hat{\Lambda}_{ij,0}(t) \exp(\hat{b}_{ij}^T \vec{z})$$

η εκτιμήτρια της συνάρτησης σωρευτικής διακινδύνευσης χρησιμοποιώντας ημι-παραμετρικά μοντέλα παλινδρόμησης κατά Cox.



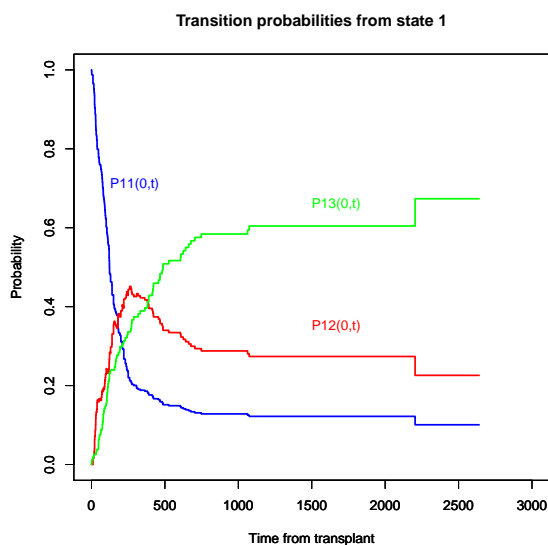
Επομένως, μέσω της R, μπορούμε να προσδιορίσουμε τις εκτιμήτριες Aalen-Johansen για τις παραπάνω πιθανότητες.

### Υπολογισμός των $P_{1j}(0, t)$ , $\forall j = 1, 2, 3$ :

Αρχικά, θα μελετήσουμε έναν ασθενή με διάνυσμα συμμεταβλητών  $(0, 0, 0, 0)$ , θεωρώντας ανάλογες βασικές συναρτήσεις διακινδύνευσης για τις μεταβάσεις  $1 \Rightarrow 3$  και  $2 \Rightarrow 3$ .

Έχουμε, λοιπόν, για τις  $\hat{P}_{11}(0, t)$ ,  $\hat{P}_{12}(0, t)$  και  $\hat{P}_{13}(0, t)$  (Σχήμα 58):

Σχήμα 58: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάνυσμα συμμεταβλητών  $(0, 0, 0, 0)$  είναι υγιής την χρονική στιγμή  $t = 0$  με την υπόθεση αναλογικής διακινδύνευσης μεταξύ των μεταβάσεων 2 και 3.

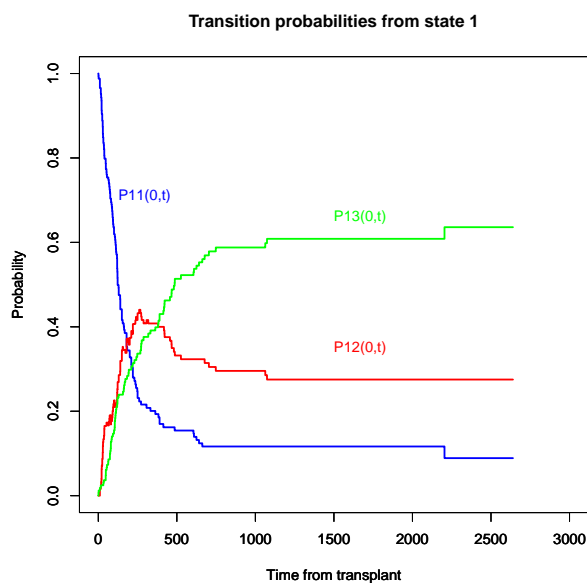


Δεδομένου, λοιπόν, ότι την χρονική στιγμή  $t = 0$  ένας ασθενής με τις παραπάνω συμμεταβλητές είναι υγιής, όσο περνάει ο χρόνος μειώνεται, με σχεδόν εκθετική μορφή, η πιθανότητα επιβίωσης του και αυξάνονται οι πιθανότητες να ασθενήσει (GVHD) και να αποτύχει (Relapse/Death). Την χρονική στιγμή  $t = 383$  και έπειτα η πιθανότητα αποτυχίας του ασθενή γίνεται μεγαλύτερη των άλλων και παίρνει μέγιστη τιμή 0.6734 για  $t = 2204$  και έπειτα (Σχήμα 58).

Ομοίως, αλλά αυτή τη φορά θα προσαρμόσουμε τρία στρωματοποιημένα (stratified) Cox μοντέλα στις ανά μετάβαση συναρτήσεις διακινδύνευσης του illness-death μοντέλου μας.

Για τις  $\hat{P}_{11}(0, t)$ ,  $\hat{P}_{12}(0, t)$  και  $\hat{P}_{13}(0, t)$  (Σχήμα 58):

Σχήμα 59: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάνυσμα συμμεταβλητών  $(0, 0, 0, 0)$  είναι υγιής την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.

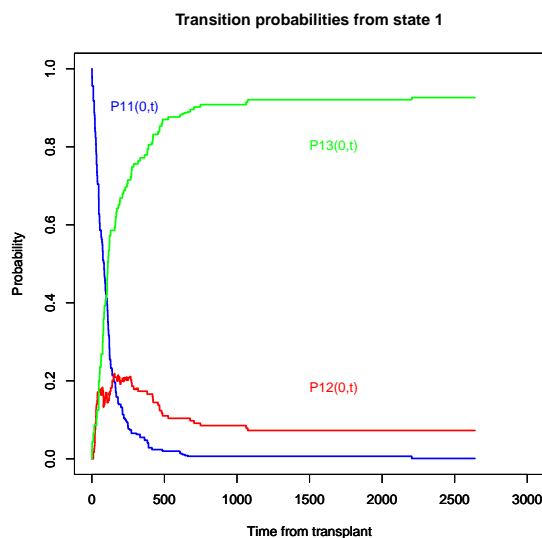


Το γράφημα του Σχήματος 59 προσεγγίζει πάρα πολύ αυτό που προέκυψε θεωρώντας την υπόθεση αναλογικότητας που περιγράψαμε. Οι μόνες ποσοτικές διαφορές είναι ότι αυτή τη φορά η πιθανότητα αποτυχίας γίνεται μεγαλύτερη των υπολοίπων για  $t \geq 381$  και παίρνει μέγιστη τιμή 0.636 για  $t = 2204$  και έπειτα.

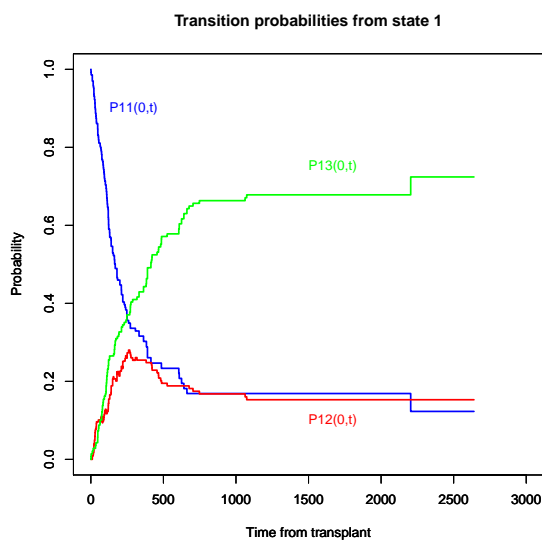
Στην συνέχεια, μπορούμε να εκτιμήσουμε τις προαναφερθείσες πιθανότητες μετάβασης στις καταστάσεις της αλυσίδας στην περίπτωση των, πιο γενικών, στρωματοποιημένων μοντέλων παλινδρόμησης Cox, για ασθενείς με διανύσματα συμμεταβλητών  $(1, 0, 1, 1)$ ,  $(0, 1, 1, 1)$ ,  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$ , αντίστοιχα.

Για τις  $\hat{P}_{11}(0, t)$ ,  $\hat{P}_{12}(0, t)$  και  $\hat{P}_{13}(0, t)$  (Σχήματα 60, 61, 62 και 63):

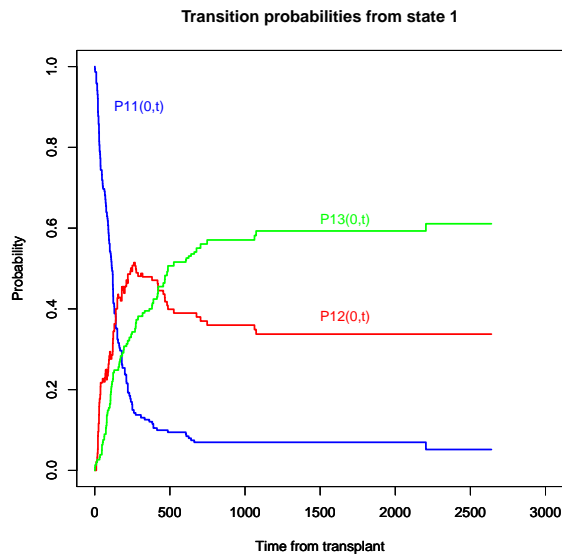
Σχήμα 60: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών (1, 0, 1, 1) είναι υγιής την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



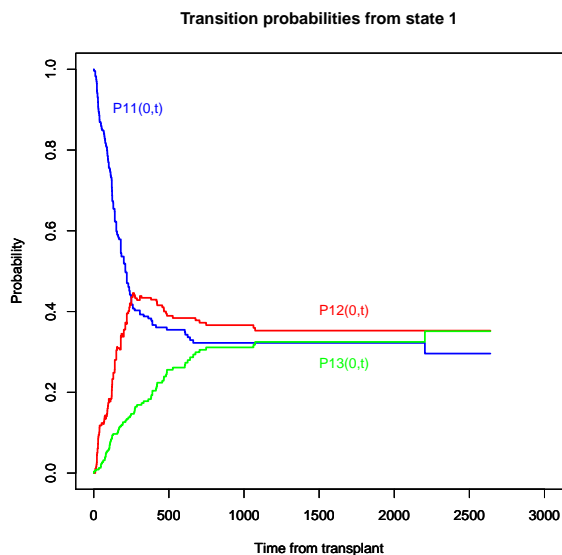
Σχήμα 61: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών (0, 1, 1, 1) είναι υγιής την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Σχήμα 62: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(1, 0, 0, 0)$  είναι υγιής την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Σχήμα 63: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(0, 1, 0, 0)$  είναι υγιής την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Από τα παραπάνω γραφήματα των Σχημάτων 60 – 63, μπορούμε να κάνουμε δύο βασικές παρατηρήσεις:

**1.** Για έναν ασθενή με συμμεταβλητές  $(1, 0, 1, 1)$  η πτώση της πιθανότητας παραμονής του στην κατάσταση 1 σε σχέση με τον χρόνο γίνεται ταχύτερα και πιο απότομα από τους υπόλοιπους συνδυασμούς συμμεταβλητών που αναφέραμε παραπάνω. Αντίστοιχα, η πιθανότητα μετάβασης των ασθενών αυτών στην απορροφητική κατάσταση 3 αυξάνεται πιο γρήγορα από αυτή των υπόλοιπων διανυσμάτων συμμεταβλητών που μελετούμε. Συγκεκριμένα, ξεπερνάει την πιθανότητα παραμονής στην 1 την  $t = 100$  με μέγιστη τιμή 0.926 για  $t \geq 2204$ . Οι ασθενείς με τις συμμεταβλητές αυτές, δηλαδή, βρίσκονται στον μεγαλύτερο κίνδυνο αποτυχίας (Relapse/Death) σε σχέση με τους ασθενείς των άλλων τριών διανυσμάτων συμμεταβλητών.

