# ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

## ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

**Ανάλυση Επιβίωσης κατά Bayes**

Κατερίνα Δημ. Τσακανίκα

ΕΡΓΑΣΙΑ
Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος
Ειδίκευσης στη Στατιστική
Μερικής Παρακολούθησης (Part-time)

Αθήνα
Οκτώβριος 2007

# ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω καταρχήν, τον επιβλέποντα καθηγητή αυτής της διπλωματικής εργασίας, τον κ. Ιωάννη Ντζούφρα. Χωρίς την βοήθεια που μου προσέφερε με υποδείξεις και διορθώσεις και χωρίς την σταθερή υποστήριξή του και υπομονή σε όλα τα στάδια συγγραφής, η παρούσα εργασία δεν θα μπορούσε να είχε ολοκληρωθεί επιτυχώς.

Επίσης, ευχαριστώ τα μέλη της οικογένειας μου για την στήριξη που μου παρείχαν σε όλη την διάρκεια των προπτυχιακών και μεταπτυχιακών σπουδών μου. Σίγουρα η πορεία μου έως εδώ θα ήταν πολύ δυσκολότερη, εάν δεν είχα την συμπαράστασή τους.

# ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

Το 2000 απέκτησα το βασικό μου πτυχίο από το Τμήμα Μαθηματικών της Σχολής Θετικών Επιστημών του Πανεπιστημίου Ιωαννίνων, όπου ειδικεύτηκα στη Στατιστική κατά την διάρκεια σπουδών. Τον Οκτώβριο του 2004, ξεκίνησα την φοίτησή μου στο Μεταπτυχιακό Πρόγραμμα «Εφαρμοσμένη Στατιστική για Εκπαιδευτικούς» του Οικονομικού Πανεπιστημίου Αθηνών.

Διατηρώ δική μου επιχείρηση παράδοσης μαθημάτων θετικών επιστημών, όπως επίσης συνεργάζομαι και με πανεπιστημιακά φροντιστήρια.

# ABSTRACT

Katerina Tsakanika

## BAYESIAN SURVIVAL ANALYSIS

October 2007

The statistical analysis of the life – time data has been studied extensively in the past. Numerous statistical methods have been developed and adjusted for the analysis of survival time data.

Recently, interest of statisticians is focused on the implementation of Bayesian methods, since the development of computer technology made this implementation feasible for a wide variety of models.

In the present thesis we firstly present the theoretical distributions that have been widely used to describe survival times. Such distributions are the exponential, the Weibull and Gamma, the lognormal and the Paretto distributions.

The second Chapter deals with a presentation of Bayes theorem and its applications, such as Bayes factor, the Akaike' s Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Deviance Information Criterion (DIC).

Parametric and non- parametric methods of estimating survival functions and hazard functions are presented using the Cox proportional hazard regression model.

The analysis of all the methods discussed above, is comprehensively described through some illustrative examples of medical data using the statistical software WINBUGS, which facilitates MCMC methods to simulate values from the posterior distribution.

# ΠΕΡΙΛΗΨΗ

Κατερίνα Τσακανίκα

## ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ ΚΑΤΑ BAYES

Οκτώβριος 2007

Η Ανάλυση Επιβίωσης, δηλαδή η μελέτη της διάρκειας ζωής, είναι ένας τομέας στον οποίο η επιστήμη της Στατιστικής έχει συμβάλλει εδώ και πολλά χρόνια. Για τη ανάλυση δεδομένων, που αφορούν χρόνους επιβίωσης, έχουν αναπτυχθεί και εφαρμοστεί πολλές στατιστικές μέθοδοι.

Τα τελευταία χρόνια το ενδιαφέρον των ερευνητών φαίνεται να επικεντρώνεται στις μεθόδους κατά Bayes, καθώς με την εξέλιξη των δυνατοτήτων της πληροφορικής και των ηλεκτρονικών υπολογιστών είναι δυνατή η εφαρμογή τους σε ένα μεγάλο εύρος μοντέλων.

Στην παρούσα εργασία γίνεται καταρχήν αναφορά στις θεωρητικές κατανομές, οι οποίες προσεγγίζουν καλύτερα τα δεδομένα χρόνου επιβίωσης, καθώς και αναλυτική παρουσίαση του θεωρήματος Bayes και των εφαρμογών του. Στη συνέχεια αναλύεται το μοντέλο αναλογικού κινδύνου του Cox, το οποίο μελετάει την επίδραση επεξηγηματικών μεταβλητών πάνω στην συνάρτηση κινδύνου των δεδομένων επιβίωσης. Επίσης παρουσιάζεται και το παραμετρικό μοντέλο, βασισμένο στην Weibull κατανομή.

Η ανάλυση των κατανομών και των παραπάνω μεθόδων, όπως και οι εφαρμογές του θεωρήματος Bayes παρουσιάζονται μέσα από παραδείγματα πάνω σε ιατρικά δεδομένα, που αναλύονται με την προσομοίωση τυχαίων τιμών από την εκ των υστέρων κατανομή μέσω του στατιστικού προγράμματος WINBUGS, το οποίο αποτελεί τα τελευταία χρόνια ένα σημαντικό εργαλείο για κάθε ερευνητή.

# ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

# ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

# ΚΑΤΑΛΟΓΟΣ ΓΡΑΦΗΜΑΤΩΝ

# ΚΕΦΑΛΑΙΟ 1

## 1.1 ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ

Η Ανάλυση Επιβίωσης είναι, στην πιο απλή της ερμηνεία, η μελέτη της διάρκειας ζωής, δηλαδή ο χρόνος μεταξύ της γέννησης και του θανάτου ενός ανθρώπου, μίας μηχανής κτλ. Συχνά χρησιμοποιούμε τον όρο για τη μελέτη της διάρκειας του χρόνου μεταξύ δύο συγκεκριμένων γεγονότων.

Το επίπεδο γνώσης που κατέχουμε σήμερα για την Ανάλυση Επιβίωσης προκύπτει από μία διαδικασία εξέλιξης που διήρκησε αρκετά χρόνια και η οποία είχε ιδιαίτερη ανάπτυξη τα τελευταία χρόνια. Διάφοροι τομείς έρευνας, όπως Βιολογία, Ιατρική, Φαρμακευτική ακόμα και κλάδοι της Βιομηχανίας κ.α. συνέβαλαν σε αυτή την ανάπτυξη, προσπαθώντας να βρουν λύση για να αντιμετωπίσουν τα διάφορα προβλήματά τους.

Το 1662 στην Μ. Βρετανία δημοσιεύτηκε το πρώτο βιβλίο, το οποίο είχε καταγεγραμμένους καταλόγους με γεννήσεις και θανάτους, που αναφέρονταν στις προηγούμενες δεκαετίες κι έτσι ήταν η πρώτη φορά που οι "θάνατοι" αντιμετωπίστηκαν ως "γεγονότα", για τα οποία έγιναν αναλυτικές μελέτες (Δημάκη - 2006, σελ. 2)

Οι παγκόσμιοι πόλεμοι που ακολούθησαν, έδωσαν το έναυσμα για ανάπτυξη της έρευνας στον τομέα της αξιοπιστίας και στη μελέτη της διάρκειας ζωής των στρατιωτικών στρατευμάτων, αλλά και αργότερα η έρευνα επικεντρώθηκε στην μελέτη κάποιων ιδιαίτερων πιθανοθεωρητικών προβλημάτων σχετιζομένων με την "παύση λειτουργίας" και την "αντικατάσταση" εξαρτημάτων μηχανικών ή ηλεκτρικών κυκλωμάτων, όπως κάποιας βαλβίδας ή ενός θερμοστάτη σε ένα μηχανικό κύκλωμα, μίας λυχνίας ή μίας αντίστασης σε ένα ηλεκτρικό. Υπήρξε δηλαδή μεγάλη πρόοδος στη Βιομηχανία των ηλεκτρονικών συσκευών.

Έπειτα οι μέθοδοι της Ανάλυσης Επιβίωσης είχαν τεράστιες εφαρμογές σε κλινικά δεδομένα και σκοπός ήταν να απαντηθούν ερωτήματα όπως ποια είναι η πιθανότητα ένας ασθενής να ζήσει μέχρι μια συγκεκριμένη χρονική στιγμή ή με ποιο ρυθμό θα πεθάνουν κάποιοι ασθενείς, οι οποίοι έχουν ήδη επιβιώσει μέχρι ένα συγκεκριμένο χρονικό σημείο, κ.α. (Δημάκη – 2006, σελ. 2)

Η ανάπτυξη των επιχειρησιακών ερευνών κατέδειξε ότι υπάρχουν και πολλά άλλα προβλήματα και μοντέλα διαφορετικής υφής, στα οποία η Ανάλυση Επιβίωσης μπορεί να εφαρμοστεί και να προσφέρει λύσεις. Ακόμα περισσότερο με την εξέλιξη των ηλεκτρονικών υπολογιστών η εφαρμογή και η ανάπτυξη της γίνεται ολοένα και μεγαλύτερη, καθώς παράγονται όλο και περισσότερα λογισμικά πακέτα για το σκοπό αυτό.

## 1.2 Η Ανάλυση Επιβίωσης: Εισαγωγικές έννοιες

### 1.2.1 Χρόνος Επιβίωσης και Συνάρτηση Επιβίωσης

Ως χρόνος επιβίωσης μπορεί να ορισθεί ο χρόνος μέχρι να συμβεί ένα συγκεκριμένο γεγονός. Το γεγονός μπορεί να είναι η εμφάνιση μιας ασθένειας, η εξέλιξη ή η επιτυχία μιας θεραπείας σε κάποια ασθένεια, ο θάνατος του ασθενούς, η παύση λειτουργίας μιας ηλεκτρικής μηχανής κ.α. Ο χρόνος επιβίωσης ονομάζεται και χρόνος ως το "γεγονός" ή την "αποτυχία".

Παράδειγμα 1.2α:

Έστω ότι θεωρούμε έναν πληθυσμό, ο οποίος αποτελείται από όμοια εξαρτήματα, π.χ. ηλεκτρικούς λαμπτήρες. Κάθε ένα από τα όμοια αυτά εξαρτήματα χαρακτηρίζεται από μια μη-αρνητική τυχαία μεταβλητή $T$, που παριστάνει το χρονικό διάστημα από τη στιγμή που το συγκεκριμένο εξάρτημα τίθεται σε χρήση μέχρι τη χρονική στιγμή που αυτό παύει να λειτουργεί.

Παράδειγμα 1.2β:

Σε μια κλινική έρευνα εξετάζεται ένας πληθυσμός ασθενών με καρκίνο. Η μη-αρνητική τυχαία μεταβλητή $T$ παριστάνει το χρόνο ιατρικής παρακολούθησης του κάθε ασθενή από τη στιγμή εκδήλωσης της νόσου μέχρι το θάνατό του.

Η τυχαία μεταβλητή του χρόνου $T$ μπορεί να είναι είτε διακριτή είτε συνεχής.

- Συνεχής τυχαία μεταβλητή

Στην περίπτωση συνεχούς μεταβλητής, η $f_T(t)$ παριστάνει την συνάρτηση πυκνότητας πιθανότητας, η οποία ονομάζεται και **πυκνότητα αποτυχίας**, όπου $t \geq 0$ ενώ η $F_T(t) = \int_0^t f_T(x)dx = P(T \leq t)$ , $t \geq 0$ παριστάνει την αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής $T$, ή αλλιώς **αθροιστική συνάρτηση κατανομής αποτυχίας.**

Σε μια έρευνα ενδιαφερόμαστε για την πιθανότητα η συνιστώσα του συστήματος να μην έχει "αποτύχει" έως τη χρονική στιγμή $t$. Η συνάρτηση που καθορίζει αυτή την πιθανότητα ονομάζεται **συνάρτηση επιβίωσης** (survivor function) και συμβολίζεται ως $S(t) = P(T \geq t) = 1 - F_T(t)$ , $t \geq 0$.

Προφανώς αφού $F_T(0)=0$ και $F_T(1)= 0$ θα είναι $S(0)=1$ και $S(1)=0$.

- Διακριτή τυχαία μεταβλητή

Στην περίπτωση διακριτής μεταβλητής, η $p_T(t) = P(T = t)$ παριστάνει την συνάρτηση πιθανότητας, που ονομάζεται και **πυκνότητα αποτυχίας**, όπου $t = 0, 1, 2, \ldots$ ενώ η $F_T(t) = P(T \leq t) = \sum_{x \leq t} p_T(x)$ , $t \geq 0$ παριστάνει την αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής $T$.

### 1.2.2 Συνάρτηση Κινδύνου

- Χρόνος συνεχής

Σύμφωνα με τον ορισμό της αθροιστικής συνάρτησης κατανομής ισχύει, όπως είδαμε, ότι

$$F_T(t) = P(T \leq t) = \int_0^t f_T(x)dx \, , \, t \geq 0.$$

Έτσι αν $\Delta = [\, t \, , \, t + dt]$ τότε θα ισχύει η προσεγγιστική σχέση $P(\, t \leq T < t + dt) \approx f_T(t)dt$, δοθέντος όμως ότι $T \geq t$, έχουμε την δεσμευμένη πιθανότητα $P(\, t \leq T < t + dt \,/\, T \geq t)$ για $t \geq 0$, που εκφράζει την πιθανότητα ένα ενδεχόμενο να συμβεί την αμέσως επόμενη χρονική στιγμή (Congdon - 2001, σελ. 425).

Ορίζεται έτσι η **Συνάρτηση Κινδύνου** $h_T(t)$ (Hazard rate function) που εκφράζει την πιθανότητα η συνιστώσα του συστήματος που εξετάζουμε η οποία έχει "επιζήσει" για χρόνο *t,* να τεθεί εκτός λειτουργίας την αμέσως επόμενη χρονική στιγμή, δηλαδή στο αμέσως επόμενο μικρό χρονικό διάστημα *[t , t + dt]*. Είναι δηλαδή μια δεσμευμένη πιθανότητα.

Ισχύει $h_T(t) = \lim_{dt\to 0} \frac{P(t \le T < t+dt \mid T \ge t)}{dt} = \lim_{dt\to 0} \frac{P(t \le T < t+dt)}{dtP(T \ge t)} = \frac{f_T(t)}{S(t)}$.

Αφού $F_T(t) = \int_0^t f_T(x)dx \ , t \ge 0 \Leftrightarrow f_T(t) = F'_T(t) = -S'(t), t \ge 0$

έχουμε:

$$h_T(t) = \frac{-S'(t)}{S(t)} \Leftrightarrow$$

$$-h_T(t) = \left[\ln\left(S(t)\right)\right]' \Leftrightarrow$$

$$-\int_0^t h_T(x)dx = \ln\left[S(t)\right] - \ln\left[S(0)\right] \Leftrightarrow$$

$$S(t) = \exp\left[-\int_0^t h_T(x)dx\right], t \ge 0$$

Προφανώς η συνάρτηση πυκνότητας πιθανότητας $f_T(t) = h_T(t)\cdot S(t)$ της συνεχούς μεταβλητής *T* γίνεται:

$$f_T(t) = h_T(t) \exp\left[-\int_0^t h_T(x)dx\right], t \ge 0,$$

δηλαδή ορίζεται μονοσήμαντα από την συνάρτηση κινδύνου $h_T(t)$.

Παράδειγμα 1.2γ:

Έστω ότι η συνάρτηση βαθμού κινδύνου του χρόνου λειτουργίας ενός είδους μηχανών είναι η $h_T(t) = \frac{4}{t}$, $t \ge 1$. Τότε η συνάρτηση επιβίωσης θα είναι

$$S(t) = \exp\left[-\int_1^t \frac{4}{t}dx\right] = \exp\left[-4\left(\ln t - \ln 1\right)\right] = \exp\left(-4\ln t\right) = t^{-4} = \frac{1}{t^4}, \text{ με } t \ge 1.$$

- Χρόνος διακριτός

Στην περίπτωση της διακριτής κατανομής η συνάρτηση βαθμού κινδύνου θα είναι η $h_T(t) = \frac{P(T=t)}{P(T \geq t)} = \frac{P(T=t)}{S(t)}$, $t$ =0, 1, 2,.....

Ισχύει τότε:

$$h_T(t) = \frac{P(T \geq t) - P(T \geq t+1)}{P(T \geq t)} \Leftrightarrow$$

$$h_T(t)P(T \geq t) = P(T \geq t) - P(T \geq t+1) \Leftrightarrow$$
$$P(T \geq t+1) - [1 - h_T(t)]P(T \geq t) = 0$$

Η παραπάνω εξίσωση διαφορών έχει μοναδική λύση την

$$P(T \geq t) = P(T \geq 0)\prod_{i=0}^{t-1}(1 - h_T(i)),$$ για $t$ = 0, 1, 2,... (Δημάκη, σελ 13 - 2006)

Όμως $P(T \geq 0) = S(0) = 1$, άρα

$$P(T \geq t) = \prod_{i=0}^{t-1}(1 - h_T(i)),$$ για $t$ = 0, 1, 2,...

Οπότε η συνάρτηση επιβίωσης είναι:

$$S(t) = \prod_{i=0}^{t-1}(1 - h_T(i)),$$ για $t$ = 0, 1, 2,...

Παράδειγμα 1.2δ:

Έστω ότι η συνάρτηση βαθμού κινδύνου του χρόνου λειτουργίας ενός είδους μηχανών είναι η $h_T(t) = a$ , $0 < a < 1$ και για $t$ =0, 1, 2…..... Τότε η συνάρτηση επιβίωσης θα είναι

$$S(t) = \prod_{i=0}^{t-1}(1 - a) = (1 - a)^t$$ , για $t$ =0, 1, 2....

### 1.2.3 Δεδομένα Επιβίωσης – Περικομμένα Δεδομένα

Τα δεδομένα επιβίωσης περιλαμβάνουν όλους τους χρόνους επιβίωσης όλων των μονάδων που συμμετέχουν στο εκάστοτε πείραμα. Είναι όμως δύσκολο όλοι οι χρόνοι να είναι συγκεκριμένοι και γνωστοί. Λόγου χάρη, στο Παρ. 1.2β ενδέχεται

ορισμένοι ασθενείς να εξακολουθούν να ζουν ή να έχει επιτευχθεί πλήρης ίαση ως το τέλος της έρευνας, συνεπώς οι ακριβείς χρόνοι επιβίωσης να μην είναι γνωστοί. Οι χρόνοι αυτοί ονομάζονται περικομμένοι (censored observations). Γενικά περικομμένες παρατηρήσεις προκύπτουν όταν μερικές από τις μονάδες του πειράματος χάνονται κατά τη διάρκειά του.

Υπάρχουν οι παρακάτω κατηγορίες περικομμένων παρατηρήσεων:

1. Όταν ο ερευνητής επιλέγει να προκαθορίσει τη χρονική διάρκεια της έρευνας, οι χρόνοι επιβίωσης των υπό εξέταση μονάδων που "κατέληξαν" εντός της συγκεκριμένης προκαθορισμένης διάρκειας είναι οι ακριβείς και ονομάζονται μη περικομμένοι χρόνοι. Αντιθέτως οι πραγματικοί χρόνοι των υπό εξέταση μονάδων που δεν "κατέληξαν" στην διάρκεια της έρευνας δεν είναι γνωστοί (ή ακόμα και αν είναι γνωστοί είναι μεγαλύτεροι από τη διάρκεια της έρευνας) και ονομάζονται περικομμένοι (left cencoring). Οι χρόνοι αυτοί θεωρούνται ίσοι με το χρόνο που διαρκεί το πείραμα, όμως δεν αντιστοιχούν στο χρόνο θανάτου, αλλά στο γεγονός ότι ήταν "ζωντανοί" μέχρι εκείνη τη στιγμή.
2. Πάλι στην περίπτωση που είναι προκαθορισμένη η χρονική διάρκεια της έρευνας, αλλά κάποιες μονάδες "καταλήγουν" εντός της συγκεκριμένης διάρκειας για άλλους λόγους π.χ. ένας ασθενής αποφασίζει να μη συμμετέχει άλλο στην έρευνα και αποχωρεί πριν αυτή τελειώσει. Εδώ οι παρατηρούμενοι χρόνοι είναι μικρότεροι από τους πραγματικούς χρόνους επιβίωσης και ονομάζονται και αυτοί περικομμένοι.
3. Στην περίπτωση που ο ερευνητής επιλέγει να προκαθορίσει ένα συγκεκριμένο ποσοστό επιτυχίας και μόλις το επιτύχει να σταματήσει την έρευνα, οι περικομμένες παρατηρήσεις θεωρούνται ίσες με το χρόνο επιβίωσης της μεγαλύτερης μη περικομμένης παρατήρησης.

Η πυκνότητα πιθανότητας, όπως είδαμε, είναι $f_T(t) = h_T(t) \cdot S(t)$. Σύμφωνα με τον Congdon (2001, σελ. 428) μπορούμε να χρησιμοποιήσουμε ένα δείκτη $\delta_i$ $(i = 0, 1)$, για την περιγραφή της συνάρτησης πυκνότητας πιθανότητας των περικομμένων δεδομένων και θα είναι $f_T(t) = h_T(t)^{\delta i} \cdot S(t)$. Θέτουμε $\delta_i = 0$ για τα περικομμένα δεδομένα που

"καταλήγουν" πριν το τέλος της έρευνας (right censoring) και $\delta_i = 1$ για τα δεδομένα για τα οποία έχει παρατηρηθεί χρόνος "αποτυχίας".

### 1.2.4 Μέση Υπολοιπόμενη Ζωή

Η τυχαία μεταβλητή $T$ εκφράζει το χρόνο μέχρι την "αποτυχία" μίας συνιστώσας ενός συστήματος. Η κατανομή του χρόνου έχει συνάρτηση πυκνότητας πιθανότητας την $f_T(t)$.

Ως **Μέση υπολοιπόμενη ζωή** (Mean residual life at time $t$) ορίζεται η συνάρτηση $\mu_T(t) = E( T-t \mid T > t)$ , όπου $t \geq 0$ και εκφράζει την αναμενόμενη ζωή μιας συνιστώσας που έχει ήδη ηλικία $t$, δηλαδή έχει επιβιώσει ως τη χρονική στιγμή $t$ και εξακολουθεί να λειτουργεί. Είναι δηλαδή δείκτης γήρανσης ενός ατόμου ή ενός εξαρτήματος (Δημάκη, σελ.18 - 2006).

Η μέση υπολοιπόμενη ζωή $\mu_T(t) = E( T-t \mid T > t)$ μπορεί να υπολογιστεί μέσω της συνάρτησης επιβίωσης $S(t)$.

Στην περίπτωση που ο χρόνος $T$ είναι συνεχής μεταβλητή ισχύει

$$\mu_{\mathrm{T}}(t) = \frac{1}{S(t)} \int_t^{\infty} S(x)dx \text{ , για } t \geq 0,$$

ενώ στην περίπτωση που είναι διακριτή μεταβλητή ισχύει:

$$\mu_{\mathrm{T}}(t) = \frac{1}{S(t)} \sum_{x=t}^{\infty} S(t) \text{ , για } t = 0, 1, 2, \ldots\ldots \text{ (Δημάκη – 2006, σελ.19)}$$

### 1.3 Στατιστικά Μοντέλα

Η ανάλυση των δεδομένων επιβίωσης επικεντρώνεται κυρίως στην εύρεση της συνάρτησης επιβίωσης $S(t)$, όπου ο χρόνος $t$ μπορεί να είναι συνεχής ή διακριτός και όπως αναφέρθηκε ήδη, η συνάρτηση αυτή εκφράζει την πιθανότητα η συνιστώσα του συστήματος να μην έχει "αποτύχει" έως τη χρονική στιγμή $t$. Στη φύση όμως υπάρχουν πολλές αιτίες, οι οποίες μπορεί να οδηγήσουν στην "αποτυχία" ενός συστήματος που

μελετάται. Προφανώς είναι πολύ δύσκολο, αν όχι αδύνατον, τα φυσικά αυτά αίτια να απομονωθούν και να ληφθούν υπόψη στην επιλογή ενός μοντέλου, το οποίο θα περιγράφει ακριβώς τα χαρακτηριστικά του υπό μελέτη συστήματος. Έτσι ορισμένες θεωρητικές στατιστικές κατανομές, οι οποίες ικανοποιούν έναν επαρκή αριθμό ιδιοτήτων και αποτελούν μια καλή προσέγγιση σε διάφορα φυσικά φαινόμενα, είναι αυτές που χρησιμοποιούνται πιο συχνά για τη μελέτη δεδομένων επιβίωσης.

Παρακάτω θα δούμε αναλυτικά τις κατανομές αυτές και τις ιδιότητές τους.

### 1.3.1 Η εκθετική κατανομή – *Exp(λ)*

Η απλούστερη και σημαντικότερη κατανομή είναι η εκθετική. Οι σπουδαίες μαθηματικές ιδιότητες της κατανομής αυτής την καθιστούν πολύ ενδιαφέρουσα στις πρακτικές εφαρμογές.

Όταν ο χρόνος επιβίωσης *T* ακολουθεί την εκθετική κατανομή με παράμετρο *λ*, η συνάρτηση πυκνότητας είναι η $f_T(t) = \lambda e^{-\lambda t}$, $t \geq 0$ και η αθροιστική συνάρτηση κατανομής αποτυχίας είναι η $F_T(t) = 1 - e^{-\lambda t}$, $t \geq 0$. Έτσι η συνάρτηση επιβίωσης είναι η $S(t) = e^{-\lambda t}$ και η συνάρτηση βαθμού κινδύνου θα είναι η $h_T(t) = \lambda$ για $t \geq 0$, δηλαδή είναι σταθερός αριθμός, ανεξάρτητος από την ηλικία του ατόμου που μελετάται, γι' αυτό δεν υπάρχει φθορά και η "αποτυχία" ή ο "θάνατος" είναι ένα τυχαίο γεγονός ανεξάρτητο του χρόνου.

Η μέση υπολοιπόμενη ζωή θα είναι

$$\mu_T(t) = \frac{1}{S(t)}\int_t^\infty S(x)dx = \frac{1}{e^{-\lambda t}}\int_t^\infty e^{-\lambda x}dx = -\frac{1}{\lambda e^{-\lambda t}}(0 - e^{-\lambda t}) \text{ άρα } \mu_T(t) = \frac{1}{\lambda}, \lambda > 0.$$

Εάν $T \sim exp(\lambda)$ με $\lambda > 0$ τότε η μέση τιμή της κατανομής είναι $E(T) = \frac{1}{\lambda}$ και η διακύμανσή της είναι $Var(T) = \frac{1}{\lambda^2}$.

Η πιο βασική ιδιότητα της εκθετικής κατανομής είναι η ιδιότητα της "αμνησίας". Η λέξη αμνησία για λογικά όντα χρησιμοποιείται για να δηλώσει την

απώλεια μνήμης μερικώς ότι δηλαδή κάποιος δεν θυμάται ένα μέρος της ζωής του ή ολικώς, δηλαδή κάποιος δεν θυμάται τίποτε από τη ζωή του (Παπαϊωάννου - 1982, σελ.161).

Έστω τώρα ότι θέλουμε να βρούμε την κατανομή του χρόνου $T$, δοθέντος ότι ο χρόνος αυτός είναι μεγαλύτερος δεδομένου χρόνου $t_0$. Δηλαδή ζητάμε την κατανομή της τυχαίας μεταβλητής $T \mid T>t_0$.

Είναι:

$$P\left(T \le t \mid T > t_0\right) = \frac{P\left(t_0 < T \le t\right)}{P\left(T > t_0\right)} = \frac{F_T\left(t\right) - F_T\left(t_0\right)}{1 - P\left(T \le t_0\right)} = \frac{F_T\left(t\right) - F_T\left(t_0\right)}{1 - F_T\left(t_0\right)} =$$

$$= \frac{1 - e^{-\lambda t} - (1 - e^{-\lambda t_0})}{1 - (1 - e^{-\lambda t_0})} = \frac{e^{-\lambda t_0} - e^{-\lambda t}}{e^{-\lambda t_0}} = \frac{e^{-\lambda t_0}(1 - e^{-\lambda (t - t_0)})}{e^{-\lambda t_0}} = 1 - e^{-\lambda (t - t_0)} = 1 - e^{-\lambda \delta t},$$

αφού ισχύει $t = t_0 + \delta t$.

Παραγωγίζοντας ως προς $t$ βρίσκουμε ότι η πυκνότητα πιθανότητας (αποτυχίας) της $T \mid T>t_0$ είναι:

$$f_{T|T>t_0}(t) = \begin{cases} 0, & t \le t_0 \\ \lambda e^{-\lambda \delta t}, & t > t_0 \end{cases}$$

Βλέπουμε δηλαδή ότι ο επιπρόσθετος χρόνος έχει την ίδια κατανομή με τον αρχικό χρόνο, δηλαδή εκθετική με παράμετρο λ.

Έτσι θα ισχύει και η ακόλουθη σχέση: $P(T > t+t_0) = P(T > t_0)$. Με άλλα λόγια η πιθανότητα ο χρόνος ως την αποτυχία Τ που συνεχίζεται σε χρόνο $t_0$ να συνεχιστεί πέρα από το χρόνο $t+t_0$ είναι ανεξάρτητη της προηγούμενης διάρκειας. Αυτή είναι η ιδιότητα της απώλειας μνήμης της εκθετικής κατανομής. Αν ο χρόνος ζωής ενός ατόμου ή ενός εξαρτήματος μηχανής ακολουθεί την εκθετική κατανομή, τότε είναι σαν να θεωρούμε ότι το άτομο αυτό δεν "γερνάει" και εφ' όσον το άτομο ζει, έχει την ίδια πιθανότητα να αποσυντεθεί την επόμενη στιγμή. Προφανώς ισχύει και το αντίστροφο, δηλαδή ότι αν είναι γνωστό ότι ένα φαινόμενο έχει την ιδιότητα της αμνησίας τότε η κατανομή της διάρκειάς του πρέπει να είναι εκθετική.

Φυσικά, η παραπάνω υπόθεση δεν είναι ρεαλιστική για ζωντανούς οργανισμούς, αλλά είναι ρεαλιστική για τη μέτρηση του χρόνου ζωής ορισμένων μηχανικών εξαρτημάτων.

### 1.3.2 Η κατανομή Weibull – *Weib(r, λ)*

Η κατανομή Weibull είναι η γενίκευση της εκθετικής κατανομής. Σε αντίθεση όμως με την εκθετική δεν χαρακτηρίζεται από σταθερή συνάρτηση κινδύνου και γι' αυτό έχει ευρύτερες εφαρμογές.

Όταν ο χρόνος επιβίωσης *T* ακολουθεί την κατανομή Weibull με παραμέτρους *r* και *λ*, η συνάρτηση πυκνότητας είναι η $f_T(t) = r\lambda^r t^{r-1} e^{-(\lambda t)^r}$, $t \geq 0$ και η αθροιστική συνάρτηση κατανομής αποτυχίας είναι η $F_T(t) = 1 - e^{-(\lambda t)^r}$, $t \geq 0$. Έτσι η συνάρτηση επιβίωσης είναι η $S(t) = e^{-(\lambda t)^r}$ και η συνάρτηση κινδύνου θα είναι η $h_T(t) = r\lambda^r t^{r-1}$ για $t \geq 0$.

Όταν *r=1* η συνάρτηση βαθμού κινδύνου παραμένει σταθερή, δηλαδή η κατανομή Weibull γίνεται ίση με την εκθετική κατανομή με παράμετρο *λ*. Ο ρυθμός κινδύνου αυξάνεται όταν $r > 1$ και μειώνεται όταν $r < 1$, καθώς ο χρόνος *t* αυξάνεται.

Εάν $T \sim Weib(r, \lambda)$ με $r, \lambda > 0$ τότε η μέση τιμή της κατανομής είναι $E(T) = \frac{\Gamma(1+1/r)}{\lambda}$ και η διακύμανσή της είναι $Var(T) = \frac{1}{\lambda^2}\left[\Gamma\left(1+\frac{2}{r}\right) - \Gamma^2\left(1+\frac{1}{r}\right)\right]$, όπου *Γ(r)* με $r > 0$ είναι η συνάρτηση *Γ* που ορίζεται ως το γενικευμένο ολοκλήρωμα $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx = (r-1)!$

### 1.3.3 Η κατανομή Γάμμα– *G(α, β)*

Η κατανομή Γάμμα αποτελεί γενίκευση της εκθετικής και της $\chi^2$ κατανομής. Χαρακτηρίζεται από 2 παραμέτρους, $\alpha > 0$ και $\beta > 0$. Η συνάρτηση πυκνότητας είναι η $f_T(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$, $t > 0$ και η αθροιστική συνάρτηση κατανομής αποτυχίας είναι

$$F_T(t) = \int_0^t \frac{\beta^\alpha}{\Gamma(\alpha)} x^{a-1} e^{-\beta x} dx = \frac{1}{\Gamma(\alpha)} \int_0^{\beta t} y^{a-1} e^{-y} dy \,,\ t > 0.$$

Όταν *α=1* η κατανομή Γάμμα γίνεται η εκθετική κατανομή με παράμετρο *β*.

Όταν $a=\frac{n}{2}$ και $\beta=\frac{1}{2}$ η κατανομή Γάμμα γίνεται η $\chi^2$ με $n$ βαθμούς ελευθερίας.

Η συνάρτηση επιβίωσης είναι η $S(t)=\int_t^{\infty}\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{a-1}e^{-\beta x}dx$ .

Εάν $T \sim G(\alpha, \beta)$ με $\alpha, \beta > 0$ τότε η μέση τιμή της κατανομής είναι $E(T)=\frac{a}{\beta}$ και η διακύμανσή της είναι $Var(T)=\frac{\alpha}{\beta^2}$.

### 1.3.4 Η Λογαριθμοκανονική Κατανομή – $\Lambda(\mu, \sigma^2)$

Η Λογαριθμοκανονική κατανομή στην πιο απλή μορφή της μπορεί να οριστεί ως η κατανομή μιας μεταβλητής της οποίας ο λογάριθμος ακολουθεί την κανονική κατανομή. Ο χρόνος επιβίωσης $T$ ακολουθεί την Λογαριθμοκανονική κατανομή και συμβολίζεται ως $T \sim \Lambda(\mu, \sigma^2)$ όταν η μεταβλητή $X=lnT$ ακολουθεί την κανονική με μέσο $\mu$ και διακύμανση $\sigma^2$.

Όταν ο χρόνος επιβίωσης $T$ ακολουθεί την λογαριθμοκανονική κατανομή με παραμέτρους $\mu$ και $\sigma^2$, η συνάρτηση πυκνότητας είναι η

$$f_T(t)=\frac{1}{t\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(\ln t-\mu)^2\right],\ t>0,\ \mu\geq 0,\ \sigma>0$$

και η αθροιστική συνάρτηση κατανομής αποτυχίας είναι η

$$F_T(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^t\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}\left(\ln x-\mu^2\right)\right]dx\ ,\ t\geq 0.$$

Έτσι η συνάρτηση επιβίωσης είναι η

$$S(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_t^{\infty}\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}\left(\ln x-\mu^2\right)\right]dx\ .$$

Η συνάρτηση κινδύνου θα είναι η $h_T(t) = \dfrac{\dfrac{1}{t\sigma\sqrt{2\pi}}\exp\left[-\dfrac{(\ln at)^2}{2\sigma^2}\right]}{1-\Phi\left(\ln\dfrac{at}{\sigma}\right)}$, για $t > 0$,

όπου $\Phi(y)$ είναι η αθροιστική συνάρτηση της τυποποιημένης κανονικής κατανομής, δηλαδή $\Phi(y) = \dfrac{1}{\sqrt{2\pi}}\int_0^y \exp(-u^2/2)du$.

Η τιμή της παραμέρου $\sigma^2$ ελέγχει την ασυμμετρία της κατανομής: όσο μεγαλύτερη η τιμή του $\sigma^2$, τόσο μεγαλύτερη ασυμμετρία.

Η συνάρτηση κινδύνου αυξάνει αρχικά σε κάποιο μέγιστο και έπειτα μηδενίζεται καθώς ο χρόνος $t$ τείνει στο άπειρο.

Εάν $T \sim \Lambda(\mu, \sigma^2)$ τότε η μέση τιμή της κατανομής είναι $E(T) = \exp\left(\mu + \dfrac{\sigma^2}{2}\right)$

και η διακύμανση $Var(T) = \exp(\sigma^2 - 1)\exp(2\mu + \sigma^2)$.

Συχνά (σύμφωνα με τον Condon - 2001, σελ. 12) είναι χρήσιμο η παράμετρος $\sigma^2$ να αντικαθίσταται από την $\tau$, όπου $\tau = \dfrac{1}{\sigma^2}$, δηλαδή είναι η αντίστροφη τιμή της διακύμανσης. Έτσι όσο μεγαλύτερη τιμή έχει η $\tau$, οι παρατηρήσεις θα έχουν όλο και πιο μικρές αποκλίσεις από την μέση τιμή τους $\mu$.

### 1.3.5 Η Κατανομή Pareto – *Pareto(α,c)*

Μια συνεχής τυχαία μεταβλητή $T$ ακολουθεί την κατανομή Pareto με παραμέτρους $c > 0$ και $\alpha > 0$, εάν η συνάρτηση πυκνότητας πιθανότητας είναι η $f_T(t) = a\dfrac{c^\alpha}{t^{a+1}}$ με $t \geq c$. Η αθροιστική συνάρτηση κατανομής είναι η $F_T(t) = 1 - \left(\dfrac{c}{t}\right)^a$ και η συνάρτηση επιβίωσης $S(t) = \left(\dfrac{c}{t}\right)^a$.

Η συνάρτηση κινδύνου θα είναι $h_T(t) = a\dfrac{\dfrac{c^{\alpha}}{t^{a+1}}}{\left(\dfrac{c}{t}\right)^{a}} = a\dfrac{1}{t}$ και η μέση υπολοιπόμενη

ζωή $\mu_{\mathrm{T}}(t) = \dfrac{1}{\left(\dfrac{c}{t}\right)^{a}}\int_{t}^{\infty}\left(\dfrac{c}{t}\right)^{a} dx = t^{a}\int_{t}^{\infty} t^{-a}dx = t^{a}\,\dfrac{1}{a-1}t^{1-a} \Leftrightarrow \mu_{\mathrm{T}}(t) = \dfrac{t}{1-a}$ όπου $\alpha > 1$.

Εάν $T \sim Pareto(a, c)$ τότε η μέση τιμή της κατανομής θα είναι $E(T) = \dfrac{ac}{\alpha - 1}$ με

$\alpha > 1$ και η διακύμανση $Var(T) = \dfrac{\alpha c^{2}}{(\alpha-1)^{2}(\alpha-2)}$ με $\alpha > 2$.

### 1.3.6 Η Γεωμετρική Κατανομή – *Geo(p)*

Στην περίπτωση που ο χρόνος $T$ είναι μία διακριτή μεταβλητή ποσότητα αντί για την εκθετική κατανομή χρησιμοποιείται η Γεωμετρική κατανομή με παράμετρο $p$. Είναι η κατανομή του χρόνου αναμονής μέχρι την εμφάνιση της πρώτης "αποτυχίας" και $p$ είναι η πιθανότητα "αποτυχίας".

Η συνάρτηση πιθανότητας της κατανομής είναι η $P(T=t)=pq^{t}$ με $t = 0,1,2\ldots$ όπου $p+q=1,\ 0 \le p \le 1,\ 0 \le q \le 1$. Για διαδοχικές τιμές του $t$ οι παραπάνω πιθανότητες αποτελούν μία φθίνουσα γεωμετρική πρόοδο (γι' αυτό και η ονομασία γεωμετρική).

Η αθροιστική συνάρτηση κατανομής της Γεωμετρικής είναι

$$F_T(t) = \sum_{y=0}^{[t]} pq^{y} = p(1+q+q^{2}+\ldots q^{t}) = p\frac{1-q^{t+1}}{1-q}$$

Άρα $F_T(t) = 1-q^{t+1}$, όπου $t \ge 0$.

Έτσι η συνάρτηση επιβίωσης θα είναι $S(t) = q^{t+1}$ και η συνάρτηση κινδύνου

$$h_T(t) = \frac{P(T=t)}{P(T \ge t)} = \frac{pq^{t}}{\sum_{x=t}^{\infty} P(T=x)} = \frac{pq^{t}}{\sum_{x=t}^{\infty} pq^{x}} = \frac{pq^{t}}{1-\sum_{x=0}^{t-1} pq^{x}} = \frac{pq^{t}}{1-p\frac{1-q^{t}}{1-q}} = \frac{pq^{t}}{1-1+q^{t}} = p$$

Άρα $h_T(t) = p$, δηλαδή είναι σταθερή συνάρτηση.

Η μέση υπολοιπόμενη ζωή για τη διακριτή μεταβλητή $T$ είναι $\mu_{\mathrm{T}}(t)=\frac{1}{S(t)}\sum_{x=t}^{\infty}S(t)=\frac{1}{q^{t+1}}\sum_{x=t}^{\infty}q^{x+1}$ . Όμως:

$$\sum_{x=0}^{\infty}q^{x+1}=\frac{q}{1-q}=\sum_{x=0}^{t-1}q^{x+1}+\sum_{x=t}^{\infty}q^{x+1}\Leftrightarrow$$

$$\sum_{x=t}^{\infty}q^{x+1}=\frac{q}{1-q}-\sum_{x=0}^{t-1}q^{x+1}=\frac{q}{p}-q\frac{1-q^{t}}{1-q}=\frac{q}{p}(1-1+q^{t})=\frac{q^{t+1}}{p}$$

Άρα τελικά $\mu_{\mathrm{T}}(t)=\frac{1}{p}$.

Όπως και στην περίπτωση της εκθετικής κατανομής έτσι και στην περίπτωση της γεωμετρικής ισχύει η ιδιότητα της "αμνησίας".

$$P(T > t+t_0 \mid T > t_0)=\frac{P(T>t+t_0)}{P(T>t)}=\frac{S(t+t_0)}{S(t)}=\frac{q^{t+t_0+1}}{q^{t+1}}=q^{t_0}=P(T>t_0-1)=P(T\geq t_0).$$

Δηλαδή: $P(T > t+t_0 \mid T > t_0) = P(T \geq t_0)$.

Άρα η πιθανότητα ο χρόνος ως την αποτυχία $T$ που συνεχίζεται σε χρόνο $t_0$ να συνεχιστεί πέρα από το χρόνο $t+t_0$ είναι ανεξάρτητη της προηγούμενης διάρκειας. Αν ο διακριτός χρόνος ζωής ενός ατόμου ή ενός εξαρτήματος μηχανής ακολουθεί την γεωμετρική κατανομή τότε υποθέτουμε ότι αυτό δεν "γερνάει". Και σ' αυτή την περίπτωση ισχύει και το αντίστροφο, δηλαδή ότι αν είναι γνωστό ότι ένα φαινόμενο έχει την ιδιότητα της αμνησίας τότε η κατανομή της διάρκειάς του πρέπει να είναι γεωμετρική, εφ' όσον ο χρόνος είναι διακριτή μεταβλητή (Παπαϊωάννου - 1982, σελ.163).

Εάν $T \sim Geo(p)$ τότε η μέση τιμή της κατανομής είναι $E(T)=\frac{q}{p}$ και η διακύμανση $Var(T)=\frac{q}{p^2}$ με $t = 0,1,2,\ldots\ldots$ και $0 < p < 1$, $0 < q < 1$.

| **Κατανομή** | **Συνάρτηση Πιθ/τας $f_T(t)$** | **Αθροιστική Συνάρτηση $F_T(t)$** | **Συνάρτηση Επιβίωσης $S(t)$** | **Συνάρτηση Κινδύνου $h_T(t)$** | **Μέση τιμή $E(T)$** | **Διακύμανση $Var(T)$** |
|---|---|---|---|---|---|---|
| Εκθετική<br>$exp(\lambda)$, $\lambda > 0$ | $\lambda e^{-\lambda t}$ | $1-e^{-\lambda t}$ | $e^{-\lambda t}$ | $\lambda$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Weibull<br>$Weib(r,\lambda)$<br>$r, \lambda > 0$ | $r\lambda^r t^{r-1} e^{-(\lambda t)^r}$ | $1-e^{-(\lambda t)^r}$ | $e^{-(\lambda t)^r}$ | $r\lambda^r t^{r-1}$ | $\frac{\Gamma(1+1/r)}{\lambda}$ | $\frac{1}{\lambda^2}\left[\Gamma\left(1+\frac{2}{r}\right)-\Gamma^2\left(1+\frac{1}{r}\right)\right]$ |
| Γάμμα<br>$G(\alpha,\beta)$ $\alpha, \beta > 0$ | $\frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$ | $\int_0^t \frac{\beta^\alpha}{\Gamma(\alpha)} x^{a-1} e^{-\beta x} dx$ | $\int_t^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{a-1} e^{-\beta x} dx$ | $\frac{f_T(t)}{S(t)}$ | $\frac{a}{\beta}$ | $\frac{\alpha}{\beta^2}$ |
| Λογαριθμο-κανονική<br>$\Lambda(\mu, \sigma^2)$<br>$\mu \geq 0$, $\sigma > 0$ | $\frac{e^{-\frac{1}{2\sigma^2}\left(\ln t-\mu^2\right)}}{t\sigma\sqrt{2\pi}}$ | $\frac{\int_0^t \frac{1}{x} e^{-\frac{1}{2\sigma^2}\left(\ln x-\mu^2\right)} dx}{\sigma\sqrt{2\pi}}$ | $\frac{\int_t^\infty \frac{1}{x} e^{-\frac{1}{2\sigma^2}\left(\ln x-\mu^2\right)} dx}{\sigma\sqrt{2\pi}}$ | $\frac{\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln at)^2}{2\sigma^2}}}{1-G(\ln\frac{at}{\sigma})}$<br>όπου<br>$G(y)=\frac{\int_0^y e^{-u^2/2} du}{\sqrt{2\pi}}$ | $\exp\left(\mu+\frac{\sigma^2}{2}\right)$ | $\exp(2\mu+2\sigma^2-1)$ |
| Paretto<br>$Paretto(a, c)$<br>$a, c > 0$ | $a\frac{c^\alpha}{t^{a+1}}$ | $1-\left(\frac{c}{t}\right)^a$ | $\left(\frac{c}{t}\right)^a$ | $\frac{a}{t}$ | $\frac{ac}{\alpha-1}$,<br>$\alpha > 1$ | $\frac{\alpha c^2}{(\alpha-1)^2(\alpha-2)}$,<br>$\alpha > 2$ |
| Γεωμετρική<br>$Geo(p)$<br>$0 \leq p \leq 1$ | $pq^t$ | $1-q^{t+1}$ | $q^{t+1}$ | $p$ | $\frac{q}{p}$<br>$p \neq 0$ | $\frac{q}{p^2}$<br>$p \neq 0$ |

**Πίνακας 1.1:** Σύγκριση κατανομών για την ανάλυση δεδομένων επιβίωσης

# ΚΕΦΑΛΑΙΟ 2

## 2.1 ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ ΚΑΤΑ BAYES

### 2.1.1 Ιστορικά στοιχεία

Όπως ήδη αναφέρθηκε προηγουμένως τα τελευταία χρόνια με την εξέλιξη των ηλεκτρονικών υπολογιστών η εφαρμογή και η ανάπτυξη της Ανάλυσης Επιβίωσης γίνεται όλο και μεγαλύτερη. Συγκεκριμένες θεωρίες και μέθοδοι, οι οποίες είχαν πολύ δύσκολη εφαρμογή, λόγω των πολλών αριθμητικών πράξεων που χρειάζονταν, άρχισαν να εξελίσσονται κι αυτές σε πρακτικό, πλέον, επίπεδο με τη χρήση των στατιστικών προγραμμάτων και μία από αυτές είναι και η μέθοδος κατά Bayes.

Ο Thomas Bayes ήταν ένας Άγγλος κληρικός και μαθηματικός (σύμφωνα με τον Press - 1989, σελ. 15) που γεννήθηκε το 1702, όπως έχει υπολογιστεί, καθώς δεν είναι γνωστή η ακριβής ημερομηνία γέννησής του και έζησε ως τις 7 Απριλίου 1761, όπως επίσης δεν είναι γνωστές και πολλές λεπτομέρειες για την προσωπική του ζωή.

Η εργασία του Bayes αφορά τη στατιστική συμπερασματολογία για τις παραμέτρους της διωνυμικής κατανομής δεδομένων παρατηρήσεων προερχόμενων από την κατανομή αυτή. Η εργασία του αργότερα επεκτάθηκε και σε άλλες κατανομές, για παράδειγμα από τον LaPlace, το 1774 στην γενική του μορφή.

Η μέθοδος κατά Bayes προσφέρει πολλά πλεονεκτήματα.(Congdon - 2001, σελ.2). Σε σχέση με την κλασική στατιστική υπάρχει διαφορά, καταρχήν στο γεγονός ότι η παράμετρος, έστω $\theta$, ενός πληθυσμού δεν θεωρείται σταθερός αριθμός, αλλά τυχαία ποσότητα με κατανομή την $\pi(\theta)$, η οποία ονομάζεται εκ των προτέρων κατανομή (prior). Η ονομασία αυτή δίνεται, γιατί η κατανομή αυτή απεικονίζει τη γνώση που έχουμε για την παράμετρο του πληθυσμού πριν συλλέξουμε δεδομένα, δηλαδή βασίζεται σε πληροφορίες που μπορεί να έχουμε από προηγούμενες έρευνες ή στις πεποιθήσεις μας. Μετά τη συλλογή των δεδομένων ενός δείγματος και μέσω της συνάρτησης πιθανοφάνειας $L(\theta) = \prod_{i=1}^{n} f_T(t_i, \theta)$, όπου $t_i$, $(i = 1, 2 \ldots n)$ είναι οι χρόνοι επιβίωσης των ατόμων ή των εξαρτημάτων του δείγματος, λαμβάνεται η εκ των υστέρων κατανομή (posterior), η $g(\theta|t)$. Με άλλα λόγια η μέθοδος αυτή μας δίνει έναν τρόπο για το πώς τροποποιούνται οι υποκειμενικές πιθανότητες κάτω από το πρίσμα

νέας πληροφορίας ή των "αντικειμενικών" δεδομένων. Και λέγοντας υποκειμενική πιθανότητα εννοούμε την προσωπική θεώρηση του κάθε ατόμου που αφορά την αβεβαιότητα ενός γεγονότος, η οποία διαμορφώνεται ως πιθανότητα λαμβάνοντας υπόψη επιπλέον πληροφορίες (τα δεδομένα του πειράματος) που για κάθε άνθρωπο έχει διαφορετική ερμηνεία.

Δηλαδή με τη χρήση της μεθόδου βελτιώνεται η εκ των προτέρων γνώση χρησιμοποιώντας και την εκ των υστέρων γνώση των δεδομένων.

### 2.2 Κανόνας του Bayes

Έστω $H_1,\ldots,H_k$, $k$ ασυμβίβαστα μεταξύ τους ενδεχόμενα, το άθροισμα των οποίων καλύπτει όλο τον δειγματικό χώρο $\Omega$ (δηλ. $\Omega = H_1 \cup H_2 \cup \ldots \cup H_k$) και επίσης $P(H_i) > 0$ για κάθε $i = 1, 2, \ldots k$. Τότε για κάθε ενδεχόμενο $A$ με $P(A) > 0$ έχουμε: $P(H_i \mid A) = \dfrac{P(A \cap H_i)}{P(A)}$.

Όμως από το θεώρημα ολικής πιθανότητας ισχύει ότι για κάθε ενδεχόμενο $A$ έχουμε: $P(A) = \sum_{i=1}^{k} P(A \cap H_i) = \sum_{i=1}^{k} P(A \mid H_i) P(H_i)$

Άρα το θεώρημα Bayes γίνεται:

$$P(H_i \mid A) = \frac{P(A \mid H_i) P(H_i)}{\sum_{i=1}^{k} P(A \mid H_i) P(H_i)} \text{ για } i = 1, 2, \ldots 8$$

(βλ. Παπαϊωάννου - 1997, σελ.6)

Οι πιθανότητες $P(H_i)$ για $i = 1, 2, \ldots k$, ονομάζονται **εκ των προτέρων πιθανότητες (a priori)** και οι πιθανότητες $P(H_i \mid A)$, δηλαδή οι πιθανότητες μετά την εκτέλεση του πειράματος, ονομάζονται **εκ των υστέρων πιθανότητες (a pasteriori).**

### 2.3 Ο κανόνας του Bayes στην Ανάλυση Επιβίωσης

Συχνότερα, ενδιαφερόμαστε για την εκτίμηση της παραμέτρου $\theta$ ενός πληθυσμού με a priori συνάρτηση πιθανότητας $\pi(\theta)$. Ειδικά στην ανάλυση επιβίωσης, ενδιαφερόμαστε για την εκτίμηση της παραμέτρου $\theta$ της κατανομής που ακολουθούν οι

χρόνοι $t_i$ όπου $i = 1, 2,...n$. Η $\pi(\theta)$ είναι γνωστή από παλαιότερα αποτελέσματα ερευνών.

Έτσι, μπορεί να γίνει εφαρμογή του παραπάνω θεωρήματος, για πυκνότητες πιθανότητας, αντί για ενδεχόμενα.

Έστω $t_1,\ldots,t_n$ ένα τυχαίο δείγμα χρόνων επιβίωσης από έναν πληθυσμό με συνάρτηση πυκνότητας πιθανότητας $f_T(t|\theta)$, όπου $\theta$ είναι η παράμετρος της κατανομής του χρόνου $T$.

Για παράδειγμα εάν η κατανομή του χρόνου $T$ είναι η εκθετική τότε η παράμετρος $\theta$ αποτελείται από μία συνιστώσα και είναι $\theta = \lambda$. Εάν η κατανομή του χρόνου είναι η Weibul, η $\theta$ θα αποτελείται από δύο συνιστώσες, την $r$ και την $\lambda$ δηλαδή $\underset{\sim}{\theta} = \{r, \lambda\}$.

Αφού η παράμετρος θεωρείται τυχαία ποσότητα θα έχει συνάρτηση πυκνότητας πιθανότητας, όπως είδαμε, την $\pi(\theta)$. Έστω $g(\theta \mid t)$ η πυκνότητα πιθανότητας της παραμέτρου $\theta$, μετά την εφαρμογή των δεδομένων, δηλαδή η εκ των υστέρων πιθανότητα.