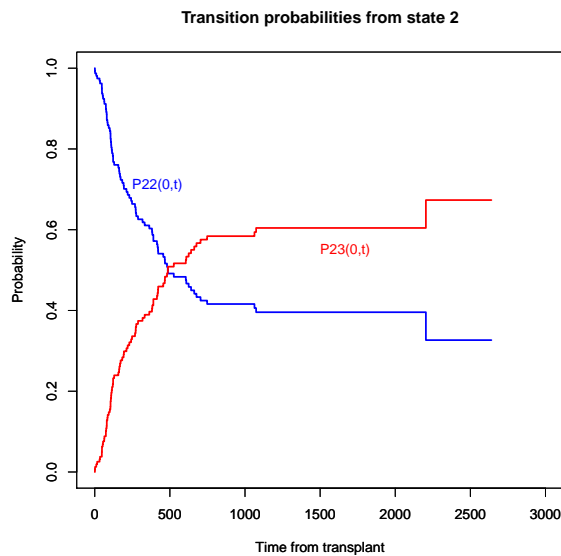
**2.** Για τους ασθενείς με διάνυσμα συμμεταβλητών  $(0, 1, 0, 0)$  παρατηρείται η πιο αργή μείωση πιθανότητας παραμονής στην αρχική κατάσταση σε σχέση με τον χρόνο. Επίσης, αντίστοιχα, βλέπουμε και την πιο αργή αύξηση της πιθανότητας αποτυχίας, η οποία αν και ξεπερνάει την πιθανότητα παραμονής την  $t = 1074$  παραμένει συνεχώς μικρότερη της πιθανότητας μετάβασης του ασθενή στην παροδική κατάσταση 2 (ασθένεια /GVHD). Μάλιστα, για  $t \geq 2640$  οι πιθανότητες μετάβασης στις καταστάσεις 1, 2 και 3 σταθεροποιούνται στο αντίστοιχο διάνυσμα  $(0.296, 0.353, 0.351)$ . Στην περίπτωση αυτή, λοιπόν, οι παραπάνω ασθενείς έχουν την μικρότερη πιθανότητα αποτυχίας ή οποία μάλιστα είναι και συστηματικά μικρότερη σε σχέση με τον χρόνο της πιθανότητας ασθένειας.

### Υπολογισμός των $P_{2j}(0, t)$ , $\forall j = 2, 3$ :

Αρχικά, θα υπολογίσουμε τις πιθανότητες παραμονής και μετάβασης των ασθενών με συμμεταβλητές  $(0, 0, 0, 0)$  από την κατάσταση 2, θεωρώντας ότι οι συναρτήσεις  $\lambda_{13}(t|\vec{z})$  και  $\lambda_{23}(t|\vec{z})$  έχουν ανάλογες βασικές (baseline) συναρτήσεις.

Για τις  $\hat{P}_{22}(0, t)$  και  $\hat{P}_{23}(0, t)$  (Σχήμα 64):

Σχήμα 64: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με συμμεταβλητές  $(0, 0, 0, 0)$  βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 0$  με την υπόθεση αναλογικότητας διακινδύνευσης μεταξύ των μεταβάσεων 2 και 3.

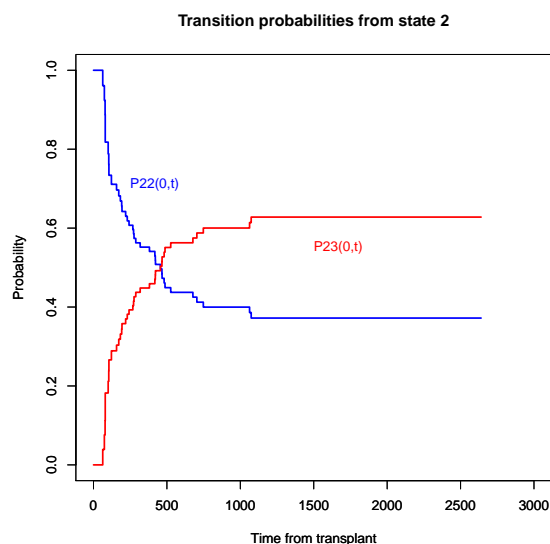


Για έναν ασθενή με το παραπάνω διάνυσμα συμμεταβλητών ο οποίος βρίσκεται στην κατάσταση 2 την  $t = 0$ , παρατηρούμε, από το Σχήμα 64, ότι μέχρι την χρονική στιγμή  $s = 486$  η πιθανότητα να παραμείνει στην κατάσταση που βρίσκεται είναι μεγαλύτερη από την πιθανότητα αποτυχίας. Ωστόσο, για  $t \geq 486$  η πιθανότητα να βρεθεί ο ασθενής στην κατάσταση 3 γίνεται μεγαλύτερη και παίρνει μέγιστη τιμή 0.673 για χρονικές στιγμές  $t \geq 2204$ .

Ομοίως με το Σχήμα 64 αλλά τώρα θα θεωρήσουμε ότι έχουμε προσαρμόσει για τους ασθενείς με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  τρία στρωματοποιημένα μοντέλα Cox στις ανά μετάβαση συναρτήσεις διακινδύνευσης του προβλήματός μας.

Για τις  $\hat{P}_{22}(0, t)$  και  $\hat{P}_{23}(0, t)$  (Σχήμα 64):

Σχήμα 65: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.

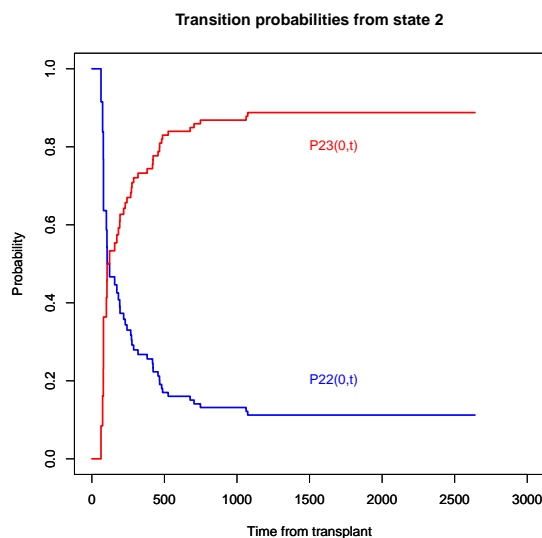


Οι πιθανότητες παραμονής και μετάβασης στην κατάσταση 3 από την κατάσταση 2, στην οποία βρίσκεται ένας ασθενής με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  την  $t = 0$ , συναρτήσεως του χρόνου, είναι σχεδόν ίδιες είτε τα μοντέλα είναι στρωματοποιημένα είτε έχουν ανάλογες βασικές συναρτήσεις διακινδύνευσης. Η μόνη στοιχειώδης διαφορά είναι ότι για  $t \geq 456$ , και όχι 486, η πιθανότητα αποτυχίας γίνεται μεγαλύτερη της πιθανότητας παραμονής στην κατάσταση 2 με μέγιστη τιμή 0.628 για  $t \geq 1074$  (Σχήμα 64).

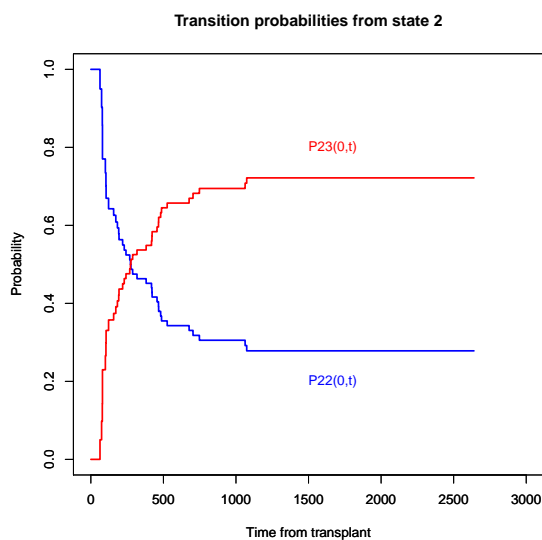
Στην συνέχεια, κάνοντας χρήση των εκτιμητριών Aalen-Johansen μπορούμε να βρούμε να επαναλάβουμε την ανάλυση των Σχημάτων 64 και 65 για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 0, 1, 1)$ ,  $(0, 1, 1, 1)$ ,  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$ , αντίστοιχα.

Για τις  $\hat{P}_{22}(0, t)$  και  $\hat{P}_{23}(0, t)$  (Σχήματα 66, 67, 68 και 69):

Σχήμα 66: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών (1, 0, 1, 1) βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.

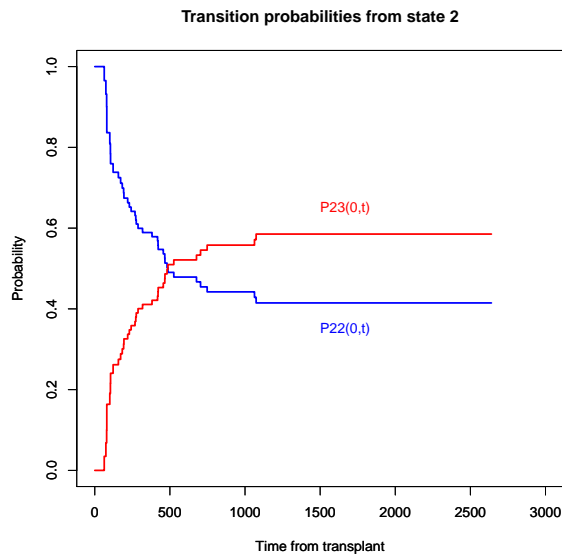


Σχήμα 67: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών (0, 1, 1, 1) βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.

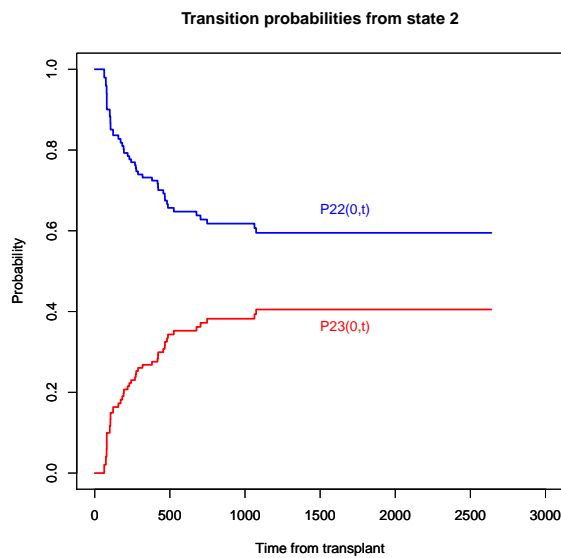




Σχήμα 68: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(1, 0, 0, 0)$  βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Σχήμα 69: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(0, 1, 0, 0)$  βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 0$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Από τις παραπάνω γραφικές παραστάσεις, των Σχημάτων 66 – 69, μπορούμε να κάνουμε τις εξής παρατηρήσεις:

**1.** Οι ασθενείς με διάνυσμα συμμεταβλητών  $(1, 0, 1, 1)$  έχουν την πιο ταχεία αύξηση της πιθανότητας αποτυχίας σε σχέση με τον χρόνο, η οποία, όπως είναι λογικό, συνοδεύεται και από την ταχύτερη μείωση της πιθανότητας παραμονής στην κατάσταση 2. Μάλιστα, η πιθανότητα μετάβασης στην κατάσταση 3 γίνεται μεγαλύτερη της πιθανότητας παραμονής στην 2 για  $t \geq 122$  με μέγιστη τιμή 0.888 για  $t = 1074$  και έπειτα.