Το θεώρημα του Bayes είναι το εξής (βλ. Press - 1989, σελ.24 & 39):

α) Αν η παράμετρος είναι διακριτή μεταβλητή

$$g(\theta \mid t) = \frac{f(t \mid \theta)\pi(\theta)}{f(t)} = \frac{f(t \mid \theta)\pi(\theta)}{\sum_{\theta} f(t \mid \theta)\pi(\theta)}$$

β) Αν η παράμετρος είναι συνεχής μεταβλητή

$$g(\theta \mid t) = \frac{f(t \mid \theta)\pi(\theta)}{f(t)} = \frac{f(t \mid \theta)\pi(\theta)}{\int_{\theta} f(t \mid \theta)\pi(\theta)d\theta}$$

Δηλαδή η εκ των υστέρων κατανομή της παραμέτρου $\theta$, η $g(\theta \mid t)$, είναι ανάλογη με το γινόμενο της συνάρτησης πιθανοφάνειας $f(t \mid \theta) = L(\theta) = \prod_{i=1}^{n} f_T(t_i, \theta)$ και της εκ των προτέρων κατανομής της παραμέτρου $\theta$, της $\pi(\theta)$.

Στην περίπτωση που η παράμετρος $\theta$ αποτελείται από περισσότερες από μία συνιστώσες, έστω $k$, δηλαδή $\underset{\sim}{\theta} = \{\theta_1, \theta_2, \ldots \theta_k\}$ οι οποίες είναι και ανεξάρτητες μεταξύ τους ανά 2, η εκ των προτέρων κατανομή της $\underset{\sim}{\theta}$ θα είναι η $\pi(\underset{\sim}{\theta}) = \pi(\theta_1, \theta_2 \ldots, \theta_k) = \pi_1(\theta_1)\pi_2(\theta_2)\ldots\pi_k(\theta_k)$, όπου $\pi_1(\theta_1), \pi_2(\theta_2), \ldots, \pi_k(\theta_k)$ είναι οι περιθώριες κατανομές των συνιστωσών $\theta_1, \theta_2, \ldots \theta_k$ αντίστοιχα και έτσι η εκ των υστέρων κατανομή του $\underset{\sim}{\theta}$ γίνεται:

α) Αν $\underset{\sim}{\theta}$ διακριτή μεταβλητή:

$$g(\underset{\sim}{\theta} \mid t) = \frac{f(t \mid \underset{\sim}{\theta})\pi(\theta_1, \theta_2 \ldots \theta_k)}{\sum_{\theta_1}\sum_{\theta_2}\ldots\sum_{\theta_k} f(t \mid \underset{\sim}{\theta})\pi(\theta_1\theta_2 \ldots \theta_k)} =$$

$$\frac{f(t \mid \underset{\sim}{\theta})\pi_1(\theta_1)\pi_2(\theta_2)\ldots\pi_k(\theta_k)}{\sum_{\theta_1}\sum_{\theta_2}\ldots\sum_{\theta_k} f(t \mid \underset{\sim}{\theta})\pi_1(\theta_1)\pi_2(\theta_2)\ldots\pi_k(\theta_k)}$$

β) Αν $\underset{\sim}{\theta}$ συνεχής μεταβλητή:

$$g(\underset{\sim}{\theta} \mid t) = \frac{f(t \mid \underset{\sim}{\theta})\pi(\theta_1, \theta_2 \ldots \theta_k)}{\int_{\theta_1}\int_{\theta_2}\ldots\int_{\theta_k} f(t \mid \underset{\sim}{\theta})\pi(\theta_1\theta_2 \ldots \theta_k)d\theta_1 d\theta_2 \ldots d\theta_k} =$$

$$\frac{f(t \mid \underset{\sim}{\theta})\pi_1(\theta_1)\pi_2(\theta_2)\ldots\pi_k(\theta_k)}{\int_{\theta_1}\int_{\theta_2}\ldots\int_{\theta_k} f(t \mid \underset{\sim}{\theta})\pi_1(\theta_1)\pi_2(\theta_2)\ldots\pi_k(\theta_k)d\theta_1 d\theta_2 \ldots d\theta_k}$$

Η παραπάνω κατανομή $g(\underset{\sim}{\theta} \mid t)$ ουσιαστικά είναι η από κοινού εκ των υστέρων κατανομή των $\theta_1, \theta_2 \ldots \theta_k$.

Έστω $g_i(\theta_i \mid t)$ οι περιθώριες εκ των υστέρων κατανομές των συνιστωσών $\theta_i$, για $i = 1, 2, \ldots k$.

α) Στη διακριτή περίπτωση θα είναι:

$$g_i(\theta_i \mid t) = \sum_{\theta_1} \cdots \sum_{\theta_{i-1}} \sum_{\theta_{i+1}} \cdots \sum_{\theta_k} g(\underset{\sim}{\theta} \mid t)$$

β) Στη συνεχή περίπτωση θα είναι:

$$g_i(\theta_i \mid t) = \int_{\theta_1} \cdots \int_{\theta_{i-1}} \int_{\theta_{i+1}} \cdots \int_{\theta_k} g(\underset{\sim}{\theta} \mid t) d\theta_1 \ldots d\theta_{i-1} d\theta_{i+1} \ldots d\theta_k$$

### 2.4 Εφαρμογές του Θεωρήματος Bayes

Χρησιμοποιώντας το παραπάνω θεώρημα μπορεί να γίνει έλεγχος υποθέσεων ή να υπολογιστούν διαστήματα εμπιστοσύνης για τις παραμέτρους ενός πληθυσμού.

- Έστω ότι η παράμετρος $\theta$ είναι μία $k$ -διάστατη τυχαία μεταβλητή δηλαδή $\underset{\sim}{\theta} = \{\theta_1, \theta_2, \ldots \theta_k\}$. Εφόσον η $\theta$ είναι μία τυχαία μεταβλητή για την αναμενόμενη τιμή και για τη διακύμανσή της θα ισχύουν:

  $E(\underset{\sim}{\theta}) = \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_k} \underset{\sim}{\theta} g(\underset{\sim}{\theta} \mid t)$ και

  $Var(\underset{\sim}{\theta}) = E\left(\underset{\sim}{\theta} - E(\underset{\sim}{\theta})\right)^2 = = \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_k} \left(\underset{\sim}{\theta} - E(\underset{\sim}{\theta})\right)^2 g(\underset{\sim}{\theta} \mid t)$, αν η εκ των υστέρων κατανομή $g(\theta \mid t)$ είναι διακριτή,

  ενώ $E(\underset{\sim}{\theta}) = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_k} \underset{\sim}{\theta} g(\underset{\sim}{\theta} \mid t) d\theta_1 d\theta_2 \ldots d\theta_k$ και

  $Var(\underset{\sim}{\theta}) = E\left(\underset{\sim}{\theta} - E(\underset{\sim}{\theta})\right)^2 = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_k} \left(\underset{\sim}{\theta} - E(\underset{\sim}{\theta})\right)^2 g(\underset{\sim}{\theta} \mid t) d\theta_1 d\theta_2 \ldots d\theta_k$, αν η $g(\theta \mid t)$ είναι συνεχής κατανομή.

- Εκτός από την αναμενόμενη τιμή $E(\underset{\sim}{\theta})$ και τη διακύμανση $Var(\underset{\sim}{\theta})$ μπορούν να βρεθούν και αλλά μέτρα θέσης και διασποράς, όπως είναι η διάμεσος, τα εκατοστιαία σημεία, η επικρατούσα τιμή ή η τυπική απόκλιση. Π.χ. για το $p$ - ποσοστιαίο σημείο $\theta_p$, με $0 \le p \le 1$, ισχύει $P(\theta \le \theta_p) = G(\theta_p \mid t) = p$, αν η $\theta$ συνεχής, όπου η $G(\theta_p \mid t) = \int_{\theta} g(\theta \mid t) d\theta$ είναι η αθροιστική συνάρτηση κατανομής της $g(\theta \mid t)$, ενώ αν η $\theta$ είναι διακριτή θα ισχύει: $P(\theta \le \theta_p) \ge p$ και $P(\theta \ge \theta_p) \ge 1 - p$.

- Σύμφωνα με τους Carlin και Louis (1996, σελ. 42), για την παράμετρο $\theta$ μπορούμε να βρούμε εκ των υστέρων διαστήματα αξιοπιστίας, τα οποία θα είναι της μορφής $(L, U)$. Το "ποσοστό φορών" που ένα διάστημα $(L, U)$ θα περιέχει το $\theta$ θα είναι ο βαθμός αξιοπιστίας $100(1 - \alpha)\%$.
  Δηλαδή: *βαθμός αξιοπιστίας* $= P(L < \theta < U) = 1 - \alpha = \int_{\theta} g(\theta \mid t) d\theta$.

- Επίσης σύμφωνα με τον Congdon (2001, σελ. 470) μπορεί να χρησιμοποιηθεί και η μέθοδος, με την οποία για κάθε μοντέλο λαμβάνεται υπόψη κι ένας συντελεστής βαρύτητας $w_k = Pr(\text{το κατάλληλο μοντέλο είναι το } M_k \mid t)$, όπου $k = 0, 1$ και να σχηματιστεί ένας συνολικός μέσος για την παράμετρο $\theta$, συνδυάζοντας εκτιμήσεις από διαφορετικά μοντέλα, σύμφωνα με τους παραπάνω συντελεστές βαρύτητας. Θα είναι:
  $$w_k = P(M_k \mid t) = \frac{P(t \mid M_k) P(M_k)}{P(t)}$$
  κι αν οι εκ των προτέρων είναι ίσες για κάθε μοντέλο, τότε θα ισχύει η παρακάτω ισότητα:
  $$w_k = \frac{P(t \mid M_k) P(M_k)}{P(t \mid M_0) P(M_0) + P(t \mid M_1) P(M_1)} = \frac{P(t \mid M_k)}{P(t \mid M_0) + P(t \mid M_1)}, \; k = 0, 1$$

Το θεώρημα Bayes χρησιμοποιείται και για προβλέψεις μελλοντικών παρατηρήσεων και για την επίδρασή τους πάνω στις a priori κατανομές. Αν έστω $t^*$ η μελλοντική παρατήρηση, η τιμή της θα βασίζεται κυρίως στην πυκνότητα πιθανότητας $f_T(t^* \mid \underset{\sim}{t}, \underset{\sim}{\theta})$, όπου $\underset{\sim}{\theta}$ η παράμετρος της κατανομής των χρόνων $t_i$.

**2.5 Παράγοντας Bayes**

Αν τώρα θέλουμε να ελέγξουμε τις υποθέσεις για το διάστημα στο οποίο βρίσκεται η τιμή της παραμέτρου *θ,* δηλαδή τη μηδενική υπόθεση $H_0$: $\theta \in (\alpha_{0,} b_0)$ έναντι της εναλλακτικής $H_1$: $\theta \in (\alpha_{1,} b_1)$ με $(\alpha_{0,} b_0) \cap (\alpha_{1,} b_1) = \varnothing$, το θεώρημα του Bayes χρησιμοποιείται ως εξής (Congdon - 2001, σελ.15):

Έστω $P(H_0)$ η εκ των προτέρων γνώση για την ισχύ της μηδενικής υπόθεσης $H_0$ και $P(H_1)$ η αντίστοιχη εκ των προτέρων γνώση για την εναλλακτική υπόθεση $H_1$ με $P(H_0) + P(H_1) = 1$. Η εκ των υστέρων γνώση για την $H_{0,}$ δοθέντων των δεδομένων χρόνων επιβίωσης $\underset{\sim}{t} = (t_1, \ldots, t_n)$ θα είναι:

$P(H_0 \mid \underset{\sim}{t}) = \dfrac{P(\underset{\sim}{t} \mid H_0) P(H_0)}{P(\underset{\sim}{t})}$, που σύμφωνα με το θεώρημα ολικής πιθανότητας θα ισχύει:

$$P(H_0 \mid \underset{\sim}{t}) = \frac{P(\underset{\sim}{t} \mid H_0) P(H_0)}{P(\underset{\sim}{t} \mid H_0) P(H_0) + P(\underset{\sim}{t} \mid H_1) P(H_1)}$$

Αντίστοιχα για την εναλλακτική υπόθεση έχουμε:

$$P(H_1 \mid \underset{\sim}{t}) = \frac{P(\underset{\sim}{t} \mid H_1) P(H_0)}{P(\underset{\sim}{t} \mid H_0) P(H_0) + P(\underset{\sim}{t} \mid H_1) P(H_1)}$$

Σχηματίζοντας το λόγο των εκ των υστέρων πιθανοτήτων έχουμε:

$$\frac{P(H_0 \mid \underset{\sim}{t})}{P(H_1 \mid \underset{\sim}{t})} = \frac{P(H_0)}{P(H_1)} \cdot \frac{P(\underset{\sim}{t} \mid H_0)}{P(\underset{\sim}{t} \mid H_1)},$$

όπου ο λόγος $\dfrac{P(\underset{\sim}{t} \mid H_0)}{P(\underset{\sim}{t} \mid H_1)}$ ονομάζεται **παράγοντας Bayes** της υπόθεσης $H_0$ έναντι της υπόθεσης $H_1$ και συμβολίζεται με $\boldsymbol{B_{01}}$. Χρησιμοποιήθηκε για πρώτη φορά από τον Jeffreys το 1935, όπως αναφέρει ο Raftery (1995)

Εάν $B_{01} > 1$ τότε δεν απορρίπτουμε την μηδενική υπόθεση $H_0$, ενώ αν $B_{01} \leq 1$ η $H_0$ απορρίπτεται.

Εάν η εναλλακτική υπόθεση καλύπτει όλες τις δυνατές τιμές για την παράμετρο *θ,* τότε θα ισχύει επίσης ότι $P(H_0 \mid \underset{\sim}{t}) + P(H_1 \mid \underset{\sim}{t}) = 1$.

Έτσι η εκ των υστέρων πιθανότητα για την $H_0$ θα υπολογίζεται από την αντίστοιχη εκ των προτέρων και τον παράγοντα Bayes και θα είναι:

$$\frac{P(H_0 \mid \underset{\sim}{t})}{P(H_1 \mid \underset{\sim}{t})} = \frac{P(H_0)}{P(H_1)} B_{01} \Leftrightarrow \frac{P(H_0 \mid \underset{\sim}{t})}{1 - P(H_0 \mid \underset{\sim}{t})} = \frac{P(H_0)}{P(H_1)} B_{01} \Leftrightarrow$$

$$P(H_0 \mid \underset{\sim}{t}) = \frac{P(H_0)}{P(H_1)} B_{01} - P(H_0 \mid \underset{\sim}{t}) \frac{P(H_0)}{P(H_1)} B_{01} \Leftrightarrow$$

$$P(H_0 \mid \underset{\sim}{t}) \left[ 1 + \frac{P(H_0)}{P(H_1)} B_{01} \right] = \frac{P(H_0)}{P(H_1)} B_{01} \Leftrightarrow$$

$$P(H_0 \mid \underset{\sim}{t}) = \frac{\frac{P(H_0)}{P(H_1)} B_{01}}{\left[ 1 + \frac{P(H_0)}{P(H_1)} B_{01} \right]} \Leftrightarrow$$

$$P(H_0 \mid \underset{\sim}{t}) = \frac{1}{\left[ \frac{1}{B_{01}} \frac{P(H_1)}{P(H_0)} + 1 \right]}$$

**2.6 Σύγκριση μοντέλων**

Ακριβώς η ίδια διαδικασία ακολουθείται όταν θέλουμε να ελέγξουμε τις υποθέσεις για την ισχύ δύο διαφορετικών μοντέλων $M_0$ ή $M_1$. (Raftery - 1995)

Δηλαδή:

$H_0$: Ισχύει το μοντέλο $M_0$ με συνάρτηση πιθανοφάνειας: $f_T(\underset{\sim}{t} \mid \underset{\sim}{\theta}_0)$

$H_1$: Ισχύει το μοντέλο $M_1$ με συνάρτηση πιθανοφάνειας: $f_T(\underset{\sim}{t} \mid \underset{\sim}{\theta}_1)$

Προφανώς χρειάζονται εκ των προτέρων κατανομές για το σύνολο των παραμέτρων $\underset{\sim}{\theta}_0$ και $\underset{\sim}{\theta}_1$, όπως και για τα μοντέλα $M_0$ και $M_1$.

Ο παράγοντας Bayes $\frac{P(\underset{\sim}{t} \mid M_0)}{P(\underset{\sim}{t} \mid M_1)} = \frac{P(M_0 \mid \underset{\sim}{t}) / P(M_0)}{P(M_1 \mid \underset{\sim}{t}) / P(M_1)}$ είναι πολύ χρήσιμος, αφού επιτρέπει συγκρίσεις 2 μοντέλων $M_0$ και $M_1$, όχι κατ' ανάγκη "φωλιασμένων", δηλαδή τα μοντέλα που συγκρίνονται δεν χρειάζεται να έχουν τις ίδιες ανεξάρτητες μεταβλητές.

Στην περίπτωση που τα μοντέλα είναι "φωλιασμένα" τότε $P(M_0) = P(M_1) = 0.5$ και τότε ο παράγοντας Bayes $B_{01}$ συγκλίνει προς το λόγο $\frac{P(M_0 \mid \underset{\sim}{t})}{P(M_1 \mid \underset{\sim}{t})}$.

Σύμφωνα με τον παρακάτω πίνακα αποφασίζουμε για το καταλληλότερο μοντέλο (Congdon - 2001, σελ.471):

| **Τιμή $B_{01}$** | **Απόφαση** |
|---|---|
| < 1 | Υπέρ $M_1$ |
| 1 - 3 | Υπέρ $M_0$ (χωρίς ισχυρές ενδείξεις) |
| 3 - 20 | Υπέρ $M_0$ |
| 20 – 150 | Υπέρ $M_0$ (ισχυρές ενδείξεις) |
| > 150 | Υπέρ $M_0$ (πολύ ισχυρές ενδείξεις) |

**Πίνακας 2.1:** Παράγοντας Bayes

### 2.7 Τα κριτήρια πληροφορίας του Akaike και του Bayes (Bayes & Akaike 's Information Criterion)

Ένα ερώτημα που θα μπορούσε να προκύψει είναι εάν ο παράγοντας Bayes υπολογίζεται ανεξάρτητα από τις εκ των προτέρων πιθανότητες των $\underset{\sim}{\theta}_m$ $(m = 0, 1)$ στα μοντέλα $M_0$ και $M_1$, διότι οι υπολογισμοί είναι αρκετά δύσκολοι. Απάντηση σ' αυτό το ερώτημα δόθηκε από τον Schwarz το 1978, σύμφωνα με τους Carlin & Louis (1996 – κεφ.2)

Έστω ότι $\underset{\sim}{\theta}^*$ είναι η εκτίμηση του μέσου των παραμέτρων σ' ένα μοντέλο $M$ και υποθέτουμε ότι η εκτίμηση κατά Bayes είναι περίπου ίση με τον εκτιμητή μεγίστης πιθανοφάνειας (MLS), τότε ο παρακάτω λόγος είναι ο λόγος των πιθανοφανειών:

$\Lambda = -2\ln\frac{L(\underset{\sim}{t} \mid \underset{\sim}{\theta}_0^*)}{L(\underset{\sim}{t} \mid \underset{\sim}{\theta}_1^*)}$, όπου $L(\underset{\sim}{t} \mid \underset{\sim}{\theta}_0^*)$ είναι η τιμή ένας συνάρτησης πιθανοφάνειας για το πιο απλό μοντέλο $M_0$ και $L(\underset{\sim}{t} \mid \underset{\sim}{\theta}_1^*)$ είναι η αντίστοιχη τιμή ένας συνάρτησης πιθανοφάνειας για ένα πιο σύνθετο μοντέλο $M_1$. Το στατιστικό αυτό ευνοεί το σύνθετο μοντέλο $M_1$ και ακολουθεί κατανομή $\chi^2$ με βαθμούς ελευθερίας $p_1 - p_0$, όπου $p_0$, $p_1$ το πλήθος των παραμέτρων των μοντέλων $M_0$ και $M_1$ αντίστοιχα.

Στην παρουσίαση του με τίτλο "Bayesian Hypothesis Testing and Bayes Factors" ο Jeff Grinavisky (University of Chicago) σημειώνει πως, σύμφωνα με τον Gill (2002), για μεγάλο μέγεθος δείγματος *n,* ακόμα καλύτερη αποτελέσματα προκύπτουν όταν χρησιμοποιείται το *AIC*, το οποίο προκύπτει από τα αρχικά των λέξεων Akaike `s Information Criterion και είναι:

$$AIC = -2\ln L\left(t \mid \underset{\sim}{\theta}^*\right) + 2p\,,$$

όπου $L\left(t \mid \underset{\sim}{\theta}^*\right)$ είναι η συνάρτηση πιθανοφάνειας για ένα μοντέλο *M* και *p* ο αριθμός των παραμέτρων του.

Καλύτερο είναι το μοντέλο με το μικρότερο *AIC*. Το *AIC* δείχνει μεροληψία προς τα πιο πολύπλοκα μοντέλα, γιατί η συνάρτηση πιθανοφάνειας τείνει να αυξάνεται γρηγορότερα απ' ότι ο αριθμός των παραμέτρων.

Έτσι για καλύτερα αποτελέσματα χρησιμοποιείται το Bayesian Information Criterion (*BIC*) το οποίο είναι:

$$BIC = -2\ln L\left(t \mid \underset{\sim}{\theta}^*\right) + 2p\ln n\,,$$

όπου *n* είναι το μέγεθος του δείγματος και μπορεί να εφαρμοστεί και για μη "φωλιασμένα" μοντέλα. Η προσέγγιση του παράγοντα Bayes είναι το $\exp\left(-\frac{1}{2}BIC\right)$, που δεν εξαρτάται από ένας εκ των προτέρων πιθανότητες των $\underset{\sim}{\theta}$ (Carlin & Louis – 1996, σελ.48).

Στη συνέχεια ακολουθεί ένας πίνακας, σύμφωνα με τον οποίο αποφασίζουμε για το καλύτερο μοντέλο, ανάμεσα στα $M_0$ και $M_1$, κρίνοντας από τη διαφορά των *BICs* και την εκ των υστέρων πιθανότητα του κάθε μοντέλου.

Σύμφωνα με τον Raftery (1995), ισχύει ότι η διαφορά των *BICs* για 2 μοντέλα $M_0$ και $M_1$ είναι $BIC_1 - BIC_0 = 2lnB_{01}$, όπου $B_{01}$ είναι ο παράγοντας Bayes του μοντέλου $M_0$ έναντι του μοντέλου $M_1$. Έτσι ο πίνακας απόφασης είναι ο ακόλουθος:

| **Διαφορά *BICs*** $BIC_1 - BIC_0 = 2lnB_{01}$ | **Εκ των υστέρων πιθανότητα** $P(M_0 \mid \underset{\sim}{t})$ | **Απόφαση** |
|---|---|---|
| < 0 | < 50 | Υπέρ $M_1$ |
| 0 -2 | 50 – 75 | Υπέρ $M_0$ (χωρίς ισχυρές ενδείξεις) |
| 2 -6 | 75 – 95 | Υπέρ $M_0$ |
| 6 -10 | 95 – 99 | Υπέρ $M_0$ (ισχυρές ενδείξεις) |
| >10 | > 99 | Υπέρ $M_0$ (πολύ ισχυρές ενδείξεις) |

**Πίνακας 2.2:** Σύγκριση μοντέλων με τη χρήση των διαφορών *BICs*

### 2.8 Το κριτήριο πληροφορίας διασποράς (Deviance Information Criterion)

Ένα άλλο στατιστικό το οποίο αναφέρεται από τους δημιουργούς του WINBUGS (Spiegelhalter et. al – 2004), το οποίο μπορεί να υπολογιστεί στο WINBUGS, είναι το κριτήριο πληροφορίας διασποράς (Deviance Information Criterion - *DIC)*

Το *DIC* είναι ίσο με *Dbar* + $p_D$, όπου:

***Dbar****:* η εκ των υστέρων μέση τιμή της συνάρτησης πιθανοφάνειας υπολογισμένη με τα τιμές των παραμέτρων σε κάθε επανάληψη του αλγορίθμου MCMC.

$\boldsymbol{p_D}$ = *Dbar – Dhat* ο αριθμός των αποτελεσματικών (effective) ή ουσιαστικών παραμέτρων

***Dhat****:* η συνάρτηση πιθανοφάνειας υπολογισμένη χρησιμοποιώντας τους εκ των υστέρων μέσους των παραμέτρων $\underset{\sim}{\theta}$ (ή άλλο μέτρο κεντρικής θέσης).

Το *DIC* είναι η γενίκευση του κριτηρίου *AIC*. Όμως, ενώ στο *AIC* και στο *BIC* χρησιμοποιείται ο πραγματικός αριθμός των παραμέτρων του μοντέλου, στο *DIC* χρησιμοποιείται ο αριθμός των αποτελεσματικών / ουσιαστικών (effective) παραμέτρων, εκτός αν έχουμε μοντέλα με μικρή prior πληροφορία.

Το κριτήριο *BIC* προσπαθεί να αναγνωρίσει το πραγματικό μοντέλο. Αντίθετα το κριτήριο *DIC,* δεν βασίζεται σε υποθέσεις για το πραγματικό μοντέλο και γι' αυτό διαφέρει από το *BIC*.

Το *DIC* επιλέγει ως καταλληλότερο το μοντέλο το οποίο κάνει τις καλύτερες προβλέψεις. Ως "καλύτερο" επιλέγουμε το μοντέλο με το μικρότερο *DIC*, εφ' όσον τα

δεδομένα είναι τα ίδια ακριβώς σε όλα τα μοντέλα, αλλιώς δεν έχει νόημα να γίνει σύγκριση των *DICs*.

# ΚΕΦΑΛΑΙΟ 3

## 3.1 ΜΗ ΠΑΡΑΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ

Το κύριο αντικείμενο ενδιαφέροντος στην Ανάλυση Επιβίωσης είναι η εκτίμηση και περιγραφή των κατανομών του χρόνου ζωής ενός ατόμου ή ενός εξαρτήματος. Το ενδιαφέρον της έρευνας επικεντρώνεται στην εκτίμηση της συνάρτησης πυκνότητας πιθανότητας $f_T(t)$, της συνάρτησης επιβίωσης $S(t)$, της συνάρτησης κινδύνου $h_T(t)$, αλλά και στην εκτίμηση των παραμέτρων $\underset{\sim}{\theta} = \{\theta_1, \ldots, \theta_k\}$ που εμφανίζονται σε κάθε κατανομή και σε παραμέτρους όπως είναι η μέση τιμή, η διάμεσος κ.α.

Οι μη παραμετρικές μέθοδοι εκτίμησης είναι βέβαια λιγότερο αποτελεσματικές από τις παραμετρικές μεθόδους, όταν οι χρόνοι επιβίωσης είναι γνωστό ότι ακολουθούν μία συγκεκριμένη κατανομή, αλλά πολύ πιο κατάλληλες όταν η θεωρητική κατανομή των χρόνων επιβίωσης δεν είναι γνωστή, που είναι και η πιο συνηθισμένη περίπτωση.

Από τις συναρτήσεις η πιο ενδιαφέρουσα για μελέτη είναι η συνάρτηση επιβίωσης $S(t)$, καθώς είναι αυτή που εκφράζει την πιθανότητα το άτομο ή το εξάρτημα που εξετάζεται να μην έχει αποτύχει μέχρι την χρονική στιγμή $t$. Όμως, από την εκτίμηση της $S(t)$ προκύπτουν εκτιμήσεις για την πυκνότητα αποτυχίας $f_T(t)$, αλλά και για τη συνάρτηση κινδύνου $h_T(t)$.

### 3.2 Μέθοδος πίνακα επιβίωσης (Life table method)

Στην ενότητα αυτή θα ασχοληθούμε με μία από τις πιο παλιές τεχνικές για τη μέτρηση της θνησιμότητας και της επιβίωσης ενός πληθυσμού. Χρησιμοποιήθηκε σε ιατρικές έρευνες, μελέτες ανάπτυξης πληθυσμών, μεταναστεύσεων, κ.α. Αποτελείται από πίνακες, οι οποίοι συνοψίζουν τις προηγούμενες εμπειρίες ενός πληθυσμού για μια συγκεκριμένη περίοδο (συνήθως δεκαετία) και εκδίδονται από τις κρατικές υπηρεσίες, βασισμένοι σε δεδομένα από απογραφές. Στο βιβλίο του ο Congdon (2001) περιγράφει τη χρήση αυτών των πινάκων και μεταξύ άλλων τονίζεται η ύπαρξη δύο ειδών τέτοιων πινάκων. Οι πίνακες που περιλαμβάνουν δεδομένα για τη θνησιμότητα ενός

συγκεκριμένου πληθυσμού για μια συγκεκριμένη χρονική περίοδο (population life – table) και οι πίνακες που περιλαμβάνουν ιστορικά δεδομένα για μια συγκεκριμένη ασθένεια, τα συμπτώματα της οποίας έχουν παρακολουθηθεί για κάποιο χρονικό διάστημα μέσα από κλινικές έρευνες (clinical life -tables).

Με τη χρήση των πινάκων γίνονται εκτιμήσεις για τη συνάρτηση επιβίωσης του πληθυσμού που μελετάται, όπως και για τη πυκνότητα πιθανότητας ή για τη συνάρτηση κινδύνου, σύμφωνα με τον Gehan (1969), ο οποίος χρησιμοποίησε και τύπους για την εύρεση διαστημάτων εμπιστοσύνης των καμπυλών επιβίωσης. Επίσης πολύ ενδιαφέροντα στη μελέτη των πληθυσμών είναι τα γραφήματα, καθώς κατασκευάζονται από αρκετά μεγάλο πλήθος παρατηρήσεων.

### 3.3 Εκτίμηση της συνάρτησης επιβίωσης

Στην πιο απλή περίπτωση εκτίμησης της συνάρτηση επιβίωσης όλοι οι ασθενείς παρακολουθούνται μέχρι το θάνατο και προφανώς όλοι οι χρόνοι επιβίωσης είναι γνωστοί. Έστω ότι *n* άτομα ή αντικείμενα παίρνουν μέρος στην έρευνα, άρα έχουμε *n* χρόνους τους οποίους διατάσσουμε κατά αύξουσα σειρά: $t_{(1)} \le t_{(2)} \le t_{(3)} \le \ldots\ldots \le t_{(n)}$

Εάν *i* είναι το πλήθος των μονάδων του δείγματος που έζησαν λιγότερο από τον χρόνο $t_{(i)}$, τότε $n - i_{)}$ θα είναι το πλήθος αυτών που έζησαν περισσότερο από $t_{(i)}$, άρα η εκτιμώμενη συνάρτηση επιβίωσης για το χρόνο $t_{(i)}$ θα είναι $\hat{S}(t_{(i)}) = \frac{n-i}{n}$. Στην περίπτωση που δύο ή περισσότεροι χρόνοι επιβίωσης είναι ίσοι π.χ. $t_{(k)} = t_{(l)}$ με $k < l$, θα είναι $\hat{S}(t_{(k)}) = \hat{S}(t_{(l)}) = \frac{n-l}{n}$. Επίσης θα ισχύει ότι $\hat{S}(t_{(\min)}) = \hat{S}(t_{(0)}) = 1$ και $\hat{S}(t_{(\max)}) = \hat{S}(t_{(n)}) = 0$.

### 3.4 Η μέθοδος Kaplan – Meier

Η μέθοδος Kaplan – Meier (ή Product Limit Method) χρησιμοποιείται για την εκτίμηση της συνάρτησης επιβίωσης και λαμβάνονται υπόψη οι ακριβείς χρόνοι επιβίωσης για κάθε άτομο (ή εξάρτημα).

Με τη μέθοδο αυτή μελετάται η πιθανότητα ένα άτομο να επιβιώσει, έστω *M ( ≥ 2)* χρόνια (ή μήνες ή εβδομάδες, ανάλογα με τη μονάδα μέτρησης του χρόνου στο εκάστοτε πείραμα) και ισχύει ότι η πιθανότητα αυτή ισούται με το γινόμενο των *m* λόγων επιβίωσης: $\hat{S}(m) = P_1 P_{2.....} P_m$, όπου

$P_1$: είναι το ποσοστό των ατόμων που επιβίωσαν τον $1^o$ χρόνο

$P_2$: είναι το ποσοστό των ατόμων που επιβίωσαν τον $2^o$ χρόνο, ενώ είχαν ήδη επιβιώσει ένα χρόνο

$P_m$: είναι το ποσοστό των ατόμων που επιβίωσαν τον $m^o$ χρόνο, ενώ είχαν ήδη επιβιώσει *m-1* χρόνια

Κι αυτό επειδή μπορεί να θεωρηθεί ότι τα άτομα που επιβίωσαν για 2 χρόνια, ουσιαστικά επιβίωσαν τον $1^o$ χρόνο κι έπειτα επιβίωσαν ένα χρόνο ακόμη. Άρα η πιθανότητα να επιβιώσουν 2 ή παραπάνω χρόνια είναι ίση με την πιθανότητα να επιβιώσουν τον $1^o$ χρόνο και επιπλέον 1 χρόνο ακόμα. Δηλαδή:

$\hat{S}(2)$ = *P(ποσοστό ατόμων που επιβίωσαν 1 χρόνο και μετά κι άλλον 1 χρόνο) = P(ποσοστό ατόμων που επιβίωσαν για 2 χρόνια δεδομένου ότι επέζησαν τον* $1^o$ *χρόνο) ·P(ποσοστό ατόμων που επιβίωσαν τον* $1^o$ *χρόνο)* $\Leftrightarrow$

$\hat{S}(2) = P_1 \cdot P_2$.

Με τον ίδιο τρόπο ορίζεται και η εκτίμηση του *S(m)*.

Στην πραγματικότητα όμως σε μία έρευνα υπάρχουν πάντα και περικομμένοι χρόνοι (για προκαθορισμένη χρονική διάρκεια του πειράματος ή για προκαθορισμένο ποσοστό επιτυχίας). Η μέθοδος Kaplan – Meier δίνει εκτιμήσεις των τιμών της συνάρτησης επιβίωσης για τους μη περικομμένους χρόνους.

Για να γίνει αυτό όλοι οι χρόνοι κατατάσσονται σε αύξουσα σειρά $t_{(1)} \le t_{(2)} \le \ldots \le t_{(n)}$, όπου *n* είναι ο συνολικός αριθμός των μονάδων που πήραν μέρος στο πείραμα κι έτσι η εκτιμήσεις της συνάρτησης επιβίωσης για κάθε μη περικομμένο χρόνο δίνονται από τον τύπο: $\hat{S}(t) = \prod_{t_{(i)} \le t} \frac{n-i}{n-(i-1)}$, όπου το $t_{(i)}$ είναι χρόνος που αντιστοιχεί σε μη περικομμένη παρατήρηση.

Έχοντας τις εκτιμήσεις της συνάρτησης επιβίωσης, κατασκευάζεται η καμπύλη επιβίωσης (το γράφημα) στο οποίο οι μη περικομμένοι χρόνοι επιβίωσης είναι χωρισμένοι σε διαστήματα και σε κάθε διάστημα αντιστοιχεί μία τιμή της $\hat{S}(t)$. Ωστόσο για τη σύγκριση των καμπυλών επιβίωσης ομάδων χρησιμοποιούνται άλλες στατιστικές τεχνικές, όπως το Logrank test ή το Cox – Mantel.

- Η $\hat{S}(t)$ είναι η εκτιμήτρια της συνάρτησης επιβίωσης του πληθυσμού και γι' αυτό χρησιμοποιώντας το τυπικό σφάλμα της, μπορεί να υπολογιστεί η δειγματική μεταβλητότητα, κατασκευάζοντας μία ζώνη εμπιστοσύνης γύρω από την $\hat{S}(t)$, η οποία θα υπολογίζεται από τον τύπο:

  $\hat{S}(t_{(i)}) \pm z_{1-\frac{a}{2}} s.e.\left[\hat{S}(t_{(i)})\right]$, για κάθε μη περικομμένο χρόνο $t_{(i)}$.

- Η διακύμανση της $\hat{S}(t)$ θα είναι: $Var\left[\hat{S}(t)\right] = \left[\hat{S}(t)\right]^2 \sum_{t_{(i)} \le t} \frac{1}{(n-i)\left[n-(i-1)\right]}$.

  Άρα $s.e.\left[\hat{S}(t)\right] = \sqrt{Var\left[\hat{S}(t)\right]}$

- Ο μέσος χρόνος επιβίωσης μπορεί να υπολογιστεί από την καμπύλη επιβίωσης και θα είναι $\mu_{\text{T}}(t) = \frac{1}{S(t)} \sum_{x=t}^{\infty} S(x)$.

  Αν $t = 0$: $\mu_{\text{T}}(0) = \frac{1}{S(0)} \sum_{x=0}^{\infty} S(x) = \sum_{0}^{\infty} S(x)$ αφού $S(0)=1$.

  Είναι, άρα, το εμβαδόν ανάμεσα στην καμπύλη επιβίωσης και τον άξονα των χρόνων $t_{i.}$, δηλαδή το άθροισμα των εμβαδών των ορθογωνίων κάτω από την καμπύλη επιβίωσης, που κατασκευάστηκε από τους μη περικομμένους χρόνους.

### 3.5 Το μοντέλο αναλογικού κινδύνου του Cox (Cox PH model)

Οι μέθοδοι που έχουν ήδη αναφερθεί δεν μπορούν να χρησιμοποιηθούν όταν ο ερευνητής ενδιαφέρεται και για τις επιδράσεις άλλων μεταβλητών στην συνάρτηση επιβίωσης. Κάποιες άλλες μεταβλητές που θα ήταν ίσως σημαντικές σε μία κλινική έρευνα είναι η ηλικία του ασθενούς, το φύλο του, το οικογενειακό ιστορικό, η διάρκεια της θεραπείας κ.α. που ορισμένες από αυτές είναι ποιοτικές.

Το μοντέλο Cox είναι μία πολύ διαδεδομένη στατιστική τεχνική που εφαρμόζεται για την ανάλυση δεδομένων επιβίωσης. Με τη χρήση του, ο ερευνητής βρίσκει εκτιμήσεις για την επίδραση της θεραπείας στην επιβίωση των ασθενών, αφού προσαρμοστούν και οι άλλες επεξηγηματικές μεταβλητές και προφανώς εκτός από τη συνάρτηση επιβίωσης, λαμβάνονται επίσης, εκτιμήσεις για τη συνάρτηση βαθμού κινδύνου $h_T(t)$ και για τη συνάρτηση πυκνότητας πιθανότητας $f_T(t)$.

Σύμφωνα με τους Altman (1991) και Cox (1972) ισχύουν τα παρακάτω:

- Έστω $X_1,\ldots,X_p$ κάποιες μεταβλητές που ενδέχεται να επηρεάζουν την επιβίωση των ασθενών. Το γενικό μοντέλο με τη μέθοδο Cox περιγράφει και εκτιμά τη συνάρτηση βαθμού κινδύνου και είναι το εξής:
  $h_T(t \mid \underset{\sim}{X}) = h_0(t)\exp\left(b_1X_1 + b_2X_2 + \ldots b_pX_p\right)$, όπου $b_1, b_2, \ldots b_p$ είναι οι συντελεστές που αντιστοιχούν σε κάθε μία από τις μεταβλητές $X_1,\ldots,X_p$. (Cox proportional hazard regression model).

- Η συνάρτηση επιβίωσης και η συνάρτηση πυκνότητας πιθανότητας ορίζονται μονοσήμαντα από τη συνάρτηση κινδύνου (κεφ. 1) και είναι $S(t) = \exp\left[-\int_0^t h_T(x)dx\right]$ και $f_T(t) = h_T(t)\cdot\exp\left[-\int_0^t h_T(x)dx\right], t \geq 0$ (συνεχής χρόνος), άρα η εκτίμηση της συνάρτησης κινδύνου δίνει εκτιμήσεις και για τις συναρτήσεις $S(t)$ και $f_T(t)$.

- Στην περίπτωση που κάποιες μεταβλητές είναι ποιοτικές, χρησιμοποιούνται ψευδομεταβλητές (Dummy variables) για να περιγράψουν την πληροφορία που εμπεριέχεται στις ποιοτικές αυτές μεταβλητές. Οι κατηγορικές μεταβλητές

μπορεί να είναι είτε δυαδικές είτε να έχουνε περισσότερες από 2 κατηγορίες. Για κάθε κατηγορική μεταβλητή με $i \geq 2$ κατηγορίες, απαιτείται η κατασκευή *i-1* ψευδομεταβλητών. Οι ψευδομεταβλητές παίρνουν τις τιμές $\{0,1\}$.

- Αν στο παραπάνω μοντέλο γίνει χρήση νεπέριων λογαρίθμων τότε αυτό γίνεται: $\ln h_T(t \mid \underset{\sim}{X}) = \ln h_0(t) + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p$.

  Πρόκειται δηλαδή για ένα μοντέλο λογιστικής παλινδρόμησης, του οποίου οι συντελεστές $b_i, i = 1, 2, \ldots p$ εκτιμώνται μεγιστοποιώντας την περιθώρια συνάρτηση πιθανοφάνειας (Cox partial likelihood).

- Η περιθώρια συνάρτηση πιθανοφάνειας είναι η $L(\underset{\sim}{b}) = \prod_{X_i} \frac{\exp\left(X_i \underset{\sim}{b}\right)}{\sum_{X_j \geq X_i} \exp\left(X_j \underset{\sim}{b}\right)}$ η

  οποία με τη χρήση νεπέριων λογαρίθμων γίνεται:

  $l(\underset{\sim}{b}) = \ln L(\underset{\sim}{b}) = \sum_{X_i} \left\{ X_i \underset{\sim}{b} - \ln\left[ \sum_{X_j \geq X_i} \exp\left(X_j \underset{\sim}{b}\right) \right] \right\}$ $l(\underset{\sim}{b})$ (partial log-likelihood), όπου

  $\underset{\sim}{b}$ ένα διάνυσμα (*p x 1)* και οι τιμές $Y_i$ αντιστοιχούν σε μη περικομμένους χρόνους. Με τη μεγιστοποίηση της $l(\underset{\sim}{b})$, λαμβάνονται οι εκτιμητές των $b_i$, *i = 1, 2, ...p.*

- Υποθέτουμε ότι για το μοντέλο Cox, οι μεταβλητές $X_1, \ldots, X_p$ επιδρούν προσθετικά στην ποσότητα $\ln h_T(t \mid \underset{\sim}{X})$, δηλαδή δεν υπάρχουν αλληλεπιδράσεις (interactions) και ότι η ποσότητα $\ln h_T(t \mid \underset{\sim}{X})$ μεταβάλλεται γραμμικά με τα $b_i$ όπου *i = 1, 2,...p.*

- Επίσης, η μέθοδος του Cox δεν υποθέτει μια ιδιαίτερη κατανομή για τους χρόνους επιβίωσης, αλλά υποθέτει ότι οι επιδράσεις διαφορετικών μεταβλητών στην επιβίωση είναι σταθερές ως προς το χρόνο.

- Η ποσότητα $h_0(t)$ είναι το επίπεδο αναφοράς της συνάρτησης βαθμού κινδύνου και αντιστοιχεί στην πιθανότητα θανάτου, όταν όλες οι επεξηγηματικές μεταβλητές είναι μηδέν. Δηλαδή είναι η σταθερά του μοντέλου παλινδρόμησης.

- Είναι ένα ημιπαραμετρικό μοντέλο, καθώς δεν γίνονται υποθέσεις για τη μορφή της συνάρτησης κινδύνου $h_T(t)$ (το μη παραμετρικό κομμάτι του μοντέλου), αλλά υποθέτει παραμετρική μορφή όσον αφορά την επίδραση των επεξηγηματικών μεταβλητών πάνω στην $h_T(t)$. Πάντως τις περισσότερες φορές, η έρευνα επικεντρώνεται στην εκτίμηση των παραμέτρων, παρά στο σχήμα της συνάρτησης κινδύνου $h_T(t)$.

### 3.6 Λόγοι στιγμιαίων κινδύνων (hazard ratios)

Εκτός από τη συνάρτηση κινδύνου, το μοντέλο εξετάζει και τους συντελεστές για κάθε μία από τις επεξηγηματικές μεταβλητές. Αφού η σχέση των μεταβλητών $X_i$, για $i = 1, 2, ...p$ με το λογάριθμο της συνάρτησης βαθμού κινδύνου $\ln h_T(t \mid \underset{\sim}{X})$ είναι γραμμική, τότε η αύξηση του συντελεστή $b_j$ μιας επεξηγηματικής μεταβλητής $X_j$ (όπου $j = 1, 2, ...p$) κατά 1 μονάδα, θα συνεπάγεται αύξηση της τιμής του λογαρίθμου της συνάρτησης $\ln h_T(t \mid \underset{\sim}{X})$ κατά $lnh_0(t)$ μονάδες, στην περίπτωση που όλες οι άλλες επεξηγηματικές μεταβλητές $X_i$ $(i = 1,2,...p$ και $i \neq j)$ παραμένουν σταθερές. Δηλαδή η τιμή της συνάρτησης κινδύνου τώρα θα είναι:

$$\ln h_T(t \mid X_1, X_2 \ldots, X_j + 1, X_{j+1}, \ldots, X_p) = \ln h_0(t) + b_1 X_1 + \ldots b_j (X_j + 1) + \ldots + b_p X_p \Leftrightarrow$$
$$\ln h_T(t \mid X_1, X_2 \ldots, X_j + 1, X_{j+1}, \ldots, X_p) = \ln h_0(t) + b_1 X_1 + \ldots b_j X_j + b_j + \ldots + b_p X_p ,$$

άρα για το συντελεστή $b_j$ θα είναι:

$$b_j = \ln h_T(t \mid X_1, X_2 \ldots, X_j + 1, X_{j+1}, \ldots, X_p) - \ln h_T(t \mid X_1, X_2 \ldots, X_j, X_{j+1}, \ldots, X_p) \Leftrightarrow$$
$$b_j = \ln \frac{h_T(t \mid X_1, X_2 \ldots, X_j + 1, X_{j+1}, \ldots, X_p)}{h_T(t \mid X_1, X_2 \ldots, X_j, X_{j+1}, \ldots, X_p)}$$

$$\text{Άρα } \exp b_j = \frac{h_T(t \mid X_1, X_2 \ldots, X_j + 1, X_{j+1}, \ldots, X_p)}{h_T(t \mid X_1, X_2 \ldots, X_j, X_{j+1}, \ldots, X_p)} .$$

Ο παραπάνω λόγος ονομάζεται **λόγος στιγμιαίου κινδύνου** (hazard ratio) και η τιμή $expb_j\text{-}100\%$ δείχνει την αύξηση (θετικό αποτέλεσμα) ή τη μείωση (αρνητικό

αποτέλεσμα) στην τιμή της συνάρτησης βαθμού κινδύνου (σε λογαρίθμους) όταν η μεταβλητή $X_j$, *(j = 1, 2, ...p)* αυξηθεί κατά 1 μονάδα και οι άλλες παραμείνουν σταθερές. Έτσι ένας θετικός συντελεστής σημαίνει ότι ο κίνδυνος είναι μεγαλύτερος και η πρόγνωση θα είναι χειρότερη για μεγάλες τιμές της μεταβλητής στην οποία αναφέρεται. Αντιθέτως, αρνητικός συντελεστής συνεπάγεται και καλύτερη πρόγνωση για τους ασθενείς με μεγάλες τιμές σε αυτή τη μεταβλητή.

Κατά παρόμοιο τρόπο, υπολογίζονται hazard ratios για δύο διαφορετικά μοντέλα, έστω το $\ln h_T(t \mid \underset{\sim}{X})$ και το $\ln h_T(t \mid \underset{\sim}{X}^*)$ $\underset{\sim}{X}$ όπου $\underset{\sim}{X}$ το σύνολο των μεταβλητών του 1ου μοντέλου και $\underset{\sim}{X}^*$ το σύνολο των μεταβλητών του 2ου μοντέλου, τα οποία μοντέλα είναι "φωλιασμένα" (nested models). Έτσι θα είναι:

$$HR\left(\frac{\underset{\sim}{X}}{\underset{\sim}{X}^*}\right)=\frac{h_T(t \mid \underset{\sim}{X})}{h_T(t \mid \underset{\sim}{X}^*)}=\frac{h_0(t)\exp(\underset{\sim}{X}\underset{\sim}{b})}{h_0(t)\exp(\underset{\sim}{X}^*\underset{\sim}{b})}=\exp\left[(\underset{\sim}{X}-\underset{\sim}{X}^*)\underset{\sim}{b}\right]$$

Κάνοντας χρήση των εκτιμήσεων των $b_i$ , *i =1, 2, ...p,* προκύπτει εκτίμηση και για τον παραπάνω λόγο που είναι: $\widehat{HR}\left(\frac{\underset{\sim}{X}}{\underset{\sim}{X}^*}\right)=\exp\left[\left(\underset{\sim}{X}-\underset{\sim}{X}^*\right)\hat{\underset{\sim}{b}}\right]$.

Έτσι με τις παραπάνω εκτιμήσεις, κατασκευάζονται διαστήματα εμπιστοσύνης για το $HR\left(\frac{\underset{\sim}{X}}{\underset{\sim}{X}^*}\right)$ που θα είναι της μορφής:

$$\exp\left\{\left[\left(\underset{\sim}{X}-\underset{\sim}{X}^*\right)\hat{\underset{\sim}{b}}\right]\pm z_{1-\frac{a}{2}}s.e.\left[\left(\underset{\sim}{X}-\underset{\sim}{X}^*\right)\hat{\underset{\sim}{b}}\right]\right\}$$

σε επίπεδο στατιστικής σημαντικότητας *α%.*

### 3.7 Deviance

Εκτός από τους παραπάνω λόγους μπορούν να χρησιμοποιηθούν και λόγοι πιθανοφανειών για τη σύγκριση μοντέλων με διαφορετικό αριθμό παραμέτρων. Ενώ στην παλινδρόμηση χρησιμοποιείται η δοκιμασία *F* για τον έλεγχο της στατιστικής υπόθεσης ότι όλοι οι συντελεστές των ανεξάρτητων μεταβλητών $X_i$, εκτός της σταθεράς, είναι μηδέν *($H_0$: $b_1 = b_2 = ... = b_p = 0$),* στην λογιστική παλινδρόμηση, άρα

και στο μοντέλο Cox, ο αντίστοιχος έλεγχος γίνεται με το λόγο των πιθανοφανειών που είναι:

$$\text{Deviance} = \Lambda = -2\ln\left(\frac{L_0}{L_1}\right)$$

όπου $L_0$ είναι η τιμή της πιθανοφάνειας για το μοντέλο μόνο με τη σταθερά, δηλαδή το:

$$\ln h_T(t \mid \underset{\sim}{X}) = \ln h_0(t)$$

και $L_1$ είναι η τιμή της πιθανοφάνειας για το μοντέλο που δοκιμάζουμε, δηλαδή το μοντέλο με $p$ επεξηγηματικές μεταβλητές:

$$\ln h_T(t \mid \underset{\sim}{X}) = \ln h_0(t) + b_1X_1 + \ldots + b_pX_p$$

Η παραπάνω στατιστική συνάρτηση ελέγχου $\Lambda$ ακολουθεί την $\chi^2$ κατανομή με βαθμούς ελευθερίας ίσους με τον αριθμό των παραμέτρων.

Αυτό το στατιστικό test είναι το καταλληλότερο για χρήση, όταν το μοντέλο επιλέγεται με τη βηματική μέθοδο αποκλεισμού των μεταβλητών (backward elimination)

Η διαφορά δύο Deviance μπορεί να χρησιμοποιηθεί για να ελεγχθεί η στατιστική σημαντικότητα συντελεστών παλινδρόμησης σε "φωλιασμένα" μοντέλα. Η στατιστική συνάρτηση ελέγχου ($Deviance_{(2)} - Deviance_{(1)}$) ακολουθεί την $\chi^2$ κατανομή με βαθμούς ελευθερίας ($df_{(2)} - df_{(1)}$)

Έστω δηλαδή ότι θέλουμε να γίνει ο έλεγχος για τα μοντέλα:

Model 1: $\ln h_T(t \mid \underset{\sim}{X}) = \ln h_0(t) + b_1X_1 + \ldots b_lX_l$

Model 2: $\ln h_T(t \mid \underset{\sim}{X}) = \ln h_0(t) + b_1X_1 + \ldots + b_lX_l + b_{l+1}X_{l+1} + \ldots + b_pX_p$

Η υπόθεση που θα ελεγχθεί είναι η $H_0$: $b_{l+1} = b_{l+2} = \ldots = b_p = 0$ έναντι της εναλλακτικής $H_1$: $b_k \neq 0$ για τουλάχιστον ένα από τα $k = l+1, l+2, \ldots, p$.

### 3.8 Wald test

Αν απλά για το μοντέλο $\ln h_T(t \mid \underset{\sim}{X}) = \ln h_0(t) + b_1X_1 + b_2X_2 + \ldots + b_pX_p$ θέλουμε να ελέγξουμε την υπόθεση ότι ένας μόνο συντελεστής $b_j$ είναι μηδέν, τότε εφαρμόζουμε

τη δοκιμασία του Wald, η οποία είναι $\left( \frac{\hat{b}_j}{ASE\left(\hat{b}_j\right)} \right)^2$, όπου $ASE\left(\hat{b}_j\right)$ είναι το ασυμπτωτικό τυπικό σφάλμα και ακολουθεί την κατανομή $\chi^2$ με *1* βαθμό ελευθερίας.

Γενικά υπάρχει διαφορά από τις παραμετρικές τεχνικές στο γεγονός ότι οι εκτιμήσεις και τα διαστήματα εμπιστοσύνης προκύπτουν από τη μεγιστοποίηση της περιθώριας πιθανοφάνειας (partial likelihood), αντί της ολικής πιθανοφάνειας (full likelihood)

# ΚΕΦΑΛΑΙΟ 4

## 4.1 ΠΑΡΑΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ

Στο προηγούμενο κεφάλαιο μελετήθηκε η διαδικασία για την εκτίμηση των συναρτήσεων επιβίωσης των ατόμων που λαμβάνουν μέρος σε ένα πείραμα, με μη παραμετρικές μεθόδους, δηλαδή με μεθόδους που εφαρμόζονται στις περιπτώσεις που οι κατανομές των χρόνων επιβίωσης δεν είναι γνωστό ποια συγκεκριμένη θεωρητική κατανομή ακολουθούν. Όταν όμως η θεωρητική κατανομή των χρόνων είναι δεδομένη (δηλαδή είναι γνωστή από πληροφορίες που βασίζονται σε παλαιότερες έρευνες), χρησιμοποιούνται οι αντίστοιχες παραμετρικές μέθοδοι εκτίμησης.

Έχουμε ήδη αναφερθεί στις θεωρητικές στατιστικές κατανομές που χρησιμοποιούνται πιο συχνά για την ανάλυση των δεδομένων επιβίωσης (εκθετική κατανομή, κατανομή Weibull κ.α.), καθώς και στις παραμέτρους, οι οποίες τις χαρακτηρίζουν.