**2.** Οι ασθενείς με συμμεταβλητές  $(0, 1, 0, 0)$ , αντίθετα, έχουν την πιο αργή αύξηση της πιθανότητας αποτυχίας και, αντίστοιχα, την πιο αργή μείωση της πιθανότητας παραμονής. Μάλιστα, από τα υπόλοιπα είδη ασθενών τις πιθανότητες μετάβασης από την κατάσταση 2, των οποίων έχουμε δει παραπάνω, οι ασθενείς αυτοί είναι οι μόνοι που έχουν συστηματικά μεγαλύτερη πιθανότητα παραμονής στην 2 από μετάβαση στην 3. Πράγματι, με πιθανότητα 0.595 για  $t \geq 1074$  οι παραπάνω ασθενείς έχουν την μεγαλύτερη πιθανότητα επιβίωσης σε σχέση με τους υπόλοιπους.

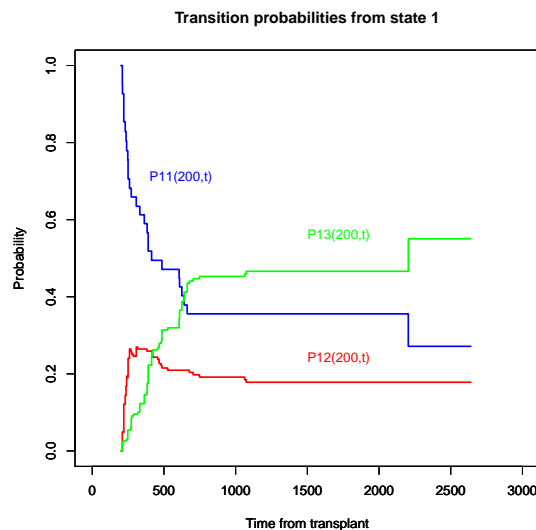
### Υπολογισμός των $P_{1j}(200, t)$ , $\forall j = 1, 2, 3$ :

Επίσης, για τους ασθενείς με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  μπορούμε να υπολογίσουμε τις ίδιες πιθανότητες με παραπάνω αλλά ξεκινώντας τις προβλέψεις μας από μια μεταγενέστερη χρονική στιγμή. Έστω δηλαδή, ότι ένας ασθενής με το παραπάνω διάγνωση συμμεταβλητών, την χρονική στιγμή  $t = 200$  βρίσκεται υγιής στην κατάσταση 1.

Θα μελετήσουμε τις πιθανότητες παραμονής και μετάβασης του στις τρεις δυνατές καταστάσεις για χρόνους  $t \geq 200$ .

Για τις  $\hat{P}_{11}(200, t)$ ,  $\hat{P}_{12}(200, t)$  και  $\hat{P}_{13}(200, t)$  (Σχήμα 70):

Σχήμα 70: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  είναι υγιής την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.

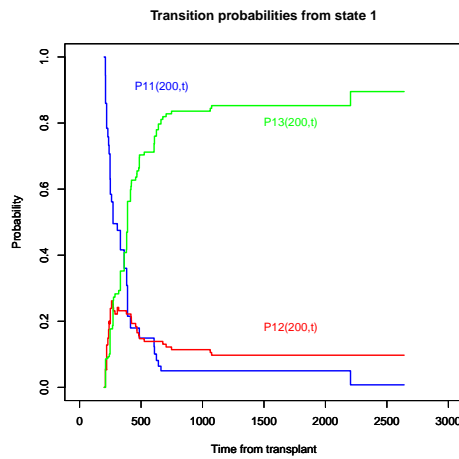


Από το Σχήμα 70 παρατηρούμε ότι μέχρι την χρονική στιγμή  $t = 420$  η πιθανότητα παραμονής του ασθενή στην αρχική του κατάσταση υπερτερεί των υπολοίπων. Ωστόσο, για  $t \geq 421$  η πιθανότητα αποτυχίας του ασθενή ξεπερνάει τις υπόλοιπες και λαμβάνει μέγιστη τιμή 0.550 για  $t = 2204$  και έπειτα. Τέλος, η πιθανότητα παραμονής στην κατάσταση 1 είναι συστηματικά μεγαλύτερη από την πιθανότητα ασθένειας (κατάσταση 2-GVHD) των παραπάνω ασθενών.

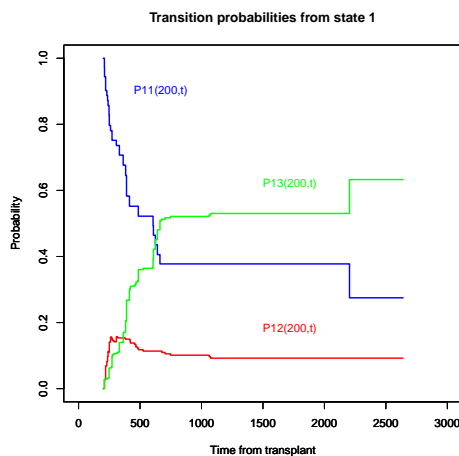
Με τον ίδιο τρόπο, για τους ασθενείς με διανύσματα συµµεταβλητών  $(1, 0, 1, 1)$ ,  $(0, 1, 1, 1)$ ,  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$ , µπορούµε να υπολογίσουµε τα εξής:

Για τις  $\hat{P}_{11}(200, t)$ ,  $\hat{P}_{12}(200, t)$  και  $\hat{P}_{13}(200, t)$  (Σχήµατα 71, 72, 73 και 74):

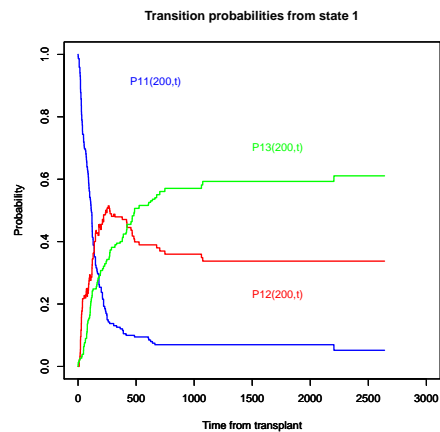
Σχήµα 71: Εκτιµήσεις των πιθανοτήτων παραµονής και µετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδοµένου ότι ένας ασθενής µε διάνυσµα συµµεταβλητών  $(1, 0, 1, 1)$  είναι υγιής την χρονική στιγµή  $t = 200$ , χρησιµοποιώντας τρία στρωµατοποιηµένα µοντέλα Cox.



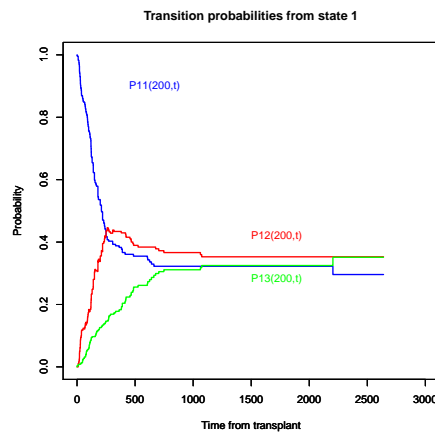
Σχήµα 72: Εκτιµήσεις των πιθανοτήτων παραµονής και µετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδοµένου ότι ένας ασθενής µε διάνυσµα συµµεταβλητών  $(0, 1, 1, 1)$  είναι υγιής την χρονική στιγµή  $t = 200$ , χρησιµοποιώντας τρία στρωµατοποιηµένα µοντέλα Cox.



Σχήμα 73: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(1, 0, 0, 0)$  είναι υγιής την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Σχήμα 74: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών  $(0, 1, 0, 0)$  είναι υγιής την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Από τις γραφικές παραστάσεις των Σχημάτων 71 – 74 μπορούμε να κάνουμε τις εξής παρατηρήσεις:

**1.** Η πιο ταχεία μείωση της πιθανότητας παραμονής στην κατάσταση 1, συνοδευόμενη και από την ταχύτερη αύξηση της πιθανότητας αποτυχίας σε σχέση με τον χρόνο, συναντάται στους ασθενείς με συμμεταβλητές  $(1, 0, 1, 1)$ . Μάλιστα, η τελευταία, ξεπερνάει τις πιθανότητες μετάβασης στις άλλες 2 καταστάσεις την χρονική στιγμή  $t = 363$  και αποκτά μέγιστη τιμή 0.895 για  $t \geq 2204$ .

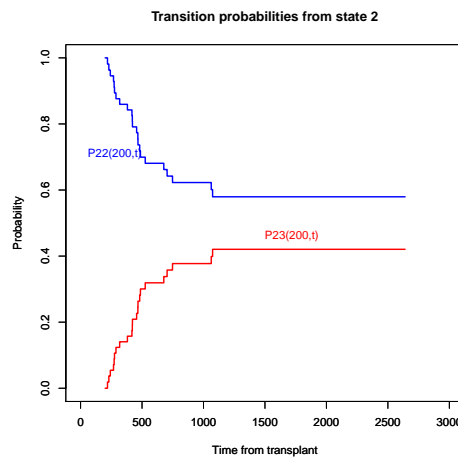
2. Αντίθετα, η πιο αργή μείωση της πιθανότητας παραμονής στην 1, ακολουθούμενη φυσικά και από την πιο αργή αύξηση της πιθανότητας αποτυχίας σε σχέση με τον χρόνο παρατηρείται στους ασθενείς με διάνυσμα συμμεταβλητών  $(0, 1, 0, 0)$ . Συγκεκριμένα, η τελευταία δεν ξεπερνάει ποτέ την πιθανότητα μετάβασης στην παροδική 2. Συγκεκριμένα για  $t \geq 2204$  το διάνυσμα πιθανοτήτων μετάβασης στις καταστάσεις 1,2 και 3, αντίστοιχα, είναι το  $(0.296, 0.353, 0.351)$ . Οι ασθενείς αυτοί, λοιπόν, έχουν την μικρότερη πιθανότητα αποτυχίας (Relapse/Death) με την πάροδο του χρόνου σε σχέση με τους ασθενείς των άλλων συμμεταβλητών που αναφέραμε παραπάνω.