### 4.2 Ιδιότητες της κατανομής Weibull

Η κατανομή, η οποία κάνει τις περισσότερες φορές την καλύτερη εφαρμογή στα δεδομένα, είναι η κατανομή Weibull με παραμέτρους $r$ και $\lambda$, που είναι η γενίκευση της εκθετικής κατανομής.

Η πυκνότητα πιθανότητας της Weibull είναι η $f_T(t) = r\lambda^r t^{r-1} e^{-(\lambda t)^r}$ , $t \geq 0$ .

Για $r = 1$ προκύπτει η εκθετική κατανομή, με παράμετρο $\lambda$ και συνάρτηση πυκνότητας πιθανότητας την $f_T(t) = \lambda e^{-\lambda t}, t \geq 0$.

Όταν $0 < r < 1$ και ο χρόνος $t \to 0$ τότε η $f_T(t) \to \infty$ , ενώ στην περίπτωση που $t \to \infty$ η συνάρτηση $f_T(t) \to 0$

Όταν όμως $r > 1$ και $t \to 0$ θα είναι $f_T(t) \to 0$ , όπως επίσης και όταν $t \to \infty$ .

### 4.3 Εκτίμηση των παραμέτρων της κατανομής Weibull.

Για την εκτίμηση των παραμέτρων $r$ και $\lambda$ χρησιμοποιούνται:

a) Η μέθοδος των ελαχίστων τετραγώνων (Least Squares method)
b) Η μέθοδος της μεγίστης πιθανοφάνειας (Maximum Likelihood)
c) Η μέθοδος του Bayes.

Σε κάθε μία από τις παρακάτω περιπτώσεις, έχουμε $n$ παρατηρήσεις, που περιγράφουν τους χρόνους επιβίωσης $t_1,\ldots,t_n$, οι οποίοι προέρχονται από πληθυσμό που ακολουθεί την κατανομή Weibull.

#### 4.3.1 Μέθοδος ελαχίστων τετραγώνων

Η μέθοδος των ελαχίστων τετραγώνων είναι κατάλληλη όταν στα δεδομένα δεν υπάρχουν περικομμένες παρατηρήσεις ή παρατηρήσεις που αντιστοιχούν σε ίσους χρόνους και όταν η συνάρτηση που χρησιμοποιείται μπορεί να γραφτεί σε γραμμική μορφή. Οι περισσότερες συναρτήσεις που χρησιμοποιούνται στην ανάλυση επιβίωσης μετασχηματίζονται σε γραμμικές, το ίδιο συμβαίνει και για την κατανομή Weibull.

Με τη χρήση της μεθόδου των ελαχίστων τετραγώνων χρειαζόμαστε την αθροιστική συνάρτηση της κατανομής Weibull που είναι η $F_T(t)=1-\exp\left[-(\lambda t)^r\right]$ όπου $t \geq 0$.

Η συνάρτηση πρέπει να μετασχηματιστεί σε γραμμική μορφή, άρα χρησιμοποιώντας λογαρίθμους θα είναι:

$$\ln\left[1-F_T(t)\right]=-(\lambda t)^r \Leftrightarrow$$

$$-\ln\left[1-F_T(t)\right]=(\lambda t)^r \Leftrightarrow$$

$$\ln\left\{-\ln\left[1-F_T(t)\right]\right\}=r\left[\ln\lambda+\ln t\right]\Leftrightarrow$$

$$\ln\left\{-\ln\left[1-F_T(t)\right]\right\}=r\ln\lambda+r\ln t$$

Αν τώρα θέσουμε $Y = \ln\left\{-\ln\left[1-F_T(t)\right]\right\}$, $a^* = r\ln\lambda$, $b^* = r$ και $X = \ln t$, τότε προκύπτει η γραμμική σχέση $Y = a^* + b^*\mathrm{X}$. Οι τιμές των $F_T(t_i)$ μπορούν να εκτιμηθούν με μη παραμετρικές μεθόδους, όπως η μέθοδος Kaplan – Meier αφού ισχύει η ισότητα $S(t_i) = 1 - F_T(t_i)$ ή η μέθοδος Median Rank, κατά την οποία αντί να βρεθεί η τάξη $k$ της κάθε παρατήρησης, βρίσκεται η διάμεσος τάξη που είναι σ' ένα συγκεκριμένο επίπεδο σημαντικότητας (50 %) και ισχύει $F_T(t_i) = MR = \dfrac{k-0,3}{n+0,4}$

Με τη μέθοδο των ελαχίστων τετραγώνων οι εκτιμητές των $\alpha^*$ και $\mathrm{b}^*$ εκλέγονται κατά τέτοιο τρόπο, ώστε το άθροισμα τετραγώνων των σφαλμάτων $S = \sum_{i=1}^{n}(Y_i - a^* - b^* X_i)^2$ να γίνεται ελάχιστο.

Από τις εξισώσεις: $\left.\begin{array}{l}\dfrac{dS}{da^*} = 0\\ \dfrac{dS}{db^*} = 0\end{array}\right\}$ προκύπτουν οι εκτιμητές ελαχίστων τετραγώνων

$\hat{b}^* = \dfrac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{X})^2}$ ή $\hat{b}^* = \dfrac{\sum_{i=1}^{n}x_i y_i - \frac{1}{n}\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}x_i^2 - \frac{1}{n}(\sum_{i=1}^{n}x_i)^2}$ και

$\hat{a}^* = \bar{Y} - \hat{b}^*\bar{X}$

Επειδή όπως είδαμε $b^* = r$ και $a^* = r\ln\lambda$, οι παράμετροι της κατανομής Weibull θα είναι $r = \hat{b}^*$ και $\lambda = \exp\dfrac{\hat{a}^*}{\hat{b}^*}$

**4.3.2 Μέθοδος μέγιστης πιθανοφάνειας.**

Η συνάρτηση $L(\underset{\sim}{\theta}) = \prod_{i=1}^{n} f_T(t_i,\underset{\sim}{\theta}) = f_T(t_1,\underset{\sim}{\theta})\,f_T(t_2,\underset{\sim}{\theta})\ldots f_T(t_n,\underset{\sim}{\theta})$ ονομάζεται συνάρτηση πιθανοφάνειας και θεωρείται συνάρτηση της παραμέτρου $\underset{\sim}{\theta} = \{\theta_1,\ldots,\theta_k\}$ μίας κατανομής με συνάρτηση πυκνότητας πιθανότητας $f_T(t,\underset{\sim}{\theta})$.

Η αρχή της μέγιστης πιθανοφάνειας ερμηνεύεται ως εξής:

Η $L(\underset{\sim}{\theta})$ εκφράζει την πιθανότητα να παρατηρηθεί το τυχαίο δείγμα $T_1,\ldots,T_n$. Εφ' όσον οι τιμές $t_1,\ldots,t_n$ έχουν πραγματοποιηθεί, πρέπει να έχουν μεγάλη πιθανότητα. Έτσι μεγιστοποίηση της $L(\underset{\sim}{\theta})$ σημαίνει επιλογή εκείνης της τιμής $\underset{\sim}{\theta}$, δηλαδή επιλογή των παραμέτρων $\theta_1,\ldots,\theta_k$, τέτοια ώστε να μεγιστοποιείται η πιθανότητα αυτή.

Χάριν ευκολίας αντί να μεγιστοποιούμε την $L(\underset{\sim}{\theta})$ μεγιστοποιούμε ισοδύναμα την $\ln L(\underset{\sim}{\theta})$, διότι το μέγιστο λαμβάνεται στο ίδιο σημείο.

Έτσι με τη λύση του συστήματος των εξισώσεων:

$$\left.\begin{aligned} &\frac{d\ln L}{d\theta_1}=0\\ &\frac{d\ln L}{d\theta_2}=0\\ &\vdots\\ &\frac{d\ln L}{d\theta_k}=0 \end{aligned}\right\}$$

προκύπτουν οι εκτιμήσεις των παραμέτρων $\hat{\theta}_1,\ldots,\hat{\theta}_k$.

Η συνάρτηση πιθανοφάνειας της κατανομής Weibull είναι η $L(\underset{\sim}{\theta})=\prod_{i=1}^{n} f_T(t_i,\underset{\sim}{\theta})$, όπου $f_T(t_i,\underset{\sim}{\theta})=r\lambda^r t_i^{r-1}\exp\left[-(\lambda t_i)^r\right]$ είναι η συνάρτηση πυκνότητας πιθανότητας της Weibull και $\underset{\sim}{\theta}=\{r,\lambda\}$ οι παράμετροι της.

Θα είναι $L(\underset{\sim}{\theta})=\prod_{i=1}^{n} r\lambda^r t_i^{r-1}\exp\left[-(\lambda t_i)^r\right]=r^n\lambda^{rn}\prod_{i=1}^{n} t_i^{r-1}\prod_{i=1}^{n}\exp\left[-(\lambda t_i)^r\right]$ άρα

$$\ln L(\underset{\sim}{\theta})=n\ln r+nr\ln\lambda+(r-1)\sum_{i=1}^{n}\ln t_i+\sum_{i=1}^{n}\left[-(\lambda t_i)^r\right]$$

Για την εύρεση των εκτιμητών $\hat{r}$ και $\hat{\lambda}$ χρησιμοποιούμε το παρακάτω σύστημα εξισώσεων:

$$\left.\begin{aligned}\frac{d\ln L}{dr}=0\\ \frac{d\ln L}{d\lambda}=0\end{aligned}\right\} \text{ δηλαδή ισοδύναμα}$$

$$\left.\begin{aligned}\frac{n}{r}+n\ln\lambda+\sum_{i=1}^{n}\ln t_i-\sum_{i=1}^{n}\left(\lambda t_i\right)^r\ln\left(\lambda t_i\right)=0\\ \frac{nr}{\lambda}-r\lambda^{r-1}\sum_{i=1}^{n}t_i^{\,r}=0\end{aligned}\right\}\Leftrightarrow$$

$$\left.\begin{aligned}\frac{n}{r}+\sum_{i=1}^{n}\ln(\lambda t_i)-\sum_{i=1}^{n}\left(\lambda t_i\right)^r\ln\left(\lambda t_i\right)=0\\ n=\lambda^r\sum_{i=1}^{n}t_i^{\,r}\end{aligned}\right\}$$

Η λύση του συστήματος (η οποία ξεφεύγει από το σκοπό αυτής της εργασίας) δίνει τους εκτιμητές των $\hat{r}$ και $\hat{\lambda}$.

### 4.3.3 Μέθοδος Bayes

Για την εκτίμηση των παραμέτρων $\alpha$ και $\lambda$ της κατανομής Weibull μπορούν να χρησιμοποιηθούν και οι μέθοδοι που προκύπτουν από τη γενική θεωρία του νόμου του Bayes.

Στην κλασσική στατιστική, όλα τα συμπεράσματα βασίζονται στα δεδομένα του δείγματος. Στη στατιστική κατά Bayes, όπως είδαμε στο Κεφάλαιο 2, η εκ των προτέρων πληροφορίες, δηλαδή οι πληροφορίες από τ' αποτελέσματα παλαιότερων ερευνών, αποτελούν τη βάση των αναλύσεων και όλες οι παράμετροι είναι τυχαίες μεταβλητές.

Το μοντέλο, καταρχήν, υποθέτει την ύπαρξη εκ των προτέρων (a priori) κατανομών για τις παραμέτρους $r$ και $\lambda$. Ειδικότερα όταν η έρευνα που διεξάγεται, εφαρμόζεται σε μικρό δείγμα δεδομένων, οι προηγούμενες γνώσεις για τις κατανομές των παραμέτρων είναι πολύ χρήσιμες.

Υποθέτουμε, καταρχήν, ότι οι 2 παράμετροι είναι ανεξάρτητες μεταξύ τους. Άρα $\pi(r, \lambda) = \pi_1(r)\cdot\pi_2(\lambda)$, όπου $\pi_1(r)$, $\pi_2(\lambda)$ είναι οι 2 εκ των προτέρων περιθώριες κατανομές των παραμέτρων $r$ και $\lambda$ και $\pi(r, \lambda)$ η από κοινού.

Εφαρμόζοντας το θεώρημα Bayes, η εκ των υστέρων κατανομή (posterior) για τις παραμέτρους $r$ και $\lambda$ θα είναι:

$$g(r,\lambda \mid t) = \frac{L(r,\lambda)\pi_1(r)\pi_2(\lambda)}{\int_0^\infty\int_0^\infty L(r,\lambda)\pi_1(r)\pi_2(\lambda)dr d\lambda}$$

όπου $L(r,\lambda)$ είναι η συνάρτηση πιθανοφάνειας.

Η $g(r, \lambda \mid t)$ είναι η από κοινού κατανομή των παραμέτρων $r$ και $\lambda$. Άρα οι δύο εκ των υστέρων περιθώριες κατανομές θα είναι:

$$g_1(r \mid t) = \int_0^\infty g(r,\lambda \mid t)d\lambda \text{ και } g_2(\lambda \mid t) = \int_0^\infty g(r,\lambda \mid t)dr .$$

Οι εκ των υστέρων κατανομές $g_1(r \mid t)$ και $g_2(\lambda \mid t)$ των παραμέτρων δεν είναι απαραίτητο ότι θα μοιάζουν στο σχήμα με τις αντίστοιχες εκ των προτέρων. Μπορεί δηλαδή, μετά την εφαρμογή των δεδομένων η πεποίθηση μας για τις κατανομές των $r$ και $\lambda$ να αλλάξει εντελώς.

Έτσι η διαδικασία που ακολουθείται για την ανάλυση της κατανομής Weibull κατά Bayes είναι η εξής:

1. Συγκέντρωση των χρόνων επιβίωσης $t_1,\ldots,t_n$, δηλαδή τα δεδομένα της έρευνας.
2. Ορισμός της εκ των προτέρων κατανομής για την παράμετρο $r$, καθώς για την $\lambda$.
3. Υπολογισμός της εκ των υστέρων κατανομής $g(r, \lambda \mid t)$.
4. Υπολογισμός των περιθώριων $g_1(r \mid t)$ και $g_2(\lambda \mid t)$.

Για τις παραμέτρους δεν θα βρεθούν συγκεκριμένες τιμές, όπως με τις προηγούμενες 2 μεθόδους. Επειδή πρόκειται πλέον, για τυχαίες μεταβλητές, μπορούν να βρεθούν οι αναμενόμενες τιμές $E(r)$ και $E(\lambda)$ ή οι διάμεσοι κ.α.

- Η αναμενόμενη τιμή της παραμέτρου *r* είναι:

  $$\mathrm{E}(r) = \int_0^\infty \int_0^\infty r g(r, \lambda \mid t) dr d\lambda = \int_0^\infty r g_1(r \mid t) dr ,$$

  ενώ η αναμενόμενη τιμή της παραμέτρου $\lambda$ είναι:

  $$E(\lambda) = \int_0^\infty \int_0^\infty \lambda g(r, \lambda \mid t) dr d\lambda = \int_0^\infty \lambda g_2(\lambda \mid t) d\lambda$$

- Οι διάμεσοι θα προκύψουν από τις παρακάτω 2 ισότητες, αν αυτές λυθούν ως προς *r* και ως προς $\lambda$:

  $$\int_0^\infty \int_o^r g(r, \lambda \mid t) dr d\lambda = 0,5 \text{ και}$$

  $$\int_0^\lambda \int_o^\infty g(r, \lambda \mid t) dr d\lambda = 0,5$$

- Επίσης από τις 2 προηγούμενες ισότητες μπορούν να υπολογιστούν εκατοστιαία σημεία, αντικαθιστώντας στο δεύτερο μέλος το *0,5* πχ. με *0,1* αν πρόκειται για το $10^o$ εκατοστιαίο σημείο ή με το *0,95* αν πρόκειται για το $95^o$ εκατοστιαίο σημείο.

- Επίσης με τη χρήση της μεθόδου Bayes βρίσκονται εκ των υστέρων διαστήματα αξιοπιστίας για τις 2 παραμέτρους.

  Είναι $P(r \le r_U) = \int_0^{r_U} g_1(r \mid t) dr$ και $P(r \ge r_L) = 1 - P(r \le r_L) = \int_{r_L}^\infty g_1(r \mid t) dr$ , όπου $r_U$ είναι ένα άνω όριο και $r_L$ ένα κάτω όριο της τιμής της παραμέτρου *r*.

  Έτσι το διάστημα $(r_L, r_U)$ θα είναι ένα π.χ. 95% διάστημα αξιοπιστίας για την παράμετρο *r* με πιθανότητα:

  $$P(r_L \le r \le r_U) = \int_{r_L}^{r_U} g_1(r \mid t) dr = 0,95$$

Ομοίως για την παράμετρο $\lambda$ θα ισχύει $P(\lambda \le \lambda_U) = \int_0^{\lambda_U} g_2(\lambda \mid t) d\lambda$ και $P(\lambda \ge \lambda_L) = 1 - P(\lambda \le \lambda_L) = \int_{\lambda_L}^{\infty} g_2(\lambda \mid t) d\lambda$, όπου $\lambda_U$ είναι ένα άνω όριο και $\lambda_L$ ένα κάτω όριο της τιμής της παραμέτρου $\lambda$, οπότε ένα 95% διάστημα αξιοπιστίας θα είναι το $(\lambda_L, \lambda_U)$ με πιθανότητα:

$$P(\lambda_L \le \lambda \le \lambda_U) = \int_{\lambda_L}^{\lambda_U} g_2(\lambda \mid t) d\lambda = 0,95 .$$

# ΚΕΦΑΛΑΙΟ 5

## 5.1 ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΟ ΤΗΝ ΕΚ ΤΩΝ ΥΣΤΕΡΩΝ ΚΑΤΑΝΟΜΗ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ WINBUGS

Όπως ήδη αναφέρθηκε, στην κλασσική στατιστική προκειμένου να εκτιμηθεί η άγνωστη παράμετρος $\underset{\sim}{\theta} = \{\theta_1, \ldots, \theta_k\}$ μίας κατανομής, χρησιμοποιείται η μέθοδος ελαχίστων τετραγώνων ή η μέθοδος της μεγιστοποίησης της συνάρτησης πιθανοφάνειας. Και στις δύο προσεγγίσεις οι παράμετροι $\theta_1, \ldots, \theta_k$ θεωρούνται σταθερές άγνωστες ποσότητες και η προσέγγισή τους γίνεται μέσω εκτιμητών, οι οποίοι είναι συναρτήσεις των τυχαίων μεταβλητών $X_1, \ldots, X_n$ του δείγματος. Αντιθέτως, στη στατιστική κατά Bayes οι $\theta_1, \ldots, \theta_k$ θεωρούνται τυχαίες μεταβλητές και η εκτίμησή τους βασίζεται στην εκ των υστέρων κατανομή $g(\underset{\sim}{\theta} \mid t)$, με δεδομένη την εκ των προτέρων κατανομή $\pi(\underset{\sim}{\theta})$. Έτσι λοιπόν, η μέθοδος κατά Bayes προσφέρει το βασικό πλεονέκτημα να μπορεί να συμπεριλάβει πληροφορία από παλαιότερες έρευνες. Παρόλα αυτά, οι εκ των υστέρων κατανομές δεν μπορούν πάντα να υπολογιστούν αναλυτικά και συνεπώς η εφαρμογή της προσέγγισης αυτής παρουσιάζει υπολογιστικά προβλήματα.

Με την ανάπτυξη της επιστήμης της πληροφορικής και την εξέλιξη των υπολογιστών τα τελευταία χρόνια, βρέθηκε λύση στο πρόβλημα αυτό. Ο βασικός τρόπος υπολογισμού των εκ των υστέρων κατανομών στηρίζεται στην προσομοίωση Μόντε Κάρλο με τη χρήση Μαρκοβιανών Αλυσίδων (Markov Chain Monte Carlo – MCMC). Μέσω αυτής της προσέγγισης παράγεται ένα δείγμα από προσομοιωμένες τιμές από τις εκ των υστέρων κατανομές των αγνώστων παραμέτρων του μοντέλου. Η ανάλυση αυτών των τιμών θα μας δώσει και τις εκτιμήσεις μας για το μοντέλο.

### 5.2 Στοχαστικές Διαδικασίες - Μαρκοβιανές Αλυσίδες

Όπως είναι γνωστό η θεωρία των στοχαστικών διαδικασιών μελετά συστήματα ή φαινόμενα, των οποίων η εξέλιξη στον χρόνο διέπεται από τους νόμους των

πιθανοτήτων. Η εξέλιξη αυτή στον χρόνο δεν είναι στατική, αλλά δυναμική και στην περίπτωση αυτή για να περιγραφεί η εξέλιξη του φαινομένου, χρειάζονται περισσότερες από μία τυχαίες μεταβλητές. Συνεπώς απαιτείται μία οικογένεια τυχαίων μεταβλητών, κάθε μία από τις οποίες θα περιγράφει το φαινόμενο σε συγκεκριμένη χρονική στιγμή.

Μία στοχαστική διαδικασία είναι μία οικογένεια τυχαίων μεταβλητών $\{X(t) : t \in T\}$ όπου *t* είναι μία παράμετρος που παίρνει τιμές σε ένα κατάλληλα ορισμένο σύνολο *T*. Ο δειγματικός χώρος *Ω*, των τυχαίων μεταβλητών *X(t)* ονομάζεται χώρος καταστάσεων. Ανάλογα με το σύνολο των τιμών της *t* καθώς και τη μορφή του χώρου καταστάσεων της τυχαίας μεταβλητής *X(t)* μπορούμε να κατατάξουμε τις στοχαστικές διαδικασίες σε 4 κατηγορίες. Συγκεκριμένα η στοχαστική διαδικασία με διακριτό (πεπερασμένο ή αριθμήσιμο) χώρο καταστάσεων και με διακριτό σύνολο τιμών *(t = 0, 1, 2...)* λέγεται στοχαστική αλυσίδα και συμβολίζεται με $\{X_n, n = 0,1,\ldots\}$.

Μία στοχαστική διαδικασία ονομάζεται Μαρκοβιανή διαδικασία, αν δοθείσης της κατάστασής της τη χρονική στιγμή *t,* η κατάστασή της τη στιγμή *s* με *s > t,* δεν εξαρτάται από τις καταστάσεις της σε χρονικές στιγμές μικρότερες του *t*, δηλαδή αν γνωρίζουμε ακριβώς την παρούσα τιμή της διαδικασίας, η μελλοντική της συμπεριφορά δεν εξαρτάται από την συμπεριφορά της στο παρελθόν. (Λάγκαρης - 1988, σελ.7)

Έτσι για κάθε διατεταγμένο δείγμα χρονικών στιγμών $t_1 < t_2 < \ldots < t_n <$ $t_{n+1}$ ισχύει ότι:

$$P\left(X(t_{n+1}) = i \mid X(t_1) = j_1, X(t_2) = j_2, \ldots X(t_n) = j_n\right) = P\left(X(t_{n+1}) = i \mid X(t_n) = j_n\right).$$

Μία στοχαστική αλυσίδα η οποία έχει τη Μαρκοβιανή ιδιότητα ονομάζεται Μαρκοβιανή αλυσίδα. Δηλαδή η Μαρκοβιανή αλυσίδα είναι η ακολουθία $X_0, X_1 \ldots X_n$ διακεκριμένων τυχαίων μεταβλητών, με την ιδιότητα ότι η δεσμευμένη κατανομή του $X_{n+1,}$ δοθέντων των $X_0, X_1 \ldots, X_n$ εξαρτάται μόνο από την τιμή του $X_n$ και όχι από τα $X_0, X_1 \ldots, X_{n-1}$.

Δηλαδή: $P\left(X_{n+1} \mid X_1, X_2, \ldots X_n\right) = P\left(X_{n+1} \mid X_n\right)$

### 5.3 Προσομοίωση από την εκ των υστέρων κατανομή

Κάθε Μαρκοβιανή αλυσίδα αφού αποτελείται από μία οικογένεια τυχαίων μεταβλητών $X_0, X_1 \ldots X_n$ έχει μία στάσιμη κατανομή. Για να προσομοιώσουμε από την εκ των υστέρων κατανομή του $\underset{\sim}{\theta}$, χρησιμοποιούμε μία Μαρκοβιανή αλυσίδα, η οποία έχει στάσιμη κατανομή ίση με την εκ των υστέρων $g(\underset{\sim}{\theta} \mid t)$ που επιθυμούμε να εκτιμήσουμε. Οι τιμές αυτές της Μαρκοβιανής αλυσίδας συμβολίζονται με $\underset{\sim}{\theta}^{(t)}$, όπου για $t = 0$ έχουμε τις αρχικές τιμές $\underset{\sim}{\theta}^{(0)}$ (initial values), δηλαδή την αρχική κατάσταση εκκίνησης της αλυσίδας. Κατασκευάζουμε έναν αλγόριθμο και σύμφωνα με αυτόν προσομοιώνουμε τυχαίες τιμές από την κατανομή της Μαρκοβιανής αλυσίδας. Δύο είδη αλγορίθμων είναι τα πιο δημοφιλή: ο αλγόριθμος Metropolis – Hastings και ο δειγματολήπτης Gibbs.

Όταν σιγουρευτούμε ότι η Μαρκοβιανή αλυσίδα έχει συγκλίνει, δηλαδή ότι μας δίνει τυχαίες τιμές από την εκ των υστέρων κατανομή, τότε η διαδικασία προσομοίωσης σταματάει από τον χρήστη. Οι τιμές που έχουν προσομοιωθεί απ' αυτή τη διαδικασία, αποτελούν ένα τυχαίο δείγμα από την εκ των υστέρων κατανομή $g(\underset{\sim}{\theta} \mid t)$. Λόγω όμως της κατασκευής του αλγορίθμου, ο οποίος δεν είναι παρά μια Μαρκοβιανή αλυσίδα, το δείγμα δεν θα είναι ανεξάρτητο, αλλά θα παρουσιάζει αυτοσυσχετίσεις. Έτσι, χρειάζεται ειδική ανάλυση για την αφαίρεση των επιδράσεων των αυτοσυσχετίσεων.

### 5.4 Χρήση του BUGS - WINBUGS

Όπως αναφέρθηκε ήδη, οι υπολογισμοί για την εύρεση της εκ των υστέρων κατανομής της παραμέτρου $\underset{\sim}{\theta} = \{\theta_1, \ldots, \theta_k\}$ είναι συνήθως δύσκολοι ή σε κάποιες περιπτώσεις είναι αδύνατοι. Για να εφαρμοστεί η προηγούμενη διαδικασία, η οποία απλοποιεί τους υπολογισμούς, δημιουργήθηκε περίπου το 1995, το πρόγραμμα BUGS. Η λέξη προκύπτει από τα αρχικά των λέξεων: **B**ayesian inference **U**sing **G**ibbs **S**ampling, δηλαδή Μπεϋζιανή Συμπερασματολογία με τη χρήση του δειγματολήπτη Gibbs. Αργότερα, περίπου το 1998, κυκλοφόρησε η αντίστοιχη έκδοση για τα Windows, η οποία ονομάστηκε WINBUGS και προσφέρει πολύ περισσότερες

δυνατότητες. Η πιο βελτιωμένη έκδοση του κυκλοφόρησε πρόσφατα (WINBUGS 1.4.2).

Στο WINBUGS μπορεί να γίνει ανάλυση των διαθέσιμων δεδομένων με τη χρήση παραμετρικών μοντέλων. Όσον αφορά την ανάλυση δεδομένων επιβίωσης μπορούμε να χρησιμοποιήσουμε μοντέλα βασισμένα στην εκθετική κατανομή, την Weibull, την Γάμμα και την Λογαριθμοκανονική. (βλ. Congdon, 2001). Σε κάθε περίπτωση, τα περικομμένα δεδομένα (right censored data) συμπεριλαμβάνονται στην ανάλυση, αφού σύμφωνα με τον ορισμό του μοντέλου που χρησιμοποιείται, γίνεται κατάλληλη κωδικοποίηση των αντίστοιχων τιμών. Η ανάλυση κατά Bayes και τα αντίστοιχα αποτελέσματα του WINBUGS, δίνουν εκτιμήσεις, δηλαδή υπολογίζουν τις εκ των υστέρων κατανομές, για όλους τους χρόνους επιβίωσης, περικομμένους και μη. Δηλαδή για τις περικομμένες παρατηρήσεις εκτιμούμε το χρόνο (με μια κατανομή) που θα είχε παρατηρηθεί, αν τα αντίστοιχα άτομα δεν αποχωρούσαν από την έρευνα.

Όλες οι άγνωστες παράμετροι του υπό εκτίμηση μοντέλου θεωρούνται τυχαίες μεταβλητές. Το μοντέλο αποτελείται από συγκεκριμένες κατανομές για τα δεδομένα (περικομμένα και μη περικομμένα) και για τις παραμέτρους. Για να υπολογιστούν οι εκ των υστέρων κατανομές (posterior) των παραμέτρων, χρειάζεται να εισαχθούν στο WINBUGS οι εκ των προτέρων κατανομές (prior) για κάθε μία από τις άγνωστες παραμέτρους $\theta_i$ *(για* $i = 1, 2, \ldots k$*)*, οι οποίες συμπεριλαμβάνουν πληροφορία από προηγούμενες μελέτες και η συνάρτηση πιθανοφάνειας (likelihood).

Στην περίπτωση που δεν έχουμε διαθέσιμη εκ των προτέρων πληροφορία, γίνονται αναλύσεις των δεδομένων για κάθε περίπτωση μίας εκ των προτέρων κατανομής (εκθετικής, Weibull κτλ) ξεχωριστά, οι οποίες όμως είναι επιλεγμένες έτσι ώστε να έχουν μεγάλη διακύμανση (Vague prior – Congdon, 2001). Τέλος, για την επιλογή του καταλληλότερου μοντέλου θα χρησιμοποιήσουμε το Deviance Information Criterion - *DIC* (Spiegelhalter et al, 2003). Ως καταλληλότερο λαμβάνουμε το μοντέλο με το μικρότερο *DIC*.

### 5.5 Η διαδικασία Doodle του WINBUGS

Το WINBUGS δίνει τη δυνατότητα, πριν την εφαρμογή όλων των παραπάνω, να σχεδιαστεί μία γραφική αναπαράσταση του μοντέλου που θα κατασκευαστεί. Αυτό

γίνεται με την επιλογή του μενού Doodle. Στο γράφημα που κατασκευάζεται, φαίνεται η σχέση μεταξύ των ποσοτήτων που ορίζουν το μοντέλο. Τα δεδομένα και οι παράμετροι των κατανομών, χαρακτηρίζονται από ένα κόμβο (node). Οι σταθερές, γνωστές ποσότητες, όπως για παράδειγμα οι τιμές των παραμέτρων των εκ των προτέρων κατανομών και οι επεξηγηματικές μεταβλητές (σταθεροί κόμβοι – deterministic nodes) παριστάνονται με ένα ορθογώνιο παραλληλόγραμμο και όλες οι άγνωστες ποσότητες καθώς και τα δεδομένα παριστάνονται με μία έλλειψη (στοχαστικοί κόμβοι – stochastic nodes). Αυτά τα σχήματα συνδέονται μεταξύ τους με βέλη, τα οποία είναι δύο ειδών και δείχνουν τη σχέση που τις συνδέει. Το βέλος μονής κατεύθυνσης δείχνει στοχαστική σχέση μεταξύ των συνδεόμενων ποσοτήτων, ενώ ένα βέλος διπλής κατεύθυνσης σημαίνει μία απλή ισότητα. (Lawson, Browne & Rodeiro – 2003, σελ.51)

### 5.6 Υπολογισμός της εκ των υστέρων κατανομής

Στο 1$^{ο}$ Κεφάλαιο έγινε αναφορά για τις θεωρητικές στατιστικές κατανομές, οι οποίες χρησιμοποιούνται συχνότερα για την ανάλυση δεδομένων επιβίωσης, στις περιπτώσεις που ο χρόνος επιβίωσης $T$ είναι διακριτός ή είναι συνεχής.

Όπως αναφέρθηκε, στο WINBUGS πρέπει αρχικά να εισαχθούν οι εκ των προτέρων κατανομές (prior) για κάθε μία από τις άγνωστες παραμέτρους $\theta_i$ *(για* $i = 1, 2, \ldots k$*)* και η συνάρτηση πιθανοφάνειας.

Παρακάτω θα εξετασθούν τ' αποτελέσματα ενός πειράματος που περιλαμβάνει 20 ασθενείς, στον καθένα από τους οποίους μετριέται ο χρόνος (σε μήνες) μέχρι την υποτροπή ενός νοσήματος, μετά από θεραπεία. Οι 20 χρόνοι είναι οι εξής: 11, 5, 12, 7, 7, 9, 10, 11, 19, 8, 20, 19, 12, 14, 16, 18, 9, 12, 21, 13

Τα δεδομένα αναλύονται για κάθε κατανομή ξεχωριστά. Όλες οι εκ των προτέρων κατανομές, για τις παραμέτρους που αναλύονται, έχουν επιλεχτεί έτσι ώστε να έχουν μεγάλη διακύμανση και να εκφράζουν την έλλειψη ισχυρής εκ των προτέρων πληροφορίας (Vague prior – Congdon, 2001).

### 5.6.1 Εκθετική Κατανομή

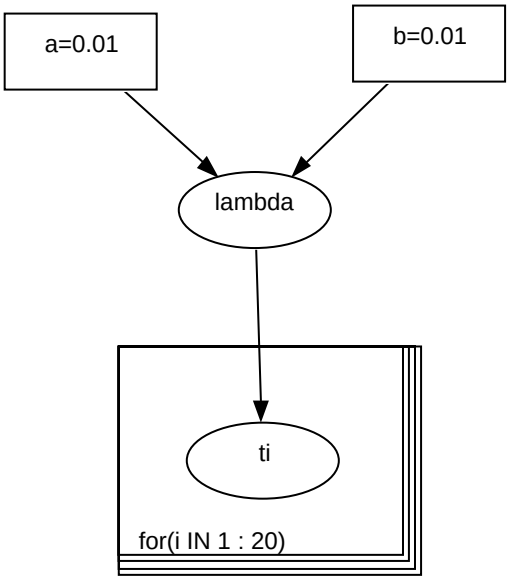
Έστω ότι οι χρόνοι επιβίωσης $t_i \geq 0$ $(i = 1, 2,...20)$, ακολουθούν την εκθετική κατανομή, με παράμετρο $\lambda$. Η εκ των προτέρων κατανομή για την παράμετρο $\lambda$ είναι γνωστό από προηγούμενες έρευνες, ότι είναι η κατανομή Γάμμα, με παραμέτρους $\alpha = 0.01$ και $\beta = 0.01$.

Δηλαδή έχουμε $t \sim exp(\lambda)$ και $\lambda \sim G(0.01, 0.01)$ και οι συναρτήσεις πυκνότητας πιθανότητας για το χρόνο $t$ και την παράμετρο $\lambda$ είναι αντίστοιχα:

$f_T(t \mid \lambda) = \lambda e^{-\lambda t}$ με $t \geq 0$ και

$$\pi(\lambda) = \frac{0.01^{0.01}}{\Gamma(0.01)} \lambda^{0.01-1} e^{-0.01\lambda} \text{ με } \lambda > 0.$$

Το μοντέλο από το Doodle αναπαριστάνεται ως εξής:




**Γράφημα 5.1:** Γραφική αναπαράσταση του απλού εκθετικού μοντέλου, όπου $t \sim exp(\lambda)$ και $\lambda \sim G(0.01, 0.01)$

Αποδεικνύεται πως η εκ των υστέρων κατανομή της $\lambda$ είναι η κατανομή Γάμμα, αφού το συγκεκριμένο μοντέλο είναι συζυγές (conjugate), δηλαδή μπορούμε να υπολογίσουμε την εκ των υστέρων κατανομή της παραμέτρου του, χωρίς την χρήση

Μαρκοβιανών αλυσίδων MCMC και είναι της ίδιας μορφής όπως η εκ των προτέρων κατανομή, δηλαδή Γάμμα.

Έτσι έχουμε ότι η εκ των υστέρων είναι:

$g(\lambda | \underset{\sim}{t}) \sim G\ (n + \alpha,\ \sum_{i=1}^{n} t_i + \beta)$

Πραγματικά:

$$g(\lambda \mid \underset{\sim}{t}) \propto f(t \mid \lambda)\cdot\pi(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} \cdot \lambda^{\alpha-1} e^{-\beta\lambda} = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} t_i\right) \cdot \lambda^{\alpha-1} \exp\left(-\beta\lambda\right)$$

$$g(\lambda \mid \underset{\sim}{t}) \propto \lambda^{n+\alpha-1} \exp\left[-\lambda\left(\sum_{i=1}^{n} t_i + \beta\right)\right]$$

Έτσι η κατανομή της παραμέτρου $\lambda$ μετά τη χρήση των δεδομένων θα είναι η Γάμμα *G(20.01, 253.01)* που θα έχει μέση τιμή $E(\lambda) = \frac{a}{\beta} = 0.079$ και διακύμανση $Var(\lambda) = \frac{a}{\beta^2} = 0.0003125,$ άρα η τυπική απόκλιση θα είναι $\sqrt{0.0003125} = 0.0177$.

Αν υποθέσουμε τώρα, ότι η εκ των προτέρων κατανομή για την παράμετρο $\lambda$ είναι η λογαριθμοκανονική με παραμέτρους $\mu = 1$ και $\tau\eta\varsigma = 0.5$, όπου $\tau = \frac{1}{\sigma^2}$ θα ισχύει:

$f_T(t \mid \lambda) = \lambda e^{-\lambda t}$ με $t \geq 0$ και

$$\pi(\lambda) = \frac{1}{2\lambda\sqrt{\pi}} \exp\left[-\frac{(\log\lambda - 1)^2}{4}\right]$$

Σε αυτή την περίπτωση είναι πιο δύσκολο να υπολογίσουμε αναλυτικά τα περιγραφικά μέτρα της εκ των υστέρων κατανομής της παραμέτρου $\lambda$, αλλά με το

WINBUGS μπορούμε να πάρουμε εύκολα ένα τυχαίο δείγμα από την εκ των υστέρων κατανομή.

Έτσι γράφοντας το μοντέλο στο πρόγραμμα WINBUGS, αυτό θα δίνεται ως εξής:

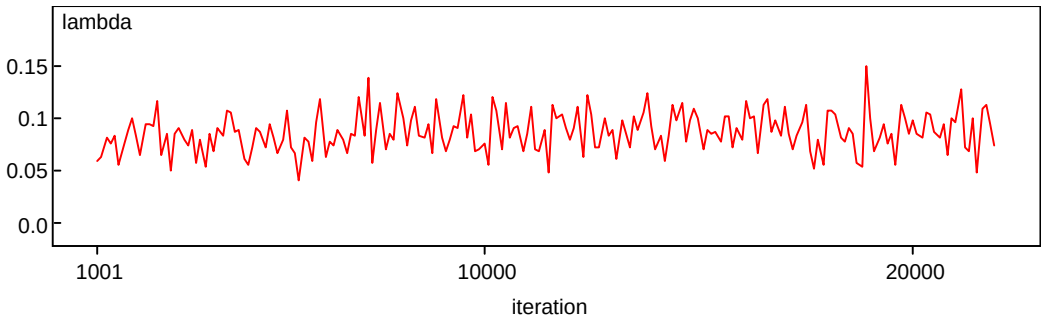
```
MODEL{

 #Prior
 lambda~dlnorm(1, 0.5)

 #Likelihood
 for (I in 1:20) {t[i]~dexp(lambda)}
}

DATA list(
t=c(11, 5, 12, 7, 7, 9, 10, 11, 19, 8, 20, 19, 12, 14, 16, 18, 9, 12, 21, 13))
```

**Πίνακας 5.1:** Μοντέλο εκθετικής κατανομής στο WINBUGS

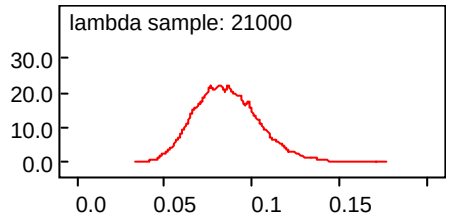
Μετά τη φόρτωση των δεδομένων και χρησιμοποιώντας 21000 προσομοιωμένες τιμές, τα αποτελέσματα για την εκ των υστέρων κατανομή του $\lambda$, δίνονται παρακάτω:

| **node** | **mean** | **sd** | **MC error** | **2.5%** | **median** | **97.5%** | **start** | **sample** |
|---|---|---|---|---|---|---|---|---|
| lambda | 0.086 | 0.0184 | 0.0001359 | 0.0538 | 0.08484 | 0.126 | 1001 | 21000 |

**Πίνακας 5.2:** Περιγραφικοί δείκτες της εκ των υστέρων κατανομής της παραμέτρου $\lambda$




**Γράφημα 5.2:** Διαγραμματική απεικόνιση του ίχνους της παραμέτρου $\lambda$ για εκθετική κατανομή




**Γράφημα 5.3:** Εκτίμηση της εκ των υστέρων συνάρτησης πυκνότητας πιθανότητας της παραμέτρου $\lambda$ για την εκθετική κατανομή

Επίσης το Deviance Information Criterion δίνει: $DIC = 143.484$

Παρατηρούμε πως σε αυτή την περίπτωση η εκ των υστέρων κατανομή της $\lambda$ φαίνεται να προσομοιάζει στην κανονική με μέσο μ = 0.086 και τυπική απόκλιση 0.0184. Συγκρίνοντας τις τιμές αυτής της εκ των υστέρων κατανομής με την εκ των υστέρων $G(20.01,\ 253.01)$ που βρέθηκε προηγουμένως ($E(\lambda) = 0.079$ και $\sqrt{Var(\lambda)} = 0.0177$), παρατηρούμε ότι οι μέσες τιμές των 2 κατανομών, όπως και οι 2 τυπικές αποκλίσεις είναι σχεδόν ίσες, άρα καταλήγουμε στα ίδια αποτελέσματα. Αυτό είναι λογικό, διότι και στις δύο περιπτώσεις χρησιμοποιήσαμε μη πληροφοριακές εκ των προτέρων κατανομές.

### 5.6.2 Κατανομή Weibull

Ας υποθέσουμε ότι οι $20$ χρόνοι επιβίωσης $t_i \geq 0$ $(i = 1, 2, \ldots 20)$, ακολουθούν την κατανομή Weibull, με παραμέτρους $r$ και $\lambda$. Έστω ότι η εκ των προτέρων κατανομή για την $r$ είναι η κατανομή Γάμμα με παραμέτρους $\alpha^* = 0.01$ και $\beta^* = 0.01$, ενώ για την $\lambda$ είναι η εκθετική με παράμετρο $\lambda^* = 0.1$.

Δηλαδή είναι $t \sim Weib(r, \lambda)$, $r \sim G(0.01, 0.01)$ και $\lambda \sim Exp(0.1)$. Άρα οι αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας είναι:

$f_T(t \mid r, \lambda) = r\lambda t^{r-1} e^{-\lambda t^r}$ για $t \geq 0$,

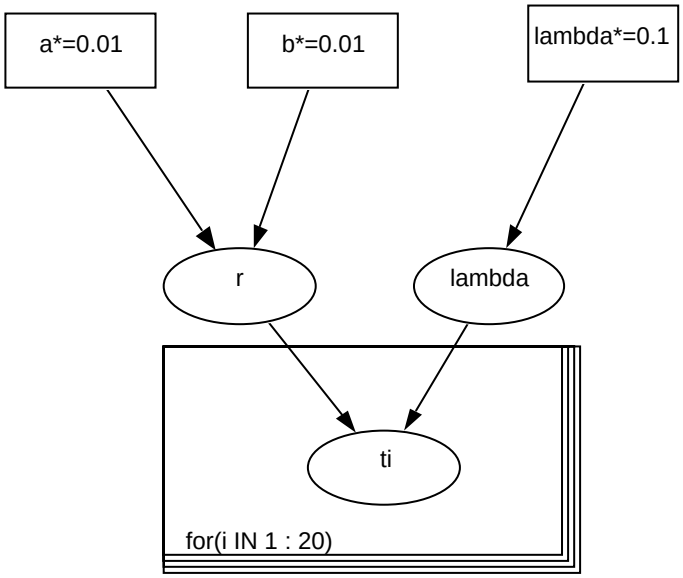
$\pi_1(r) = \dfrac{0.01^{0.01}}{\Gamma(0.01)} r^{0.01-1} e^{-0.01r}$ με $r > 0$ και

$\pi_2(\lambda) = 0.1e^{-0.1\lambda}$ με $\lambda > 0$.

Το μοντέλο στο πρόγραμμα WINBUGS, είναι το εξής:

```
MODEL{
 #Prior
 lambda~dexp(0.1)
 r~dgamma(0.01, 0.01)

 #Likelihood
 for (i in 1:20) {t[i]~dweib(r, lambda)}
}

DATA list(
t=c(11, 5, 12, 7, 7, 9, 10, 11, 19, 8, 20, 19, 12, 14, 16, 18, 9, 12, 21, 13))
```

**Πίνακας 5.3:** Μοντέλο κατανομής Weibull στο WINBUGS

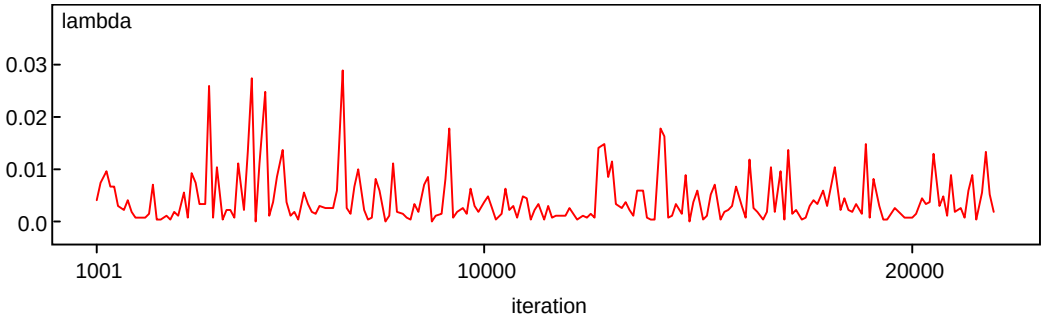



**Γράφημα 5.4:** Γραφική αναπαράσταση του μοντέλου της κατανομής Weibull, όπου *t~Weib(r, λ), r~G(0.01, 0.01)* και *λ~Exp(0.1)*

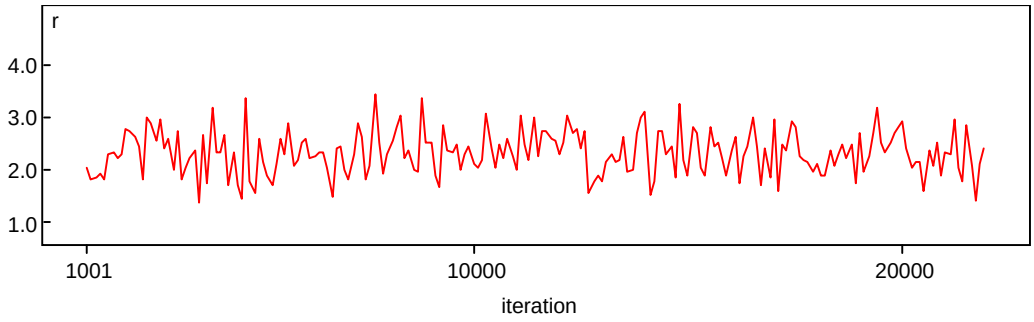
Προκύπτουν τα εξής αποτελέσματα:

| **node** | **mean** | **sd** | **MC error** | **2.5%** | **median** | **97.5%** | **start** | **sample** |
|---|---|---|---|---|---|---|---|---|
| lambda | 0.0043 | 0.0052 | 0.000199 | $1.7 \cdot 10^{-4}$ | 0.002532 | 0.019 | 1001 | 21000 |
| r | 2.324 | 0.4271 | 0.02042 | 1.573 | 2.296 | 3.239 | 1001 | 21000 |

**Πίνακας 5.4:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των παραμέτρων $\lambda$ και $r$

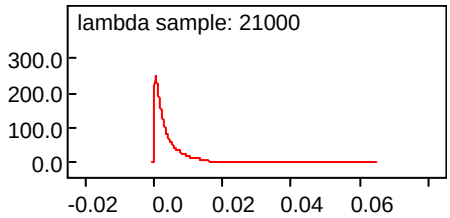



**(5.5)**

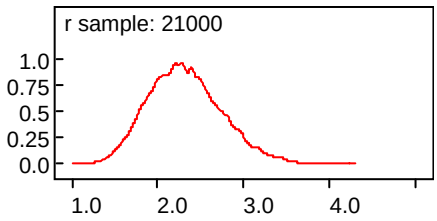



**(5.6)**

**Γραφήματα 5.5 & 5.6:** Διαγραμματική απεικόνιση του ίχνους των παραμέτρων $\lambda$ και $r$ αντίστοιχα για την κατανομή Weibull




**(5.7)**




**(5.8)**

**Γραφήματα 5.7 & 5.8:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων $\lambda$ και $r$ αντίστοιχα για την κατανομή Weibull

Επίσης ισχύει $DIC = 113.968$.

Η εκθετική κατανομή είναι ειδική περίπτωση της κατανομής Weibull και προκύπτει για $r = 1$. Έτσι εξετάζοντας την εκ των υστέρων κατανομή της παραμέτρου $r$, μπορούμε να δούμε αν πρέπει να χρησιμοποιήσουμε την εκθετική ή όχι. Από τον Πίνακα 5.4 βλέπουμε ότι το $1 \notin (1.573, 3.239)$ που είναι το 95% διάστημα αξιοπιστίας για τις τιμές της $r$, άρα η εκ των υστέρων κατανομή μας υποδεικνύει ότι τα δεδομένα δεν ακολουθούν την εκθετική κατανομή.

### 5.6.3 Κατανομή Γάμμα

Ας υποθέσουμε τώρα ότι οι 20 χρόνοι στο πειραμά μας, $t_i \geq 0$ $(i = 1, 2, \ldots 20)$, ακολουθούν την κατανομή Γάμμα, με παραμέτρους $\alpha > 0$ και $\beta > 0$ και ορίσουμε ως εκ των προτέρων κατανομή για την παράμετρο $\alpha$ την εκθετική με παράμετρο $\lambda = 0.01$ και για την παράμετρο $\beta$ είναι η Γάμμα με παραμέτρους $\alpha^* = 0.01$ και $\beta^* = 0.01$. Δηλαδή είναι $t \sim G(\alpha, \beta)$, $\alpha \sim Exp(0.01)$ και $\beta \sim G(0.01, 0.01)$ με αντίστοιχες πυκνότητες πιθανότητας:

$f_T(t \mid \alpha, \beta) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$ για το χρόνο $t > 0$,

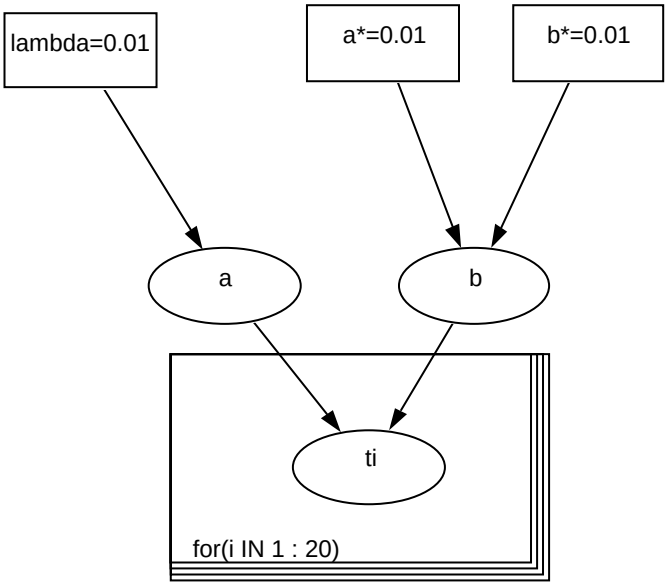
$\pi_1(a) = \dfrac{1}{100} e^{-0.01a}$ για την παράμετρο $\alpha > 0$ και

$\pi_2(\beta) = \dfrac{0.01^{0.01}}{\Gamma(0.01)} \beta^{0.01-1} e^{-0.01\beta}$ για την παράμετρο $\beta > 0$.