### Υπολογισμός των $P_{2j}(200, t)$ , $\forall j = 2, 3$ :

Ομοίως με παραπάνω, κάνουμε προβλέψεις για τις πιθανότητες μετάβασης ενός ασθενή στις δυνατές καταστάσεις μετάβασης της παραπάνω ICTMC δεδομένου ότι την χρονική στιγμή  $t = 200$  βρίσκεται στην κατάσταση 2.

Για τις  $\hat{P}_{22}(200, t)$  και  $\hat{P}_{23}(200, t)$  (Σχήμα 75):

Σχήμα 75: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάνυσμα συμμεταβλητών  $(0, 0, 0, 0)$  βρίσκονται στην κατάσταση 2 την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.

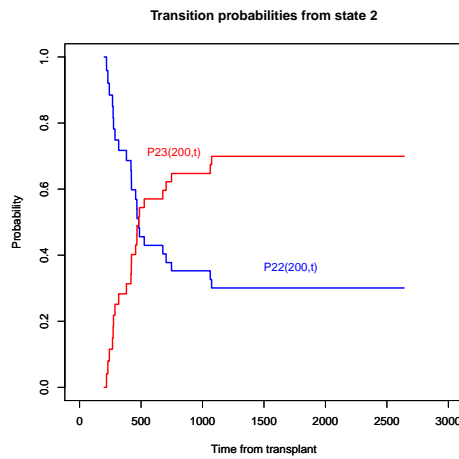


Από το γράφημα του Σχήματος 75 παρατηρούμε ότι αν ένας ασθενής με διάνυσμα συμμεταβλητών  $(0, 0, 0, 0)$  βρίσκεται στην κατάσταση 2 (GVHD) την χρονική στιγμή  $t = 200$ , είναι συστηματικά πιθανότερο, με το πέρασμα του χρόνου να παραμείνει στην κατάσταση αυτή από το να μεταβεί στην απορροφητική 3. Μάλιστα, για  $t \geq 1074$  το αντίστοιχο διάνυσμα πιθανοτήτων μετάβασης στις καταστάσεις 2 και 3, αντίστοιχα, είναι το  $(0.579, 0.421)$ .

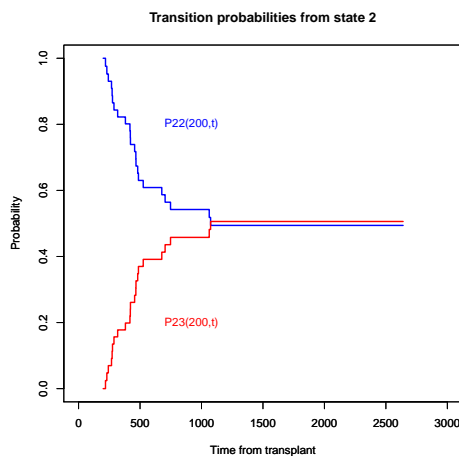
Ομοίως, για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 0, 1, 1)$ ,  $(0, 1, 1, 1)$ ,  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$ , μπορούμε να υπολογίσουμε τα εξής:

Για τις  $\hat{P}_{22}(200, t)$  και  $\hat{P}_{23}(200, t)$  (Σχήματα 76, 77, 78 και 78):

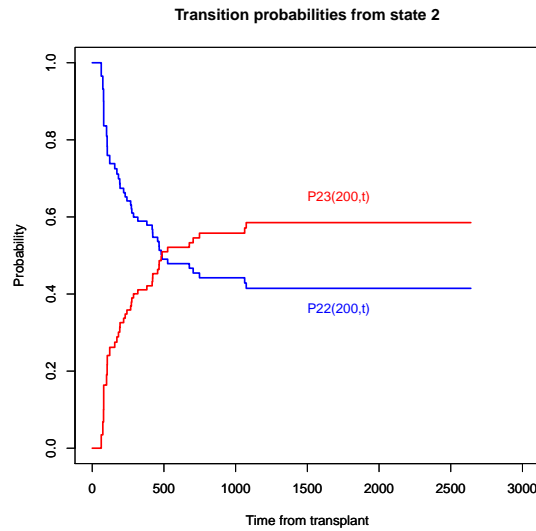
Σχήμα 76: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάνυσμα συμμεταβλητών  $(1, 0, 1, 1)$  βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



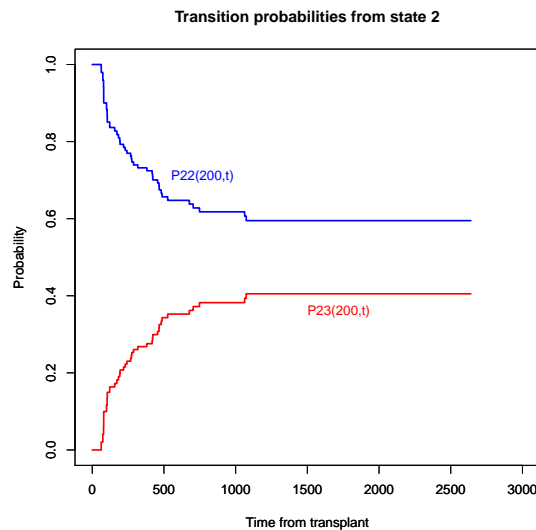
Σχήμα 77: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάνυσμα συμμεταβλητών  $(0, 1, 1, 1)$  βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.



Σχήμα 78: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών (1, 0, 0, 0) βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox



Σχήμα 79: Εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης στις άλλες δύο καταστάσεις της αλυσίδας, δεδομένου ότι ένας ασθενής με διάγνωση συμμεταβλητών (0, 1, 0, 0) βρίσκεται στην κατάσταση 2 την χρονική στιγμή  $t = 200$ , χρησιμοποιώντας τρία στρωματοποιημένα μοντέλα Cox.





Από τις γραφικές παραστάσεις, των Σχημάτων 76 – 79, μπορούμε να εξάγουμε τα εξής γενικά συμπεράσματα:

- 1.** Για τους ασθενείς με συμμεταβλητές  $(0, 1, 0, 0)$ , σε αντίθεση με τους ασθενείς με τα υπόλοιπα τρία διάνυσματα συμμεταβλητών, η πιθανότητα παραμονής στην κατάσταση 2 (GVHD) είναι συνεχώς μεγαλύτερη της πιθανότητας μετάβασης στην απορροφητική 3, με διάνυσμα πιθανοτήτων μετάβασης  $(0.595, 0.405)$  για  $t \geq 1074$ .
- 2.** Η ταχύτερη αύξηση της πιθανότητας αποτυχίας συνοδευόμενη και από την ταχύτερη μείωση της πιθανότητας παραμονής στην κατάσταση 2, με το πέρασμα του χρόνου, παρουσιάζεται στους ασθενείς με διάνυσμα συμμεταβλητών  $(1, 0, 1, 1)$ . Η πιθανότητα μετάβασης στην απορροφητική 3 ξεπερνάει την πιθανότητα παραμονής στην 2 την  $t = 481$  με μέγιστη τιμή 0.699 για  $t \geq 1074$ .
- 3.** Επίσης, αξιοσημείωτο είναι ότι για τους ασθενείς με διάνυσμα συμμεταβλητών  $(0, 1, 1, 1)$ , για  $t \geq 1074$  το αντίστοιχο διάνυσμα πιθανοτήτων παραμονής στην 2 και μετάβασης στην 3 είναι το  $(0.494, 0.506)$ . Για τους παραπάνω ασθενείς λοιπόν, σε βάθος χρόνου, είναι περίπου ισοπίθανη η παραμονή τους στην κατάσταση ασθένειας (GVHD) με την μετάβαση τους στην απορροφητική Relapse/Death.

## Χρήση εναλλακτικού illness-death μοντέλου:

### Υπολογισμός των $P_{1j}(0, t)$ , $\forall j = 1, 2, 3$ :

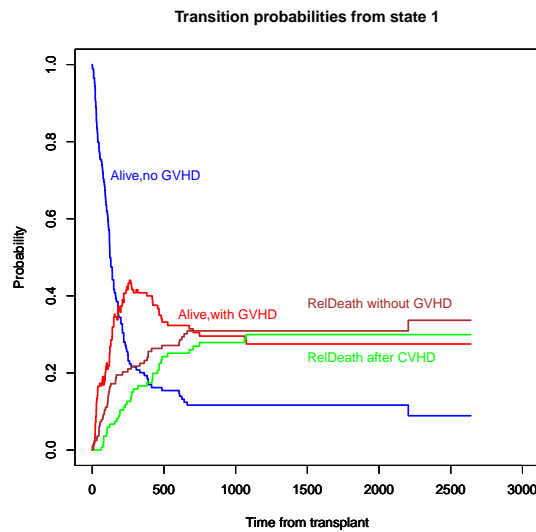
Ωστόσο, στην παραπάνω ανάλυση δεν έγινε διαχωρισμός μεταξύ των μεταβάσεων  $1 \Rightarrow 2 \Rightarrow 3$  και  $1 \Rightarrow 3$ . Όπως αναφέραμε και παραπάνω, κατασκευάζουμε ένα εναλλακτικό multi-state μοντέλο τεσσάρων καταστάσεων, δύο εκ των οποίων είναι απορροφητικές. Ο τύπος που συνδέει τις εκτιμήσεις πιθανοτήτων Aalen-Johansen είναι ο εξής:

$$\hat{P}_{13}(s, t) = \hat{P}_{13}^1(s, t) + \hat{P}_{13}^2(s, t)$$

Για έναν ασθενή με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  προσαρμόζουμε τρία στρωματοποιημένα (stratified) μοντέλα Cox και έχουμε:

Για τις  $P_{11}(0, t)$ ,  $P_{12}(0, t)$ ,  $P_{13}^2(0, t)$  και  $P_{13}^1(0, t)$  (Σχήμα 80):

Σχήμα 80: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  που είναι υγιής την χρονική στιγμή  $t = 0$  στο νέο μοντέλο multi-state.

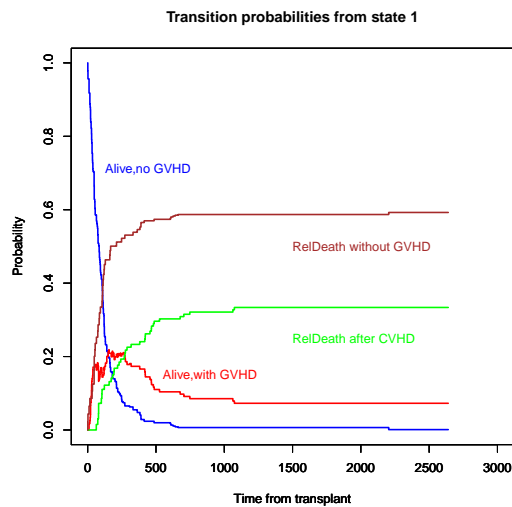


Ένας ασθενής με το παραπάνω διάγνωση συμμεταβλητών (Σχήμα 79) έχει μεγαλύτερη πιθανότητα να παραμείνει υγιής μέχρι την  $t = 180$  από όπου μέχρι την  $t = 704$  είναι πιο πιθανή η μετάβαση στην 2 και η παραμονή της εκεί. Τέλος, για  $t \geq 704$  επικρατεί η πιθανότητα θανάτου χωρίς να ασθενήσει πρώτα ο ασθενής, με μέγιστη πιθανότητα 0.337 για  $t \geq 2204$ .