Το μοντέλο στο πρόγραμμα WINBUGS, δίνεται από τον παρακάτω κώδικα:

```
MODEL{
 #Prior
 α~dexp(0.01)
 β~dgamma(0.01, 0.01)

 #Likelihood
 for (i in 1:20) {t[i]~dgamma(α, β)}
}

DATA list(
t=c(11, 5, 12, 7, 7, 9, 10, 11, 19, 8, 20, 19, 12, 14, 16, 18, 9, 12, 21, 13))
```

**Πίνακας 5.5:** Μοντέλο κατανομής Γάμμα στο WINBUGS

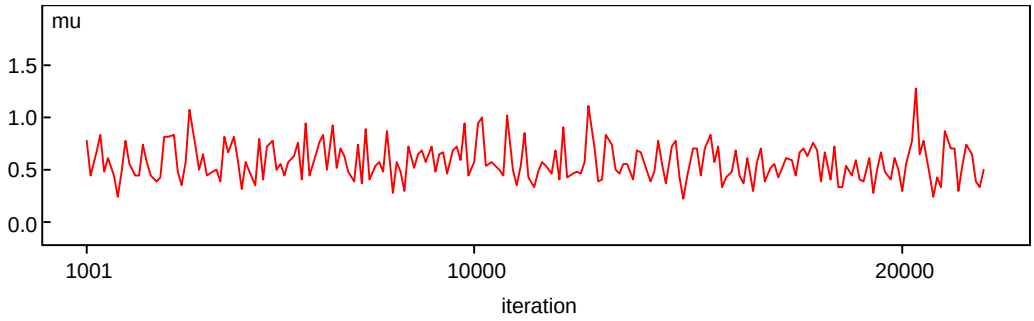



**Γράφημα 5.9:** Γραφική αναπαράσταση του μοντέλου της κατανομής Γάμμα όπου *t~G(α, β), α~Exp(0.01)* και *β~G(0.01, 0.01)*

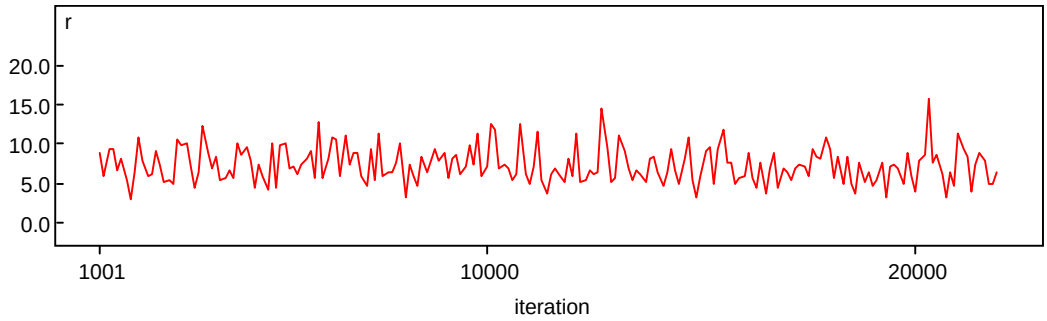
Προκύπτουν τα εξής αποτελέσματα:

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| β | 0.5855 | 0.1863 | 0.006094 | 0.281 | 0.5658 | 1.008 | 1001 | 21000 |
| α | 7.413 | 2.275 | 0.07952 | 3.699 | 7.172 | 12.6 | 1001 | 21000 |

**Πίνακας 5.6:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των παραμέτρων *α* και *β*

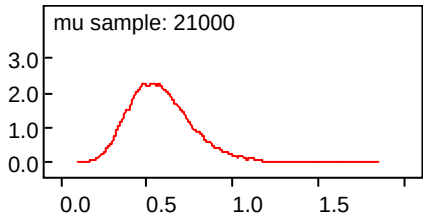



**(5.10)**

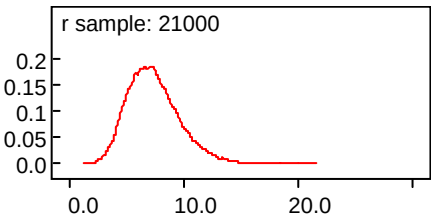



**(5.11)**

**Γραφήματα 5.10 & 5.11:** Διαγραμματική απεικόνιση του ίχνους των παραμέτρων $\alpha$ και $\beta$ αντίστοιχα για την κατανομή Γάμμα




**(5.12)**




**(5.13)**

**Γραφήματα 5.12 & 5.13:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων $\alpha$ και $\beta$ αντίστοιχα για την κατανομή Γάμμα

Επίσης προκύπτει ότι: $DIC = 121.122$

Όμως, η κατανομή Γάμμα είναι και αυτή γενίκευση της εκθετικής κατανομής, η οποία προκύπτει όταν η παράμετρος $\alpha$ είναι ίση με 1. Έτσι από την εκ των υστέρων

κατανομή της $\alpha$, μπορούμε να δούμε αν οι χρόνοι ακολουθούν την εκθετική κατανομή ή όχι. Από τον Πίνακα 5.6 βλέπουμε ότι το $1 \notin (3.699, 12.6)$ που είναι το 95% διάστημα εμπιστοσύνης για τις τιμές της $\alpha$, άρα και οι εκ των υστέρων κατανομές των παραμέτρων της κατανομής Γάμμα μας υποδεικνύουν ότι οι χρόνοι δεν ακολουθούν την εκθετική κατανομή.

### 5.6.4 Λογαριθμοκανονική κατανομή

Αν οι χρόνοι επιβίωσης $t_i \geq 0$ $(i = 1, 2,...20)$ στο πείραμα, ακολουθούν την Λογαριθμοκανονική κατανομή με παραμέτρους $\mu$ και $\sigma^2$, θα ισχύει ότι η μεταβλητή $lnT$ ακολουθεί την Κανονική κατανομή με παραμέτρους $\mu$ και $\sigma^2$.

Επειδή (σύμφωνα με τον Congdon - 2001) η παράμετρος $\sigma^2$ αντικαθιστάται από την $\tau$, όπου $\tau = \frac{1}{\sigma^2}$, η συνάρτηση πυκνότητας πιθανότητας των παρατηρήσεων με $t_i \geq 0$ θα είναι η $f_T(t \mid \mu, \sigma^2) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[ -\frac{1}{2\sigma^2} (\ln t - \mu)^2 \right]$ ή αλλιώς

$$f_T(t \mid \mu, \tau) = \frac{\sqrt{\tau}}{t\sqrt{2\pi}} \exp\left[ -\frac{\tau}{2} (\ln t - \mu)^2 \right].$$

Από προηγούμενες μελέτες, γνωρίζουμε τις κατανομές των δύο παραμέτρων, οι οποίες για την μεν παράμετρο $\mu$ είναι η κανονική κατανομή $N(1,0.01)$ και για την $\tau$ είναι η κατανομή Γάμμα $G(0.01, 0.01)$.

Άρα:

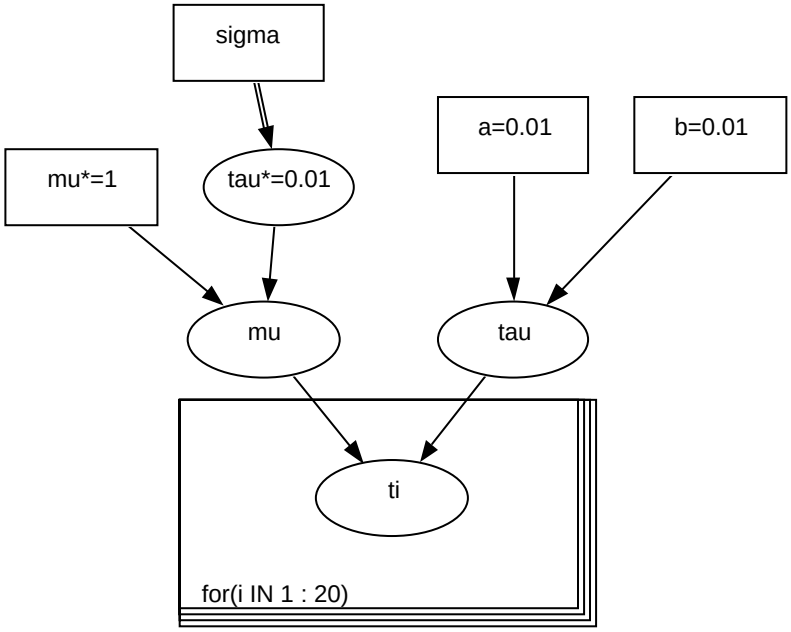
$\pi_1(\mu) = \frac{0.1}{\mu\sqrt{2\pi}} \exp\left[ -\frac{0.01}{2} (\ln \mu - 1)^2 \right]$ και

$\pi_2(\tau) = \frac{0.01^{0.01}}{\Gamma(0.01)} \tau^{0.01-1} e^{-0.01\tau}$, $\tau > 0$.

Το μοντέλο στο πρόγραμμα WINBUGS, είναι το εξής:

```
MODEL{

 #Prior
 mu~dnorm(1, 0.01)
 tau~dgamma(0.01, 0.01)

 #Likelihood
 for (i in 1:20) {t[i]~dlnorm(mu, tau)}
}

DATA list(
t=c(11, 5, 12, 7, 7, 9, 10, 11, 19, 8, 20, 19, 12, 14, 16, 18, 9, 12, 21, 13))
```

**Πίνακας 5.7:** Μοντέλο Λογαριθμοκανονικής κατανομής στο WINBUGS

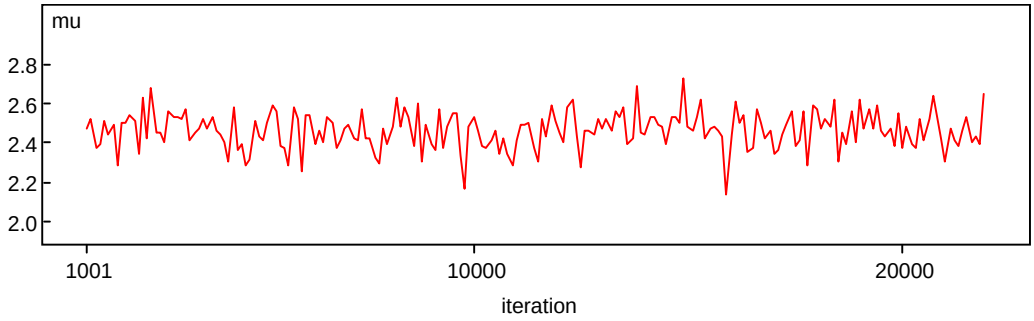



**Γράφημα 5.14:** Γραφική αναπαράσταση του μοντέλου της Λογαριθμοκανονικής κατανομής, όπου $t \sim \Lambda(\mu, \tau)$, $\mu \sim N(1, 0.01)$ και $\tau \sim G(0.01, 0.01)$

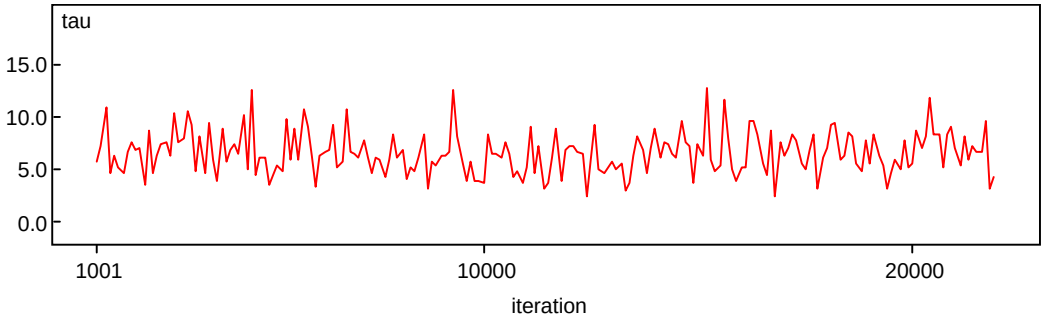
Με δείγμα προσομοιωμένων τιμών μεγέθους 21000, προκύπτουν τα παρακάτω αποτελέσματα:

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| mu | 2.466 | 0.0928 | 0.000581 | 2.283 | 2.465 | 2.651 | 1001 | 21000 |
| tau | 6.324 | 2.056 | 0.01472 | 2.985 | 6.094 | 10.97 | 1001 | 21000 |

**Πίνακας 5.8:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των παραμέτρων $\mu$ και $\tau$

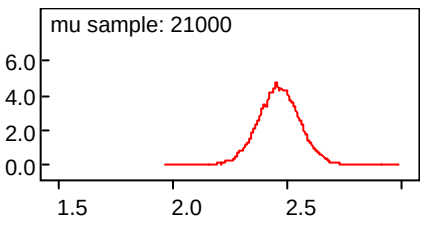



**(5.15)**

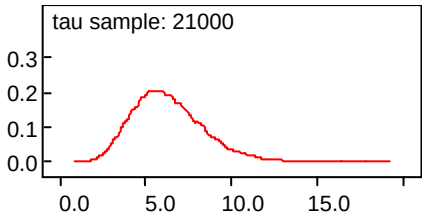



**(5.16)**

**Γραφήματα 5.15 & 5.16:** Διαγραμματική απεικόνιση του ίχνους των παραμέτρων $\mu$ και $\tau$ αντίστοιχα για την Λογαριθμοκανονική κατανομή




**(5.17)**




**(5.18)**

**Γραφήματα 5.17 & 5.18:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων $\mu$ και $\tau$ αντίστοιχα για την Λογαριθμοκανονική κατανομή

Επίσης ισχύει $DIC = 121.547$

### 5.6.5 Κατανομή Paretto

Έστω ότι οι χρόνοι επιβίωσης $t_i \geq 0$ $(i = 1, 2,...20)$ ακολουθούν την κατανομή Paretto με παραμέτρους $\alpha > 0$ και $c > 0$. Όμως πρέπει $t_i \geq c$ για κάθε $i = 1, 2,...20$ και από τα δεδομένα έχουμε ότι $t_{min} = 5$, άρα η μεγαλύτερη δυνατή τιμή της παραμέτρου $c$ θα είναι η τιμή 5 και η μικρότερη το 0, δηλαδή $0 < c \leq 5$.

Αν υποθέσουμε ότι η εκ των προτέρων κατανομή της παραμέτρου $\alpha$ είναι η κανονική με παραμέτρους $\mu = 1$ και $\sigma^2 = 100$ και η αντίστοιχη για την παράμετρο $c$ είναι η κατανομή $G(0.01, 0.01)$.

Δηλαδή:

$$f_T(t \mid \alpha, c) = a\frac{c^a}{t^{a+1}} \text{ με } t \geq c$$

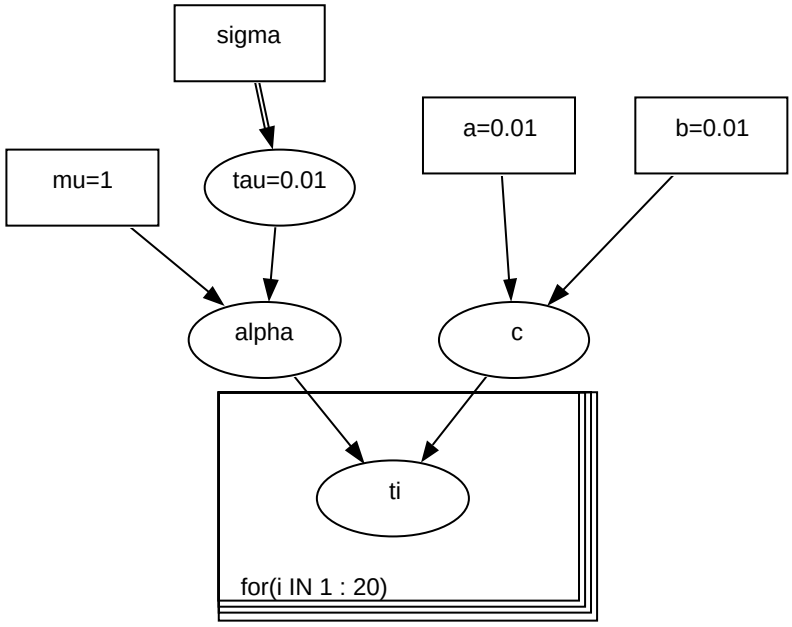
$$\pi_1(a) = \frac{1}{10\sqrt{2\pi}}\exp\left[-\frac{(t-1)^2}{200}\right] \text{ με } \alpha > 0 \text{ και}$$

$$\pi_2(c) = \frac{0.01^{0.01}}{\Gamma(0.01)}c^{0.01-1}\,\mathrm{e}^{-0.01c} \text{ με } 0 < c < t_i \ (i = 1, 2,...20)$$

Το μοντέλο τότε είναι το εξής:

```
MODEL{

 #Prior
 alpha~dnorm(1, 0.01)
 c~dgamma(0.01, 0.01)I(0, 5)

 #Likelihood
 for (i in 1:20) {t[i]~dpar(alpha, c)}
}

DATA list(
t=c(11, 5, 12, 7, 7, 9, 10, 11, 19, 8, 20, 19, 12, 14, 16, 18, 9, 12, 21, 13))
```

**Πίνακας 5.9:** Μοντέλο κατανομής Paretto στο WINBUGS

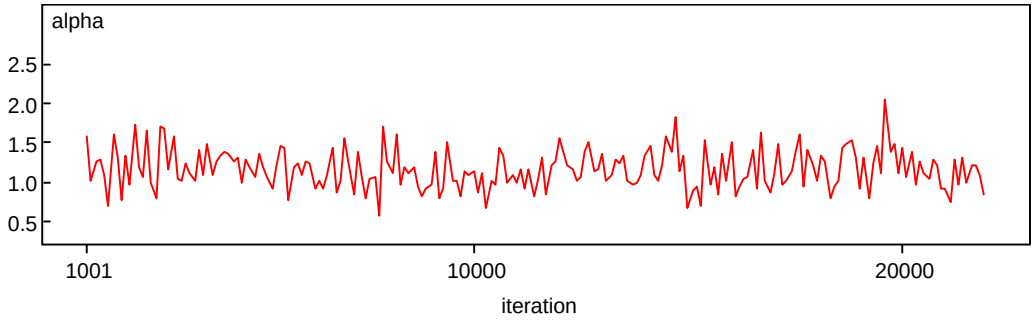



**Γράφημα 5.19:** Γραφική αναπαράσταση του μοντέλου της κατανομής Paretto, όπου *t~Paretto(α, c), α~N(1, 100)* και *c~G(0.01, 0.01)*

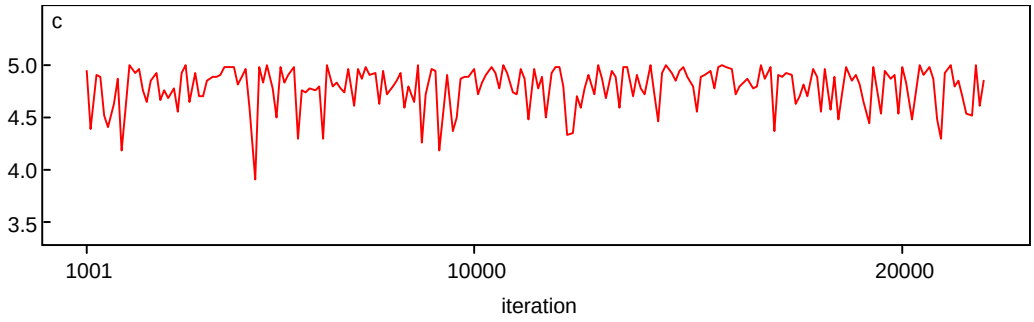
Προκύπτουν τα εξής αποτελέσματα:

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| alpha | 1.167 | 0.2617 | 0.00208 | 0.7159 | 1.148 | 1.727 | 1001 | 21000 |
| c | 4.783 | 0.2153 | 0.002593 | 4.201 | 4.849 | 4.995 | 1001 | 21000 |

**Πίνακας 5.10:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των *α* και *c*

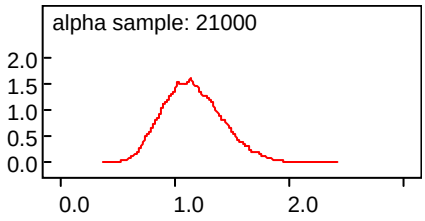



**(5.20)**

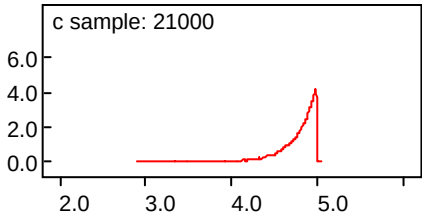



**(5.21)**

**Γραφήματα 5.20 & 5.21:** Διαγραμματική απεικόνιση του ίχνους των παραμέτρων *α* και *c* αντίστοιχα για την κατανομή Paretto




**(5.22)**




**(5.23)**

**Γραφήματα 5.22 & 5.23:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων *α* και *c* αντίστοιχα για την κατανομή Paretto

Επίσης ισχύει ότι: $DIC = 136.504$

### 5.6.6 Σύγκριση μοντέλων με τη χρήση του *DIC*

Σε όλα τα παραπάνω μοντέλα έχουν υπολογιστεί οι εκ των υστέρων κατανομές των παραμέτρων με την χρήση Μαρκοβιανών αλυσίδων MCMC και από τα σχετικά διαγράμματα προσομοιωμένων τιμών, οι αλυσίδες που κατασκευάστηκαν φαίνεται πως συγκλίνουν. Σε κάθε περίπτωση έχει υπολογιστεί το στατιστικό *DIC*, το οποίο είδαμε πως ισούται με $-2\ln L\left(t \mid \underset{\sim}{\theta}^*\right) + 2p_D$, με $p_D$: ο αριθμός των "αποδοτικών" παραμέτρων.

Έτσι για κάθε κατανομή έχουμε:

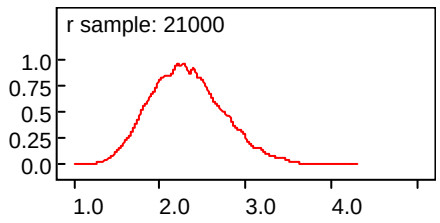
| **Κατανομή** | **Dbar** | **Dhat** | **$p_D$** | **DIC** |
|---|---|---|---|---|
| Εκθετική | 142.568 | 141.651 | 0.916 | 143.484 |
| Weibull | 120.990 | 128.012 | -7.022 | 113.968 |
| Γάμμα | 119.048 | 116.973 | 2.075 | 121.122 |
| Λογαριθμοκανονική | 119.495 | 117.443 | 2.052 | 121.547 |
| Paretto | 135.538 | 134.573 | 0.966 | 136.504 |

**Πίνακας 5.11:** Συνοπτικός πίνακας σύγκρισης κατανομών με τη χρήση του *DIC*

Το καταλληλότερο με βάση το μικρότερο *DIC* είναι το μοντέλο που υποθέτει ότι οι χρόνοι επιβίωσης ακολουθούν την κατανομή Weibull.

Οι εκ των υστέρων κατανομές για τις 2 παραμέτρους του μοντέλου, σύμφωνα με τον Πίνακα 5.4, έχουν:

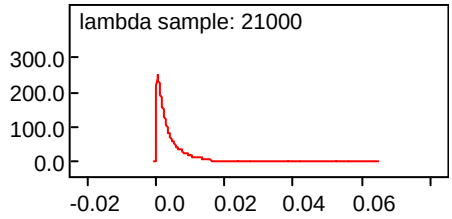
- Για την παράμετρο *r*: Μέση τιμή 2.324, τυπική απόκλιση 0.4271, διάμεσο 2.296 και 95% διάστημα εμπιστοσύνης το (1.573, 3.239) με MC σφάλμα 0.02042.




**Γράφημα 5.24:** Εκτίμηση της εκ των υστέρων συνάρτησης πυκνότητας πιθανότητας της παραμέτρου *r*

Κρίνοντας και από το γράφημα 5.24 η εκ των υστέρων κατανομή για την $r$ μπορεί να προσεγγιστεί ικανοποιητικά από μια κατανομή με μέσο $\mu$ = 2.3 και διακύμανση $\sigma^2$ = 0.18 (με ακρίβεια ± 0.02).

- Για την παράμετρο $\lambda$: Μέση τιμή 0.0043, τυπική απόκλιση 0.0052, διάμεσο 0.0025 και 95% διάστημα εμπιστοσύνης το (0.00017, 0.019) με MC σφάλμα 0.0002.




**Γράφημα 5.25:** Εκτίμηση της εκ των υστέρων συνάρτησης πυκνότητας πιθανότητας της παραμέτρου $\lambda$

### 5.7 Προσέγγιση του μοντέλου Cox (PH model) κατά Bayes

Πάρα πολλοί ερευνητές, όπως οι Kalbfleisch (1978), Prentice (1980), Clayton(1994) έχουν ασχοληθεί με την προσέγγιση του μοντέλου Cox:

$$h_T(t \mid \underset{\sim}{X}) = h_0(t)\exp\left(b_1X_1 + b_2X_2 + \ldots b_pX_p\right),$$

το οποίο ισοδύναμα είναι το:

$$\ln h_T(t \mid \underset{\sim}{X}) = \ln h_0(t) + b_1X_1 + b_2X_2 + \ldots + b_pX_p.$$

Ενώ ο Kalbfleisch μελέτησε το παραπάνω μοντέλο χρησιμοποιώντας ως επίπεδο αναφοράς της συνάρτησης κινδύνου $h_0(t)$, την κατανομή Weibull με παραμέτρους $r$ και $\lambda = 1$, ο Clayton το 1994 εκτίμησε την παραπάνω συνάρτηση με τη χρήση Μαρκοβιανών αλυσίδων.

Παράδειγμα 5.7:

Το μοντέλο μελετήθηκε πάνω σε δεδομένα που αφορούν 42 ασθενείς. Οι ασθενείς χωρίστηκαν σε 2 ομάδες των 21 ατόμων, στην κάθε μία από της οποίες χορηγήθηκε μία διαφορετική θεραπεία (το placebo και η θεραπεία 6-MP) και για τον κάθε έναν ασθενή μετρήθηκε ο χρόνος $t_i$ *(i = 1, 2…42)* μέχρι το θάνατό του.

Οι χρόνοι είναι οι εξής:

Με το placebo: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Με τη θεραπεία 6-MP: $6^*$, 6, 6, 6, 7. 9, $10^*$, 10, $11^*$, 13, 16, $17^*$, $19^*$, $20^*$, 22, 23, $25^*$, $32^*$, $32^*$, $34^*$, $35^*$

(Με το * συμβολίζονται οι περικομμένοι χρόνοι)

Για κάθε *i = 1, 2…42* παρατηρείται ο αριθμός των θανάτων $N_i(t)$ που έχει συμβεί μέχρι το χρόνο *t*. Έστω $dN_i(t)$ η αύξηση του $N_i(t)$ μέσα στο πολύ μικρό διάστημα *[t, t + dt)*. Τα $dN_i(t)$ είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson.

Ισχύει ότι:

$dN_i(t) = 1$, αν ο *I* ασθενής πεθάνει εντός του διαστήματος *[t, t + dt)*

$dN_i(t) = 0$, αν όχι

Η αναμενόμενη αύξηση των θανάτων θα είναι:

$I_i(t) = Y_i(t) \cdot h_0(t) \exp(bZ_i)$, όπου η μεταβλητή $Y_i(t)$ δείχνει αν ο ασθενής *I* είναι ζωντανός στο χρόνο πριν το διάστημα *[t, t + dt)*.

Ισχύει ότι:

$Y_i(t) = 1$, αν ο *I* ασθενής είναι ζωντανός πριν το διάστημα *[t, t + dt)*

$Y_i(t) = 0$, αν όχι

Συνεπώς, η αύξηση των θανάτων $dN_i(t)$ μέσα στο διάστημα *[t, t + dt)* έχουν μέση τιμή $I_i(t)dt$, δηλαδή $dN_i(t) \sim \text{Poisson}(I_i(t)dt)$.

Στο παραπάνω μοντέλο θέλουμε να εκτιμήσουμε την παράμετρο *b*, δηλαδή τον συντελεστή που αντιστοιχεί στην επεξηγηματική μεταβλητή *Z*, που παριστάνει το είδος

της θεραπείας και την $\Lambda_0(t) = \int_0^t h_0(u)du$, δηλαδή την αθροιστική συνάρτηση κατανομής του επιπέδου αναφοράς της συνάρτησης κινδύνου $h_0(t)$.

Με τη χρήση του θεωρήματος Bayes υπολογίζουμε την εκ των υστέρων πιθανότητα των παραμέτρων, η οποία δίνεται από τον τύπο:

$$P\left(b, \Lambda_0(t) \mid D\right) \propto P\left(D \mid b, \Lambda_0(t)\right) \cdot \Pi\left(b, \Lambda_0(t)\right),$$

όπου $\Pi\left(b, \Lambda_0(t)\right)$ η αντίστοιχη εκ των προτέρων και $D$ τα δεδομένα. Οι $b$, $\Lambda_0(t)$ είναι της ανεξάρτητες, με εκ των προτέρων πιθανότητες $\Pi_1(b)$ και $\Pi_2\left(\Lambda_0(t)\right)$, άρα θα ισχύει:

$$P\left(b, \Lambda_0(t) \mid D\right) \propto P\left(D \mid b, \Lambda_0(t)\right) \cdot \Pi_1(b)\Pi_2\left(\Lambda_0(t)\right)$$

Αφού οι αυξήσεις των θανάτων $dN_i(t)$ ακολουθούν την κατανομή Poisson($I_i(t)dt$) και αφού η εκ των προτέρων κατανομή για την μέση τιμή της κατανομής Poisson είναι η κατανομή Γάμμα (conjugate prior), σύμφωνα με τον Kalbfleisch (1978) θα ήταν λογικό να υποτεθεί ότι οι αυξήσεις $d\Lambda_0(t)$ ακολουθούν την κατανομή Γάμμα με παραμέτρους $cd\Lambda_0^*(t)$ και $c$. Άρα $d\Lambda_0(t) \sim$ Γάμμα ($cd\Lambda_0^*(t)$, $c$)

Ο αριθμός $d\Lambda_0^*(t)$ εκφράζει την εικασία που κάνουμε για την άγνωστη συνάρτηση κινδύνου και η παράμετρος $c$ εκφράζει το βαθμό εμπιστοσύνης που δίνουμε σ' αυτή την εικασία.