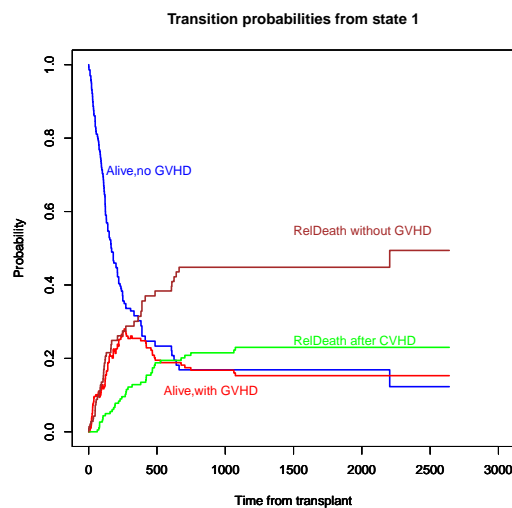
Επαναλαμβάνουμε για τους ασθενείς με συμμεταβλητές  $(1, 0, 1, 1)$ ,  $(0, 1, 1, 1)$ ,  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$ .

Για τις  $P_{11}(0, t)$ ,  $P_{12}(0, t)$ ,  $P_{13}^2(0, t)$  και  $P_{13}^1(0, t)$  (Σχήματα 81, 82, 83 και 84):

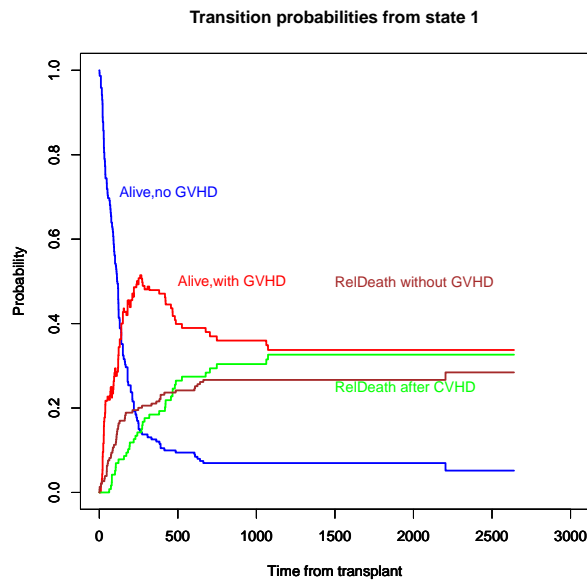
Σχήμα 81: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάνυσμα συμμεταβλητών  $(1, 0, 1, 1)$  που είναι υγιής την χρονική στιγμή  $t = 0$  στο νέο μοντέλο multi-state.



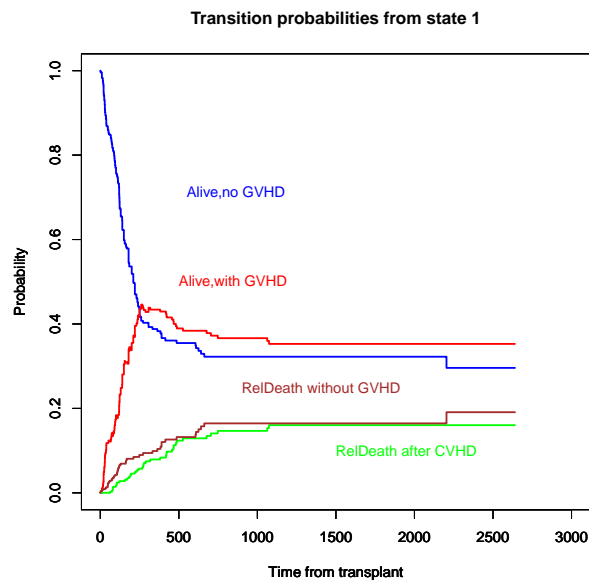
Σχήμα 82: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάνυσμα συμμεταβλητών  $(0, 1, 1, 1)$  που είναι υγιής την χρονική στιγμή  $t = 0$  στο νέο μοντέλο multi-state.



Σχήμα 83: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάγνωση συμμεταβλητών (1, 0, 0, 0) που είναι υγιής την χρονική στιγμή  $t = 0$  στο νέο μοντέλο multi-state.



Σχήμα 84: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάγνωση συμμεταβλητών (0, 1, 0, 0) που είναι υγιής την χρονική στιγμή  $t = 0$  στο νέο μοντέλο multi-state.



Με βάση τις γραφικές παραστάσεις των Σχημάτων 81 – 84 μπορούμε να καταλήξουμε στα εξής:

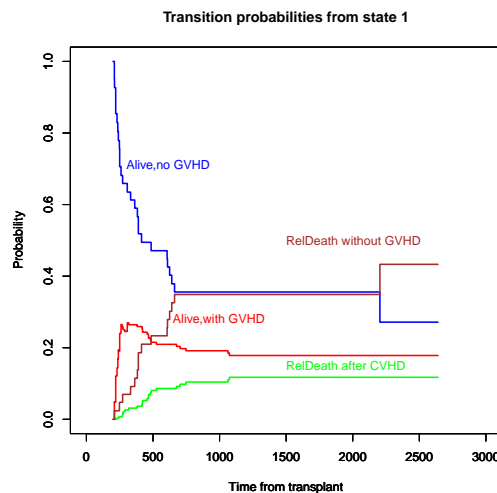
1. Οι ασθενείς με συμμεταβλητές  $(1, 0, 1, 1)$  έχουν την πιο απότομη αύξηση πιθανότητας απευθείας μετάβασης  $1 \Rightarrow 3$ , η οποία ξεπερνάει την πιθανότητα παραμονής στην 1 την  $t = 109$  και αποκτά μέγιστη πιθανότητα 0.593 για  $t \geq 2204$ .
2. Για τους ασθενείς με διανύσματα συμμεταβλητών  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$  τελικά επικρατεί η πιθανότητα μετάβασης στην 2 και παραμονής σε αυτή χωρίς να παρατηρηθεί η αποτυχία των ασθενών. Στους ασθενείς με συμμεταβλητές  $(0, 1, 0, 0)$ , ωστόσο, η παραπάνω πιθανότητα παίρνει σε βάθος χρόνου μεγαλύτερες τιμές, με μέγιστη τιμή 0.353 για  $t \geq 1074$ .

### Υπολογισμός των $P_{1j}(200, t)$ , $\forall j = 1, 2, 3$ :

Υπολογίζουμε τις πιθανότητες παραμονής και μετάβασης ενός ασθενή με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  όταν βρίσκεται την  $t = 200$  στην κατάσταση 1.

Για τις  $P_{11}(200, t)$ ,  $P_{12}(200, t)$ ,  $P_{13}^2(200, t)$  και  $P_{13}^1(200, t)$  (Σχήμα 85):

Σχήμα 85: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$  που είναι υγιής την χρονική στιγμή  $t = 200$  στο νέο μοντέλο multi-state.



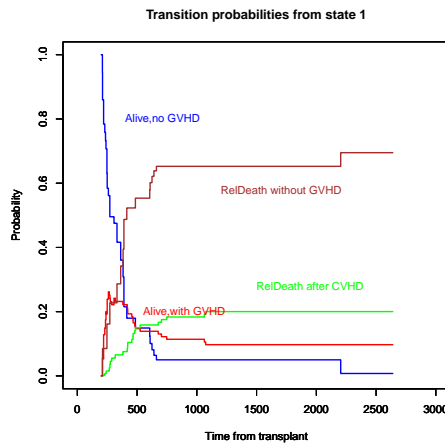
Όπως φαίνεται από το Σχήμα 85, για τους ασθενείς με συμμεταβλητές  $(0, 0, 0, 0)$  μέχρι την  $t = 2204$  είναι πιο πιθανή η παραμονή τους στην αρχική τους κατάσταση με την πιθανότητα της απευθείας αποτυχίας να ακολουθεί. Μάλιστα, στο διάστημα  $[662, 1074]$  έχουν σχεδόν αμελητέα διαφορά.

Ωστόσο, τελικά, για  $t \geq 2204$  επικρατεί με πιθανότητα 0.433 μετάβαση τους από την 1 στην 3 δίχως να περάσουν από την 2.

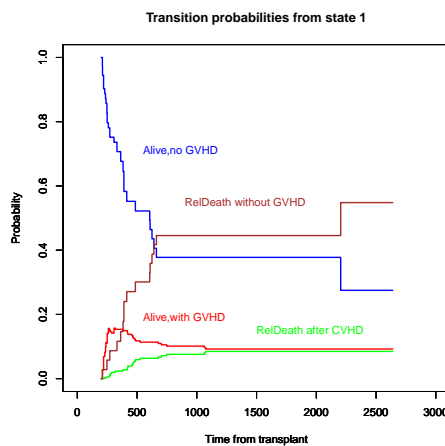
Επαναλαμβάνουμε για τους ασθενείς με συμμεταβλητές  $(1, 0, 1, 1)$ ,  $(0, 1, 1, 1)$ ,  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$ .

Για τις  $P_{11}(200, t)$ ,  $P_{12}(200, t)$ ,  $P_{13}^2(200, t)$  και  $P_{13}^1(200, t)$  (Σχήματα 86, 87, 88 και 89):

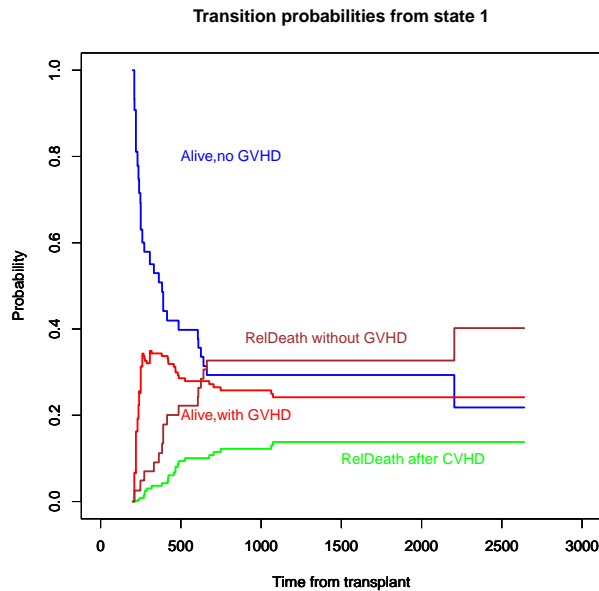
Σχήμα 86: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάγνωση συμμεταβλητών  $(1, 0, 1, 1)$  που είναι υγιής την χρονική στιγμή  $t = 200$  στο νέο μοντέλο multi-state.



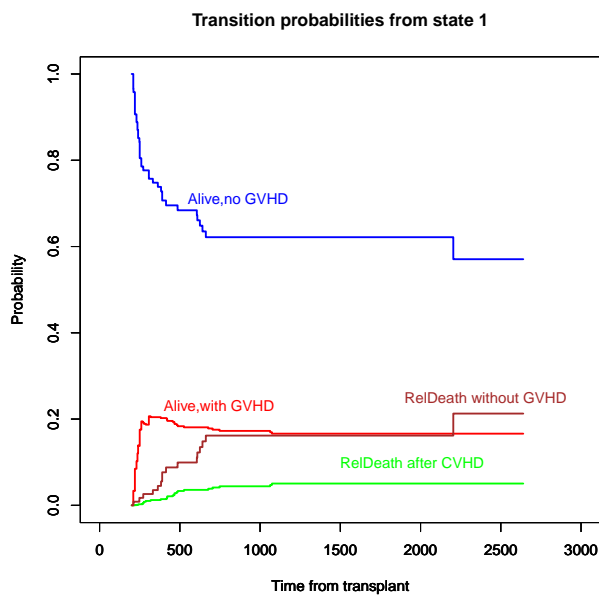
Σχήμα 87: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάγνωση συμμεταβλητών  $(0, 1, 1, 1)$  που είναι υγιής την χρονική στιγμή  $t = 200$  στο νέο multi-state μοντέλο.



Σχήμα 88: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάλυμα συμμεταβλητών  $(1, 0, 0, 0)$  που είναι υγιής την χρονική στιγμή  $t = 200$  στο νέο μοντέλο multi-state.



Σχήμα 89: Πιθανότητες παραμονής και μετάβασης στις υπόλοιπες καταστάσεις για έναν ασθενή με διάλυμα συμμεταβλητών  $(0, 1, 0, 0)$  που είναι υγιής την χρονική στιγμή  $t = 200$  στο νέο μοντέλο multi-state.



Από τα παραπάνω γραφήματα των Σχημάτων 86 – 89 μπορούμε να βγάλουμε τα εξής γενικά συμπεράσματα:

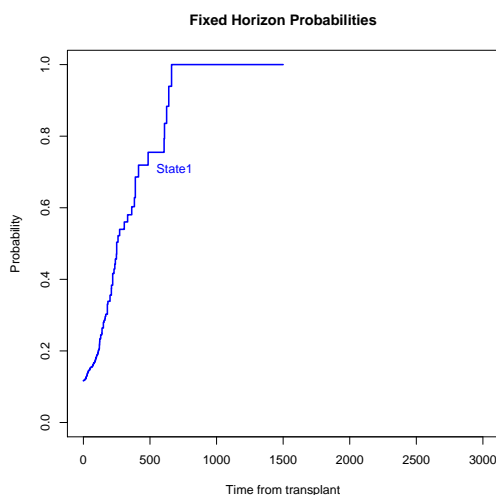
1. Οι ασθενείς με συμμεταβλητές  $(0, 1, 0, 0)$  είναι οι μόνοι από αυτούς που μελετούμε παραπάνω που έχουν συστηματικά μεγαλύτερη πιθανότητα να παραμείνουν στην αρχική τους κατάσταση από το να μεταβούν σε κάποια άλλη.
2. Για τους ασθενείς με διάνυσμα συμμεταβλητών  $(1, 0, 1, 1)$  βλέπουμε την πιο ταχεία αύξηση της πιθανότητας απευθείας μετάβασης  $1 \Rightarrow 3$ , η οποία ξεπερνάει την πιθανότητα παραμονής στην 1 την  $t = 383$  και επικρατεί των υπολοίπων αποκτώντας μέγιστη τιμή 0.695 για  $t \geq 2204$ .

Για όλες τις παραπάνω περιπτώσεις, οι πιθανότητες  $P_{22}(s, t)$  και  $P_{23}(s, t)$  έχουν ακριβώς τις ίδιες εκτιμήτριες με την προηγούμενή μας ανάλυση.

### Υπολογισμός των $P_{11}(s, 1500)$ :

Επίσης, για τους ασθενείς με συμμεταβλητές  $(0, 0, 0, 0)$ , δεδομένου ότι την χρονική στιγμή  $t = 1500$  η μαρκοβιανή αλυσίδα βρίσκεται στην κατάσταση 1 μπορούμε να εκτιμήσουμε, κινούμενοι ανάποδα, τις πιθανότητες με τις οποίες βρίσκονταν στην αρχική της κατάσταση στις προηγούμενες χρονικές στιγμές  $s \in [0, 1500]$  (Σχήμα 90).

Σχήμα 90: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(0, 0, 0, 0)$  στο χρονικό διάστημα  $[0, 1500]$ , δεδομένου ότι την χρονική στιγμή  $t = 1500$  είναι υγιής.

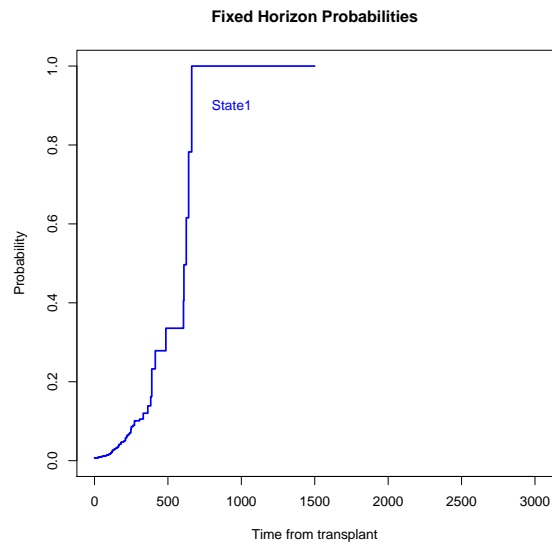


Από το Σχήμα 89 παρατηρούμε ότι η πιθανότητα παραμονής στην 1 για τους ασθενείς με διάνυσμα συμμεταβλητών  $(0, 0, 0, 0)$  αυξάνεται σχετικά γρήγορα και παίρνει την τιμή 1 για  $s \geq 662$ .

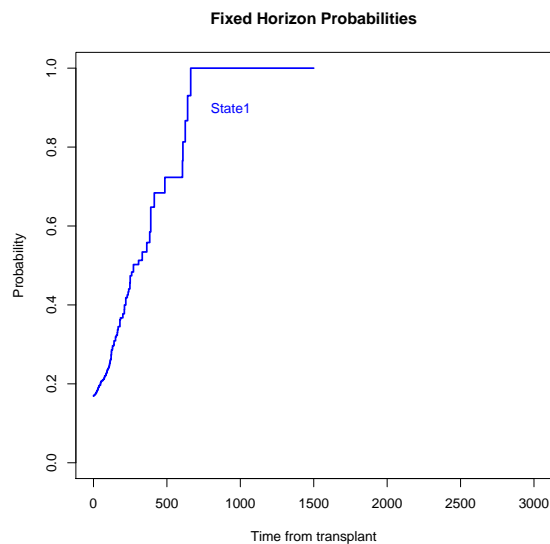


Για τους ασθενείς με συμμεταβλητές  $(1, 0, 1, 1)$ ,  $(0, 1, 1, 1)$ ,  $(1, 0, 0, 0)$  και  $(0, 1, 0, 0)$ , έχουμε (Σχήματα 91, 92, 93 και 94):

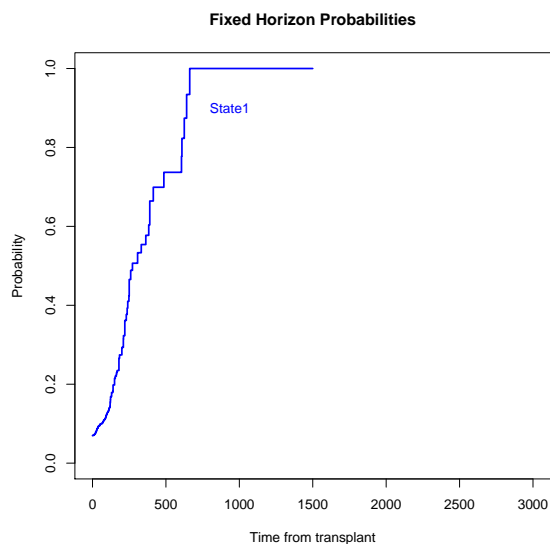
Σχήμα 91: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(1, 0, 1, 1)$  στο χρονικό διάστημα  $[0, 1500]$ , δεδομένου ότι την χρονική στιγμή  $t = 1500$  είναι υγιής.



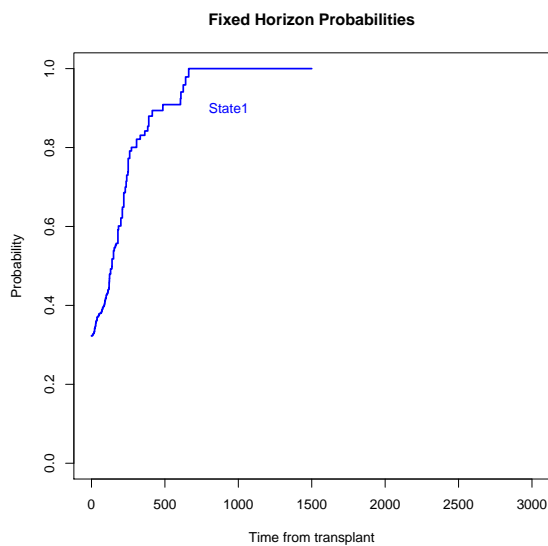
Σχήμα 92: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(0, 1, 1, 1)$  στο χρονικό διάστημα  $[0, 1500]$ , δεδομένου ότι την χρονική στιγμή  $t = 1500$  είναι υγιής.



Σχήμα 93: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(1, 0, 0, 0)$  στο χρονικό διάστημα  $[0, 1500]$ , δεδομένου ότι την χρονική στιγμή  $t = 1500$  είναι υγιής.



Σχήμα 94: Πιθανότητες παραμονής και μετάβασης ενός ασθενή με συμμεταβλητές  $(0, 1, 0, 0)$  στο χρονικό διάστημα  $[0, 1500]$ , δεδομένου ότι την χρονική στιγμή  $t = 1500$  είναι υγιής.



Από τις γραφικές παραστάσεις των Σχημάτων 91 – 94, μπορούμε να κάνουμε τις εξής παρατηρήσεις:

- 1.** Οι ασθενείς με συμμεταβλητές  $(1, 0, 1, 1)$ , την  $s = 0$  λαμβάνουν την μικρότερη, ανάμεσα στους ασθενείς των συμμεταβλητών που μελετήσαμε,  $\hat{P}_{11}(0, 1500) = 0.00661$  και αποκτούν την τιμή 1 για  $s \geq 662$ .
- 2.** Οι ασθενείς με συμμεταβλητές  $(0, 1, 0, 0)$  αποκτούν  $\hat{P}_{11}(s, 1500) = 1$  για  $s \geq 662$ , ομοίως. Ωστόσο, την  $s = 0$  είναι  $\hat{P}_{11}(0, 1500) = 0.322$ , η οποία είναι η μεγαλύτερη αρχική τιμή της παραπάνω πιθανότητας σε σχέση με τα υπόλοιπα διανύσματα συμμεταβλητών που μελετήσαμε.

## 4 Παράρτημα

Σε αυτό το σημείο θα αναφερθούμε εν συντομία στα πακέτα της γλώσσας R και σε κάποιες εντολές των οποίων χρησιμοποιήσαμε για την μοντελοποίηση και ανάλυση των προβλημάτων ανταγωνιστικών κινδύνων (competing-risks) και illness-death που μελετήσαμε παραπάνω.

### 4.1 Μελέτη των Klein και Moeschberger για τη μεταμόσχευση μυελού των οστών

Οι δύο βιβλιοθήκες της R οι οποίες υλοποιούν την μοντελοποίηση και ανάλυση του προβλήματος illness-death που μελετήσαμε στο Κεφάλαιο 3 (Klein and Moeschberger, 1997) είναι οι εξής:

**1. survival:** Είναι αναγκαία για να προβούμε σε ανάλυση προβλημάτων single-event ανάλυσης επιβίωσης.

**2. mstate:** Η βιβλιοθήκη αυτή, για την οποία η εγκατάσταση της survival είναι αναγκαία, περιέχει τις εντολές που θα μας επιτρέψουν την ανάλυση προβλημάτων επιβίωσης multi-state, και άρα και ανταγωνιστικών κινδύνων.

Ξεκινάμε με το σύνολο δεδομένων μας στο data frame, bmt, τις συμμεταβλητές μας, covs, και τον πίνακα των δυνατών μεταβάσεων του μοντέλου, tmat.

```
bmt<-data.frame(id,prtime,prstat,rftime,rfsstat,group1,group2,fab,mtx)
covs<-c("group1","group2","fab","mtx","prtime")
tmat<-trans.illdeath(names=c("Tplant","GVHD","RelDeath"))
```

Στην συνέχεια, φέρνουμε το αρχικό μας σύνολο δεδομένων σε μια επαυξημένη μορφή, msbmt, και προσδιορίζουμε τον πίνακα συχνοτήτων μεταβάσεων των ασθενών, freq.

```
msbmt<-msprep(time=c(NA,"prtime","rftime"),status=c(NA,"prstat","rfsstat"),
data=bmt,trans=tmat,keep=covs)
freq<-events(msbmt)
expcovs<-expand.covs(msbmt,covs[2:3],append=FALSE)
msbmt<-expand.covs(msbmt,covs,append=TRUE,longnames=FALSE)
```

Προσαρμογή τριών στρωματοποιημένων μοντέλων Cox στις ανά μετάβαση συναρτήσεις διακινδύνευσης:

```
c1<-coxph(Surv(Tstart,Tstop,status) group1.1+group2.1+group1.2+
group2.2+group1.3+group2.3+fab.1+fab.2+fab.3+mtx.1+mtx.2+mtx.3
+strata(trans),data=msbmt,method='breslow')
```

Προσαρμογή τριών μοντέλων Cox στις ανά μετάβαση συναρτήσεις διακινδύνευσης, θεωρώντας ότι οι συναρτήσεις διακινδύνευσης  $\lambda_{13}(t|\vec{z})$  και  $\lambda_{23}(t|\vec{z})$  έχουν ανάλογες βασικές συναρτήσεις:

```
c2<-coxph(Surv(Tstart,Tstop,status) group1.1+group2.1+group1.2+
group2.2+group1.3+group2.3+fab.1+fab.2+fab.3+mtx.1+mtx.2+mtx.3
+gvhd+strata(to),data=msbmt,method='breslow')
```

Προσαρμογή τριών μοντέλων Cox στις συναρτήσεις διακινδύνευσης του illness-death προβλήματός μας, θεωρώντας όμως ότι η αλυσίδα είναι semi-Markov:

```
c3<-coxph(Surv(Tstart,Tstop,status) group1.1+group2.1+group1.2+
group2.2+group1.3+group2.3+fab.1+fab.2+fab.3+mtx.1+mtx.2+mtx.3
+gvhd+prtime.3+strata(to),data=msbmt,method='breslow')
```

Για την ημι-παραμετρική μας ανάλυση θα αναφερθούμε μόνο στους ασθενείς με διάγνωση συμμεταβλητών (0, 0, 0, 0). Χρησιμοποιούμε, αρχικά, την εντολή `msfit` ώστε να δημιουργήσουμε ένα αντικείμενο που παριστά ένα μοντέλο multi-state, εν προκειμένω το illness-death των Klein και Moeschberger, με την Μαρκοβιανή ιδιότητα.

```
newd<-data.frame(group1=rep(0,3),group2=rep(0,3),fab=rep(0,3),mtx=rep(0,3),
trans=1:3)
bmt$group1<-as.factor(bmt$group1)
bmt$group2<-as.factor(bmt$group2)
bmt$fab<-as.factor(bmt$fab)
bmt$mtx<-as.factor(bmt$mtx)
newd$group1<-factor(newd$group1,levels=0:1,labels=levels(bmt$group1))
newd$group2<-factor(newd$group2,levels=0:1,labels=levels(bmt$group2))
newd$fab<-factor(newd$fab,levels=0:1,labels=levels(bmt$fab))
newd$mtx<-factor(newd$mtx,levels=0:1,labels=levels(bmt$mtx))
attr(newd,'trans')<-tmat
class(newd)<-c('msdata','data.frame')
newd<-expand.covs(newd,covs[1:4],longnames=F)
newd$strata=1:3
newd
msf1<-msfit(c1,newdata=newd,trans=tmat)
```

Με την εντολή `msf1$Haz` μπορούμε να προσδιορίσουμε τις εκτιμήτριες Breslow των βασικών συναρτήσεων των ανά μετάβαση σωρευτικών συναρτήσεων διακινδύνευσης. Όσον αφορά την ανά αιτία (cause-specific) ανάλυση που περιγράψαμε στην θεωρία, η πιο σημαντική εντολή, η οποία υλοποιεί όλους τους υπολογισμούς των πιθανοτήτων μετάβασης κατά Aalen-Johansen, είναι η `probtrans`.

Η εντολή `probtrans` και η εντολή `summary` αυτής, έχουν τρία ορίσματα, `predt`, `direction` και `from`, μεταβάλλοντας τα οποία μπορούμε να πάρουμε μια πληθώρα αποτελεσμάτων.

Γνωρίζουμε ότι σε μια ICTMC:

$$\hat{P}_{ij}(s, t) = \hat{P}(X(t) = j | X(s) = i)$$

Αν `direction='forward'`, τότε θέτουμε `predt=s` και μπορούμε να κάνουμε μελλοντικές προβλέψεις, ξεκινώντας από `from=i`, για τις πιθανότητες παραμονής και μετάβασης της αλυσίδας.

Αν `direction='fixedhorizon'`, τότε το όρισμα `predt` παριστάνει την χρονική στιγμή  $t$  και `from=j` είναι η κατάσταση που βρίσκεται η αλυσίδα την χρονική στιγμή  $t$ . Σε αυτή την περίπτωση, μπορούμε να κάνουμε εκτιμήσεις για τις πιθανότητες παραμονής και μετάβασης της αλυσίδας στις διάφορες καταστάσεις της για τις χρονικές στιγμές  $s \in [0, t]$ .

Για παράδειγμα, για τις πιθανότητες παραμονής και μετάβασης της αλυσίδας των ασθενών με διάγνωση συμμεταβλητών  $(0, 0, 0, 0)$ , ξεκινώντας από την κατάσταση 2 την χρονική στιγμή  $s = 200$ , έχουμε:

```
ft<-probtrans(msf1,predt=200)
f2<-summary(ft,from=2)
```

Επίσης, ορίζουμε το εναλλακτικό μοντέλο `multi-state`, με τις τέσσερις καταστάσεις δύο εκ των οποίων είναι απορροφητικές, που περιγράψαμε στη θεωρία, ώστε να υπάρχει διάκριση μεταξύ των απευθείας μεταβάσεων  $1 \Rightarrow 3$  και των  $1 \Rightarrow 2 \Rightarrow 3$ . Ο κώδικας της R που υλοποιεί το μοντέλο αυτό, είναι ο εξής:

```
msf1new<-msf1 msf1new$trans<-tmat2
tmat2<-transMat(x=list(c(2,4), c(3), c(), c()))
tmat2
```

Και στο νέο αντικείμενο `multi-state`, `msf1new`, που κατασκευάσαμε, μπορούμε να εφαρμόσουμε την ανάλυση που περιγράψαμε παραπάνω, κάνοντας χρήση των ίδιων εντολών.

Τέλος, για την μη-παραμετρική ανάλυση του προσδιορισμού των πιθανοτήτων μετάβασης κάνοντας χρήση των εκτιμητριών Nelson-Aalen, έχουμε:

```
c0<-coxph(Surv(Tstart,Tstop,status) strata(trans),data=msbmt,method='breslow')
msf0<-msfit(c0, vartype='aalen',trans=tmat)
bascumhaz2<-msf0$Haz
```

Η παραπάνω εντολή μας δίνει τις εκτιμήτριες Nelson-Aalen για τις μεταβάσεις του `illness-death` μοντέλου μας.

Όσον αφορά την εκτίμηση πιθανοτήτων χρησιμοποιούμε και πάλι την εντολή `probtrans`.

Συγκεκριμένα, για τις πιθανότητες μετάβασης  $\hat{P}_{1j}(0, t)$ , θα έχουμε:

```
tb<-probtrans(msf0,predt=0, method='aalen')
summary(tb,from=1)
```

## 4.2 Μελέτη του Lagakos για τον καρκίνο του πνεύμονα

Αρχικά, θα κάνουμε χρήση τριών βιβλιοθηκών της R οι οποίες υλοποιούν την μοντελοποίηση και ανάλυση του προβλήματος ανταγωνιστικών κινδύνων (competing risks) που είδαμε παραπάνω (Lagakos, 1978). Οι δύο εκ των τριών είναι αυτές που χρησιμοποιήσαμε στο μοντέλο illness-death, survival και mstate. Η τρίτη βιβλιοθήκη που θα χρειαστούμε είναι η εξής:

**cmprsk:** Από την βιβλιοθήκη αυτή, για την οποία η εγκατάσταση της survival είναι επίσης αναγκαία, μπορούμε να προσαρμόσουμε το μοντέλο των Fine και Gray (1999) στις συναρτήσεις διακινδύνευσης στο πρόβλημα μας.

Ξεκινάμε με το σύνολο δεδομένων μας σε ένα data frame, lagakos, και τις τρεις δυνατές καταστάσεις που μπορεί να βρίσκεται ένας ασθενής μετά το πέρας της μελέτης του, cause.

```
cause<-factor(cause,labels=c("event-free","LS","MS"))
lagakos<-data.frame(patnr,time,status,cause,performance,treatment,age)
```

Με την εντολή trans.comprisk της βιβλιοθήκης mstate κατασκευάζουμε τον πίνακα μεταβάσεων του προβλήματος μας, tmat, ως εξής:

```
tmat<-trans.comprisk(2, names=c('event-free','LS','MS'))
```

Στην συνέχεια, πρέπει να φέρουμε το σύνολο δεδομένων μας, lagakos, σε μια επαυξημένη μορφή, laglong, ώστε να προχωρήσουμε.

```
lagakos$stat1<-as.numeric(lagakos$status==1)
lagakos$stat2<-as.numeric(lagakos$status==2)
laglong<-msprep(time=c(NA,'time','time'),status=c(NA,'stat1','stat2'),
data=lagakos,keep=c('performance','treatment','age'),trans=tmat)
events(laglong)
laglong<-expand.covs(laglong,c('performance','treatment','age'))
```

Χρησιμοποιώντας το επαυξημένο αυτό σύνολο δεδομένων, laglong, μπορούμε να βρούμε και να σχεδιάσουμε τις εκτιμήτριες Naïve Kaplan-Meier αλλά και τις εκτιμήτριες των συναρτήσεων αθροιστικής επίπτωσης για τις δύο αιτίες αποτυχίας LS και MS, αντίστοιχα.

Για τις Naive Kaplan-Meier:

```
c1<-coxph(Surv(time,status) 1,data=laglong,subset=(trans==1),method='breslow')
c2<-coxph(Surv(time,status) 1,data=laglong,subset=(trans==2),method='breslow')
h1<-survfit(c1)
h2<-survfit(c2)
```

Για τις συναρτήσεις αθροιστικής επίπτωσης (CIF):

```
ci<-Cuminc(lagakos$time,lagakos$status)
```

Επιπλέον, όσον αφορά την προσαρμογή των μοντέλων Cox στις δύο συναρτήσεις διακινδύνευσης του προβλήματός μας, έχουμε:

```
c1<-coxph(Surv(time,status) performance.1+performance.2+treatment.1+
treatment.2+age.1+age.2+strata(trans),data=laglong)
```

Για τους ελέγχους ισότητας των συντελεστών των συμμεταβλητών μεταξύ των συναρτήσεων διακινδύνευσης και αναλογικότητας των βασικών (baseline) συναρτήσεων διακινδύνευσης, αντίστοιχα, έχουμε:

```
coxph(Surv(time,status) performance+treatment+age+strata(trans),data=laglong)
```

και:

```
c2<-coxph(Surv(time,status) performance.1+performance.2+treatment.1+treatment.2
+age.1+age.2+factor(trans),data=laglong,method='breslow')
```

Ακόμα, η τελευταία μέθοδος που είδαμε η οποία συμβάλει στην μείωση του συνολικού αριθμού των παραμέτρων στα προβλήματα ανταγωνιστικών κινδύνων, ο αλγόριθμος Reduced Rank, υλοποιείται ως εξής:

```
rr55<-redrank(Surv(Tstart,Tstop,status) performance+treatment+age,
data=laglong,R=1,print.level=1)
```

Επίσης, με την εντολή `cox.zph(c2)` μπορούμε να κάνουμε τον έλεγχο της υπόθεσης αναλογικότητας των μοντέλων μας κάνοντας χρήση των υπολοίπων Schoenfeld.

Όσον αφορά τις εκτιμήσεις των πιθανοτήτων παραμονής και μετάβασης από την αρχική κατάσταση της αλυσίδας, οι εκτιμήτριες που περιγράψαμε στη θεωρία δεν είναι παρά ειδικές περιπτώσεις των εκτιμητριών Aalen-Johansen.

Για το λόγο αυτό, κάνουμε και πάλι χρήση της εντολής `probtrans`.



Για παράδειγμα, για τους ασθενείς με συμμεταβλητές  $(0, 0, 0)$ , κατασκευάζουμε το αντίστοιχο αντικείμενο που αναφέρεται στο μοντέλο `lagakos`, `msf.zeros`, και υπολογίζουμε την  $\hat{P}_{01}(0, t)$ :

```
zeros<-data.frame(performance.1=c(0,0),performance.2=c(0,0),
treatment.1=c(0,0),treatment.2=c(0,0),age.1=c(0,0),age.2=c(0,0),
trans=c(1,2),strata=c(1,2))
msf.zeros<-msfit(c1,zeros,trans=tmat)
pt.zeros<-probtrans(msf.zeros,predt=0,method='aalen',
direction='forward')
summary(pt.zeros, from=1)
```

Τέλος, για την προσαρμογή του μοντέλου των Fine και Gray στα δεδομένα του Lagakos, χρησιμοποιούμε τον εξής κώδικα της βιβλιοθήκης `cmprsk`:

```
ftime<-lagakos$time
fstatus<-lagakos$status
th<-c(lagakos$performance,lagakos$treatment,lagakos$age)
cov<-matrix(th,ncol=3)
```

Για την  $\bar{\lambda}_1(t|\vec{z})$ :

```
z1<-crr(ftime,fstatus,cov)
```

Για την  $\bar{\lambda}_2(t|\vec{z})$ :

```
z2<-crr(ftime,fstatus,cov,failcode=2)
```

Χρησιμοποιώντας τις `subdistribution` συναρτήσεις διακινδύνευσης, από την θεωρία γνωρίζουμε ότι οι εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης δίνονται από τον εξής τύπο:

$$\hat{I}_k(t|\vec{z}) = 1 - \exp(-\hat{\Lambda}_{k,0}(t)\exp(\hat{b}_k^T \vec{z})) \quad \forall k \in D$$

όπου:

$$\hat{\Lambda}_{k,0} = \sum_{j=1}^N \frac{d_{kj}}{n_j}$$

Όσον αφορά τις εκτιμήσεις των συναρτήσεων αθροιστικής επίπτωσης (CIF) για τους ασθενείς με συμμεταβλητές  $(0, 0, 0)$  και  $(1, 1, 1)$ , αντίστοιχα, και για τις δύο αιτίες θανάτου, έχουμε:

```
z1.pr<-predict(z1,matrix(c(0,0,0,1,1,1),nrow=2,byrow=T))
```

```
z2.pr<-predict(z2,matrix(c(0,0,0,1,1,1),nrow=2,byrow=T))
```

```
z3.pr<-predict(z1,matrix(c(0,0,0,1,0,0),nrow=2,byrow=T))
```

```
z4.pr<-predict(z2,matrix(c(0,0,0,1,0,0),nrow=2,byrow=T))
```

Περισσότερες πληροφορίες για τους κώδικες που χρησιμοποιήσαμε μπορούν να βρεθούν στην περιγραφή των τριών προαναφερθεισών βιβλιοθηκών της R (Therneau, 2017, Putter et al., 2016 και Gray, 2015,) αλλά και σε άλλα κείμενα που υλοποιούν προβλήματα ανάλυσης επιβίωσης κάνοντας χρήση της R (Putter, 2016).

## 5 Βιβλιογραφία

- Aalen, O. O. , Borgan Ø. and Gjessing H. K. (2008). *Survival and Event History Analysis*. Oslo: Springer.
- Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, **5**, 141-150.
- Andersen, P. K. , Abildstrom, S. Z. and Rosthoj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research*. **11**, 203-215.
- Andersen, P. K. and Keiding, N. (2011). *Interpretability and importance of functionals in competing risks and multistate models*. Copenhagen: Wiley
- Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, **11**, 91-115.
- Arjas, E. and Eerola, M. (1993). On predictive causality in longitudinal studies. *Journal of Statistical Planning and Inference*, **34**, 361-386.
- Bakoyannis, G. and Touloumi, G. (2012). Practical methods for competing risk data: A review. *Statistical Methods in Medical Research*, **21**, 257-272.
- Beyersmann J. , Dettenkofer, M. , Bertz, H. and Schumacher M. (2007). A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards. *Statistics In Medicine*, **26**, 5360-5369.
- Beyersmann, J. , Allignol, A. and Schumacher, M. (2012). *Competing Risks and Multistate Models with R*. New York: Springer.
- Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.
- De Wreede, L. C. , Fiocco, M. and Putter, H. (2011). mstate: An R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, **38**, 7.
- Ferguson, N. , Datta, S. and Brock, G. (2012). msSurv: An R package for nonparametric estimation of multistate models. *Journal of Statistical Software*, **50**, 14.

- Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496-509.
- Fine, J. and Lindqvist, B. H. (2014). *Lifetime Data Analysis*. New York: Springer.
- Fiocco, M. , Putter, H. and Van Houwelingen, H. C. (2008). Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistical Methods in Medical Research*, **27**, 4340-4358.
- Fiocco, M. , Putter, H. , Van Houwelingen, J. C. and Van De Velde, J. C. (2006). Reduced rank proportional hazards model for competing risks: An application to a breast cancer trial. *Journal of Statistical Planning and Inference*, **136**, 1655-1668.
- Fiocco, M. , Putter, H. and Van Houwelingen, J. C. (2005). Reduced rank proportional hazards model for competing risks. *Biostatistics*, **6**, 465-478.
- Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression, Second Edition*. California: SAGE.
- Geskus, R. B. (2016). *Data Analysis with Competing Risks and Intermediate States*. New York: Taylor and Francis Group.
- Geskus, R. B. (2011). Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring. *Biometrics*, **67**, 39-49.
- Gray, B. (2015). *Package 'cmprsk'*. <https://cran.r-project.org/web/packages/cmprsk>.
- Haller, B. , Schmidt, G. and Ulm, K. (2013). Applying competing risks regression models: an overview. *Lifetime Data Analysis*, **19**, 33-58.
- Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, **38**, 8.
- Καρώνη, Χ. (2009). *Μοντέλα Αξιοπιστίας και Επιβίωσης*. Αθήνα: Συμεών.
- Klein, J. P. and Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

- Klein, P. J. and Shu, Y. (2002). Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research*, **11**, 117-139.
- Lagakos, S. W. (1978). A covariate model for partially censored data subject to competing causes of failure. *Journal of the Royal Statistical Society, Series C*, **27**, 235-241.
- Latouche, A. , Boisson, V. , Chevret, S. and Porcher, R. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistical Methods in Medical Research*, **26**, 965-974.
- Loprinzi, C. L. , Laurie, J. A. , Wieand, H. S. , Krook, J. E. , Novotny P. J. , Kugler, J. W. , Bartel, J. , Law, M. , Bateman, M. and Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology*, **12**, 601-607.
- Machin, D., Cheung, Y. B. and Parmar, M. K. B. (2006). *Survival Analysis: A Practical Approach, Second Edition*. London: Wiley.
- Marubin, E. and Valsecchi, M. G. (1994). *Analysing Survival Data from Clinical Trials and Observational Studies*. New York: Wiley.
- Moore, D. (2016). *Applied Survival Analysis Using R*. Switzerland: Springer.
- Norris, J. R. (1997). *Markov Chains*. Cambridge: Cambridge University Press.
- Prentice, R. L. and Kalbfleisch J. D. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541-554.
- Putter, H. (2016). Tutorial in biostatistics: Competing risks and multi-state models: Analyses using the mstate package. *Department of Medical Statistics and Bioinformatics Leiden*
- Putter, H. , Fiocco, M. and Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics In Medicine*, **26**, 2389-2430.
- Putter, H. , De Wreede, L. , Fiocco, M. and Geskus, R. (2016). *Package 'mstate'*. <https://cran.r-project.org/web/packages/mstate>).
- Sigman, K. (2016). *Lecture Notes on Stochastic Modeling I*. <http://www.ieor.columbia.edu/sigman/stochastic-I>, Columbia.

Therneau, T. M. (2017). *Package 'survival'*. <https://cran.r-project.org/web/packages/survival>.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, U.S.A.*, **72**, 20-22.