Η μέση τιμή της παραπάνω εκ των προτέρων κατανομής Γάμμα δίνεται ως $E\left[d\Lambda_0(t)\right] = \frac{cd\Lambda_0^*(t)}{c} = d\Lambda_0^*(t)$ και η διακύμανσή της ως $Var\left[d\Lambda_0(t)\right] = \frac{cd\Lambda_0^*(t)}{c^2} = \frac{d\Lambda_0^*(t)}{c}$. Άρα για μικρές τιμές του $c$ η διακύμανση αυξάνεται και αυτό σημαίνει "ασθενείς" εκ των προτέρων πεποιθήσεις.

Όσον αφορά την εκ των προτέρων κατανομή της παραμέτρου $b$, αυτή την λαμβάνουμε ως κανονική κατανομή με μέση τιμή 0 και πολύ μεγάλη διακύμανση (vague prior), για να υποδηλώσουμε ασθενή εκ των προτέρων γνώση για την παράμετρο, δηλαδή $b \sim N(0, 10^{-6})$.

Ο κώδικας WINBUGS για το παραπάνω μοντέλο είναι ο ακόλουθος:

```
MODEL{
        # Set up data
                for(i in 1:N) {
                        for(j in 1:T) {
        # risk set = 1 if obs.t >= t
                                Y[i,j] <- step(obs.t[i] – t[j] + eps)
        # counting process jump = 1 if obs.t in [ t[j], t[j+1] )
        #               i.e. if t[j] <= obs.t < t[j+1]
                                dN[i, j] <- Y[i, j] * step(t[j + 1] – obs.t[i] – eps) * fail[i]
                        }
                }
        # Model
                for(j in 1:T) {
                        for(i in 1:N) {
                                dN[i, j]   ~ dpois(Idt[i, j])           # Likelihood
                                Idt[i, j] <- Y[i, j] * exp(beta * Z[i]) * dL0[j]  # Intensity
                        }
                        dL0[j] ~ dgamma(mu[j], c)
                        mu[j] <- dL0.star[j] * c    # prior mean hazard

        # Survivor function = exp(-Integral{l0(u)du})^exp(beta*z)
                        S.treat[j] <- pow(exp(-sum(dL0[1 : j])), exp(beta * -0.5));
                        S.placebo[j] <- pow(exp(-sum(dL0[1 : j])), exp(beta * 0.5));

                }
                c <- 0.001
                r <- 0.1
                for (j in 1 : T) {  dL0.star[j] <- r * (t[j + 1] – t[j])  }
                beta ~ dnorm(0.0,0.000001)
  }
DATA
list(N = 42, T = 17, eps = 1.0E-10,
obs.t = c(1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23, 6, 6, 6, 6, 7, 9,
10, 10, 11, 13, 16, 17, 19, 20, 22, 23, 25, 32, 32, 34, 35),
fail = c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1,
1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0),
Z = c(0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5,
0.5, 0.5, 0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5,
-0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5),
t = c(1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 22, 23, 35))

INITIAL VALUES
list( beta = 0.0,
dL0 = c(1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0, 1.0,1.0))
```

**Πίνακας 5.12:** Κώδικας WINBUGS για το μοντέλο του Cox

Όπως βλέπουμε, για να οριστεί σωστά το μοντέλο στο WINBUGS ορίζουμε 2 μεταβλητές που αφορούν τους χρόνους θανάτων. Η μεταβλητή *t.obs* περιλαμβάνει όλους τους παρατηρούμενους χρόνους των 42 ασθενών (περικομμένους και μη), ενώ η μεταβλητή *t* περιλαμβάνει (κατά αύξουσα σειρά) τους 17 διακριτούς παρατηρούμενους μη περικομμένους χρόνους θανάτου και τη μέγιστη παρατηρούμενη τιμή που είναι ίση με 35.

Από τις 18 τιμές της μεταβλητής *t* κατασκευάζονται 17 διαστήματα, σε καθένα από τα οποία μετριέται η αύξηση των θανάτων *dN[i, j]* με *i =1, 2,...42* και *j = 1, 2,...17* .

Άρα η ποσότητα *dN[i, j]* με τιμές 0 και 1 δείχνει αν ο ασθενής *i* πεθαίνει κατά τη διάρκεια του διαστήματος *[j, j+1)*. Ισχύει ότι:

*dN[i, j] = 1,* αν ο *i* ασθενής πεθάνει εντός του διαστήματος *[j, j+1)*

*dN[i, j] = 0*, αν όχι

Στο διάστημα *[0, 1)* εννοείται πως όλοι οι ασθενείς είναι ζωντανοί κι έτσι το συγκεκριμένο διάστημα δεν λαμβάνεται καθόλου υπόψη.

Υπάρχουν ακόμη 2 μεταβλητές. Η μεταβλητή *fail* που δείχνει αν ένας χρόνος αναφέρεται σε θάνατο ή περικοπή, δηλαδή:

$$fail = \begin{cases} 1, & \textit{για πραγματικούς χρόνους} \\ 0, & \textit{για περικομμένους χρόνους} \end{cases}$$

και η επεξηγηματική μεταβλητή

$$Z = \begin{cases} 0.5, & \textit{για τα άτομα που έλαβαν ψευδοφάρμακο} \\ -0.5, & \textit{για τα άτομα που έλαβαν πραγματική θεραπεία } (6-\text{MP}) \end{cases}$$

Η ποσότητα *Y[i, j]= obs.t[i] - t[j] + eps* δείχνει αν ο ασθενής *i* είναι ζωντανός στο χρόνο πριν το διάστημα *[j, j+1)*. Αν η ποσότητα αυτή δεν είναι αρνητική παίρνει την τιμή 1, αλλιώς παίρνει την τιμή 0. Ισχύει ότι:

*Y[i, j] = 1,* αν ο *i* ασθενής είναι ζωντανός στην αρχή του διαστήματος *[j, j+1)*

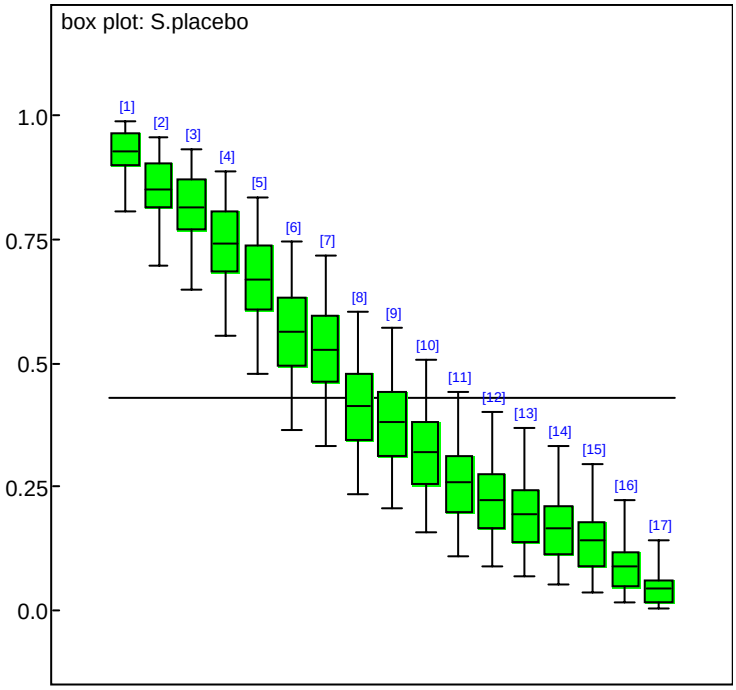
*Y[i, j] = 0*, αν ο *i* ασθενής έχει πεθάνει ή έχει αποχωρήσει στην αρχή του διαστήματος *[j, j+1)*

Στον πίνακα που ακολουθεί δίνονται οι περιγραφικοί δείκτες των εκ των υστέρων κατανομών της συνάρτησης επιβίωσης για κάθε διακριτή παρατηρούμενη τιμή και για τις 2 ομάδες του συγκεκριμένου παραδείγματος, έπειτα από 11000 επαναλήψεις, αφού αφαιρεθούν οι αρχικές 1000 ως burn – in περίοδος.

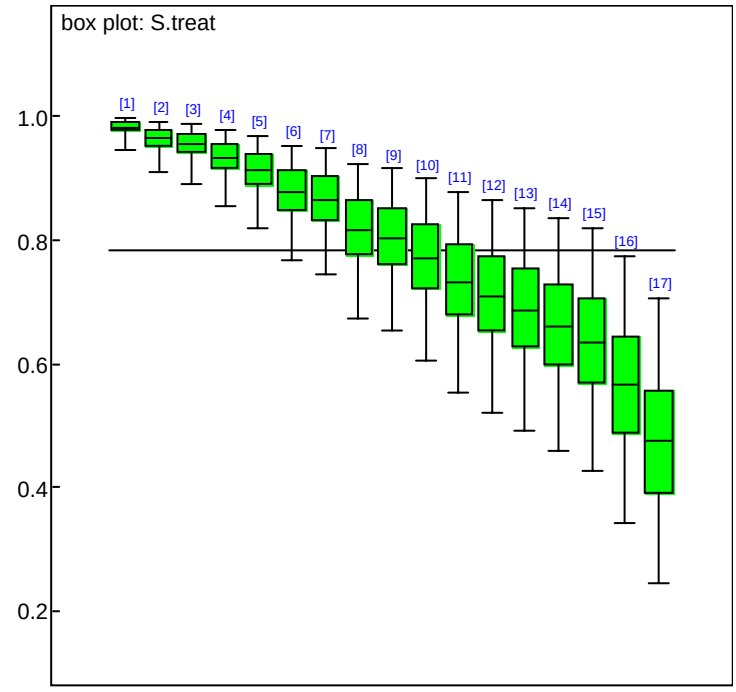
| **Παράμετρος** | **Εκ των υστέρων μέση τιμή** | **Τυπική Απόκλιση** | **Σφάλμα Monte Carlo** | **2.5% εκ των υστέρων Δ.Ε.** | **Διάμεσος** | **97.5% εκ των υστέρων Δ.Ε.** |
|---|---|---|---|---|---|---|
| *S.placebo[1]* | 0.925 | 0.051 | $5.02 \cdot 10^{-4}$ | 0.798 | 0.937 | 0.99 |
| *S.placebo[2]* | 0.851 | 0.068 | $6.33 \cdot 10^{-4}$ | 0.692 | 0.861 | 0.958 |
| *S.placebo[3]* | 0.814 | 0.074 | $6.82 \cdot 10^{-4}$ | 0.647 | 0.822 | 0.936 |
| *S.placebo[4]* | 0.741 | 0.184 | $7.56 \cdot 10^{-4}$ | 0.557 | 0.748 | 0.885 |
| *S.placebo[5]* | 0.667 | 0.091 | $9.08 \cdot 10^{-4}$ | 0.473 | 0.673 | 0.831 |
| *S.placebo[6]* | 0.56 | 0.097 | $8.4 \cdot 10^{-4}$ | 0.364 | 0.562 | 0.739 |
| *S.placebo[7]* | 0.526 | 0.097 | $8.7 \cdot 10^{-4}$ | 0.333 | 0.527 | 0.711 |
| *S.placebo[8]* | 0.411 | 0.094 | $8.23 \cdot 10^{-4}$ | 0.236 | 0.409 | 0.601 |
| *S.placebo[9]* | 0.378 | 0.093 | $8.07 \cdot 10^{-4}$ | 0.202 | 0.375 | 0.566 |
| *S.placebo[10]* | 0.318 | 0.09 | $8.05 \cdot 10^{-4}$ | 0.157 | 0.313 | 0.507 |
| *S.placebo[11]* | 0.256 | 0.085 | $6.88 \cdot 10^{-4}$ | 0.111 | 0.249 | 0.441 |
| *S.placebo[12]* | 0.223 | 0.082 | $6.87 \cdot 10^{-4}$ | 0.086 | 0.215 | 0.399 |
| *S.placebo[13]* | 0.194 | 0.078 | $6.72 \cdot 10^{-4}$ | 0.066 | 0.185 | 0.365 |
| *S.placebo[14]* | 0.164 | 0.073 | $6.82 \cdot 10^{-4}$ | 0.049 | 0.154 | 0.326 |
| *S.placebo[15]* | 0.138 | 0.067 | $6.1 \cdot 10^{-4}$ | 0.036 | 0.128 | 0.295 |
| *S.placebo[16]* | 0.085 | 0.054 | $4.75 \cdot 10^{-4}$ | 0.013 | 0.073 | 0.217 |
| *S.placebo[17]* | 0.043 | 0.038 | $3.89 \cdot 10^{-4}$ | 0.002 | 0.032 | 0.143 |
| *S.treat[1]* | 0.982 | 0.014 | $1.67 \cdot 10^{-4}$ | 0.945 | 0.976 | 0.998 |
| *S.treat[2]* | 0.964 | 0.021 | $2.58 \cdot 10^{-4}$ | 0.909 | 0.968 | 0.992 |
| *S.treat[3]* | 0.954 | 0.025 | $3.11 \cdot 10^{-4}$ | 0.891 | 0.959 | 0.988 |
| *S.treat[4]* | 0.934 | 0.032 | $4.25 \cdot 10^{-4}$ | 0.856 | 0.94 | 0.979 |
| *S.treat[5]* | 0.912 | 0.039 | $5.61 \cdot 10^{-4}$ | 0.82 | 0.918 | 0.97 |
| *S.treat[6]* | 0.876 | 0.049 | $7 \cdot 10^{-4}$ | 0.762 | 0.883 | 0.953 |

| **Παράμετρος** | **Εκ των υστέρων μέση τιμή** | **Τυπική Απόκλιση** | **Σφάλμα Monte Carlo** | **2.5% εκ των υστέρων Δ.Ε** | **Διάμεσος** | **97.5% εκ των υστέρων Δ.Ε** |
|---|---|---|---|---|---|---|
| *S.treat[7]* | 0.864 | 0.052 | $7.7 \cdot 10^{-4}$ | 0.74 | 0.872 | 0.947 |
| *S.treat[8]* | 0.817 | 0.065 | $9.48 \cdot 10^{-4}$ | 0.672 | 0.825 | 0.922 |
| *S.treat[9]* | 0.802 | 0.068 | 0.0011 | 0.649 | 0.809 | 0.914 |
| *S.treat[10]* | 0.771 | 0.076 | 0.0012 | 0.601 | 0.779 | 0.9 |
| *S.treat[11]* | 0.734 | 0.085 | 0.0012 | 0.548 | 0.742 | 0.877 |
| *S.treat[12]* | 0.711 | 0.089 | 0.0012 | 0.521 | 0.719 | 0.864 |
| *S.treat[13]* | 0.688 | 0.094 | 0.0013 | 0.488 | 0.696 | 0.851 |
| *S.treat[14]* | 0.662 | 0.098 | 0.0014 | 0.456 | 0.668 | 0.837 |
| *S.treat[15]* | 0.636 | 0.102 | 0.0014 | 0.424 | 0.642 | 0.82 |
| *S.treat[16]* | 0.566 | 0.112 | 0.0015 | 0.34 | 0.569 | 0.777 |
| *S.treat[17]* | 0.476 | 0.12 | 0.0015 | 0.244 | 0.476 | 0.709 |
| *b* | 1.546 | 0.418 | 0.0053 | 0.765 | 1.533 | 2.395 |

**Πίνακας 5.13**: Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των εκτιμώμενων χρόνων επιβίωσης για τις ομάδες που έλαβαν ψευδοφάρμακο (*S.placebo[j]),* και ενεργή θεραπεία (*S.treat[j])* για *j = 1, 2,….17* και της παραμέτρου *b*




**(5.26):** Ψευδοφάρμακο




**(5.27):** Ενεργή θεραπεία

**Γραφήματα 5.26 & 5.27:** Γραφική παράσταση των εκτιμώμενων χρόνων επιβίωσης για τις ομάδες που έλαβαν ψευδοφάρμακο (*S.placebo[j]*) και ενεργή θεραπεία (*S.treat[j]*)

Η εκ των υστέρων μέση τιμή της παραμέτρου *b* είναι 1.546 με τυπική απόκλιση 0.42. Η τιμή αυτή δείχνει ότι η αναμενόμενη αύξηση των θανάτων των ασθενών που λάμβαναν το ψευδοφάρμακο είναι μεγαλύτερη κατά 55% από την αντίστοιχη αναμενόμενη αύξηση θανάτων με την πραγματική θεραπεία 6-MP.

Επίσης, από τα παραπάνω παίρνουμε εκτιμήσεις για τη συνάρτηση επιβίωσης $S(t) = \exp\left[-\int_0^t h_0(u)\exp(bZ_i)du\right]$ σε κάθε ένα από τα διαστήματα *[j, j+1)* με *j =1, 2,...17* και για κάθε μία θεραπεία ξεχωριστά. Δηλαδή λαμβάνεται εκτίμηση για *S.placebo[j], S.treat[j]* που δείχνουν την πιθανότητα επιβίωσης μέχρι το χρόνο $t_j$ για τα άτομα που λαμβάνουν ψευδοφάρμακο και πραγματική θεραπεία αντίστοιχα.

Για παράδειγμα βλέπουμε πως *S.placebo[6] < S.treat[6],* το οποίο σημαίνει ότι η πιθανότητα επιβίωσης εντός του διαστήματος $[t_6, t_7)$ είναι μικρότερη για τους ασθενείς που λαμβάνουν ψευδοφάρμακο (56%) από την αντίστοιχη πιθανότητα για τους ασθενείς που λαμβάνουν πραγματική θεραπεία (87.6%)

Παρατηρούμε ότι σε κάθε τέτοιο διάστημα *[j, j+1)* η συνάρτηση επιβίωσης *S(t)* έχει καλύτερες τιμές για την θεραπεία 6-MP απ' ότι για το placebo. Δηλαδή η θεραπεία 6-MP δίνει καλύτερα αποτελέσματα από το placebo.

### 5.8 Το παραμετρικό μοντέλο βασισμένο στην κατανομή Weibull

Η συνάρτηση κινδύνου $h_T(t)$ , $t > 0$ μπορεί να εκτιμηθεί χρησιμοποιώντας παραμετρικές μεθόδους εκτίμησης.

Το μοντέλο που χρειάζεται για την εκτίμηση της $h_T(t)$ είναι το γενικό μοντέλο με τη μέθοδο Cox, δηλαδή:

$$h_T(t \mid X) = h_0(t)\exp\left(b_1X_1 + b_2X_2 + \ldots b_pX_p\right),$$

όπου $b_1, b_2, \ldots b_p$ είναι οι συντελεστές που αντιστοιχούν σε κάθε μία από τις ανεξάρτητες μεταβλητές $X_1, \ldots, X_p$ το οποίο ισοδύναμα είναι το:

$$\ln h_T(t \mid X) = \ln h_0(t) + b_1X_1 + b_2X_2 + \ldots + b_pX_p.$$

Στην προκειμένη περίπτωση χρησιμοποιείται η κατανομή Weibull με παραμέτρους $r$ και $\lambda = 1$, της οποίας η συνάρτηση κινδύνου είναι η:

$$h_0(t) = rt^{r-1}$$

Έτσι το μοντέλο παλινδρόμησης γίνεται:

$$h_T(t \mid X) = rt^{r-1} \exp\left(b_1 X_1 + b_2 X_2 + \ldots b_p X_p\right),$$

όπου *HR* είναι ο λόγος στιγμιαίου κινδύνου (hazard ratio) ως προς το επίπεδο αναφοράς $h_0(t)$, αφού ισχύει:

$$HR = \frac{h_T(t \mid X)}{h_0(t)} = \exp\left(b_1 X_1 + b_2 X_2 + \ldots b_p X_p\right)$$

Η εκτίμηση της συνάρτησης κινδύνου $h_T(t \mid X)$, συνεπάγεται και εκτίμηση της συνάρτησης επιβίωσης $S(t)$, αφού $S(t) = \exp\left[-\int_0^t h_T(x)dx\right], t \geq 0$ και συνεπώς έχουμε:

$$S(t) = \exp\left[-\int_0^t rx^{r-1} \cdot HR dx\right] \Leftrightarrow$$

$$S(t) = \exp\left[-t^r \cdot HR\right]$$

Το επίπεδο αναφοράς για τη συνάρτηση επιβίωσης θα είναι:

$$S_0(t) = \exp\left[-\int_0^t h_0(x)\,dx\right] = \exp\left[-\int_0^t rx^{r-1}dx\right] \text{ δηλαδή}$$

$$S_0(t) = \exp\left(-t^r\right),\ t \geq 0,\ r > 0.$$

Έτσι η συνάρτηση επιβίωσης $S(t)$, δηλαδή η πιθανότητα επιβίωσης τουλάχιστον μέχρι και τη χρονική στιγμή $t$, μπορεί να γραφτεί ως $S(t) = \left[S_0(t)\right]^{HR}$

Παράδειγμα 5.8α

Οι Krall, Uthoff & Harley (1975), όπως αναφέρει ο Woodworth (σελ. 271) ανέλυσαν χρόνους επιβίωσης 65 ασθενών με πολλαπλό μελάνωμα. Οι χρόνοι αφορούσαν το θάνατο των ασθενών μέσα σε συγκεκριμένη χρονική περίοδο. Οι 48 ασθενείς πέθαναν εντός της διάρκειας διεξαγωγής της έρευνας και οι υπόλοιποι 17 επέζησαν περισσότερο από τον προκαθορισμένο χρόνο διεξαγωγής της έρευνας, άρα οι

πραγματικοί χρόνοι (για το θάνατο) ήταν μεγαλύτεροι από τη διάρκεια της έρευνας και συνεπώς δεν ήταν γνωστοί (αριστερή περικοπή των χρόνων επιβίωσης - left censoring).

Οι αναλυτές στη διάθεσή τους είχαν 2 επεξηγηματικές μεταβλητές. Το φύλο των ασθενών *($X_1$)* και τις τιμές αζώτου στον ορό του αίματος *($X_2 = BUN$).* Απέδειξαν πως ο λογάριθμος της επεξηγηματικής μεταβλητής $X_2$ βοηθάει στην καλύτερη προσαρμογή του μοντέλου, το οποίο δίνεται από τον τύπο:

$$h_T\left(t \mid X\right) = rt^{r-1}\exp\left(b_1 X_1 + b_2 \ln X_2\right).$$

Η χρήση του WINBUGS έδωσε τη δυνατότητα να υπολογιστούν οι εκ των υστέρων κατανομές για τους συντελεστές ***b****i (i = 1, 2),* όπως και για την παράμετρο *r* της κατανομής Weibull, που υποθέτουμε ότι ακολουθούν οι χρόνοι. Η ανάλυση έγινε για 2 διαφορετικά μοντέλα, ένα για τους άντρες ασθενείς και ένα για τις γυναίκες, τα οποία είναι:

$h_T\left(t \mid X\right) = rt^{r-1}\exp\left(b_{11} X_1 + b_{21} \ln X_2\right)$, όπου ο δείκτης 1 αντιπροσωπεύει τις γυναίκες ασθενείς και

$h_T\left(t \mid X^*\right) = rt^{r-1}\exp\left(b_{12} X_1 + b_{22} \ln X_2\right)$, όπου ο δείκτης 2 αντιπροσωπεύει τους άντρες ασθενείς.

Το καθένα από αυτά τα μοντέλα εκφράζει την στιγμιαία πιθανότητα θανάτου κάποιου ασθενή (γυναίκας ή άντρα) στο χρόνο *t*, δοθέντος ότι έχει επιζήσει μέχρι το χρόνο *t.*

Το σχετικό μοντέλο είναι το παρακάτω, όπου βλέπουμε αναλυτικά και τις τιμές των 65 χρόνων επιβίωσης, περικομμένες ή μη (t.obs και t.cen), όπως επίσης και τις 4 επεξηγηματικές μεταβλητές *x[ , i], i = 1, 2, 3, 4,* ξεχωριστά για γυναίκες και άνδρες:

Ισχύει ότι: $x[,1] = \begin{cases} 1, & \textit{γυναίκες} \\ 0, & \textit{άντρες} \end{cases}$ $\quad x[,2] = \begin{cases} \ln BUN, & \textit{γυναίκες} \\ 0, & \textit{άντρες} \end{cases}$

$x[,3] = \begin{cases} 1, & \textit{άντρες} \\ 0, & \textit{γυναίκες} \end{cases}$ και $x[,4] = \begin{cases} \ln BUN, & \textit{άντρες} \\ 0, & \textit{γυναίκες} \end{cases}$

Στο WINBUGS ο κώδικας είναι:

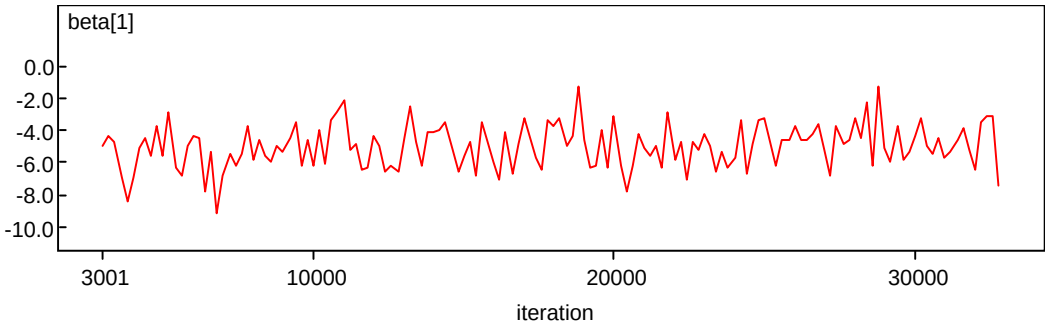
```
MODEL Weibull PHR {
 # Prior distribution of baseline hazard function
   r~dgamma(1, 0.001)

 # Prior distribution of the regression coefficients
   for (i  in 1:k) {beta[i]~dnorm(0, 0.001)}

 # Likelihood of the survival time data
   for (j in 1:n) {
   HRx[j]<-exp(inprod(x[, j], beta[ ]))
   lambda[j]<-HRx[j]
   t.obs[j]~dweib(r, lambda[j])I(t.cen[j], )
  }
}
DATA list(k=4,n=65,
t.obs=c(3, NA, 5, 6, 6, 7, 7, NA, NA, 11, NA, NA, 13, NA, 16, 18, 19, NA , 24, 26, NA,
41, NA, 52, 58, 88, 92, 1.25, 1.25, 2, 2, 2, NA, 5, 6, 6, 7, NA, 9, 11, 11, 11, 11, NA, 14,
15, 16, NA, 17, 17, 19, NA, 25, 32, 35, 37, 41, 51, NA, 54, NA, 66, 67, NA, 89),
t.cen=c(0, 4, 0, 0, 0, 0, 0, 7, 8, 0, 12, 12, 0, 13, 0, 0, 0, 19, 0, 0, 28, 0, 41, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 4, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 11, 0, 0, 0, 16, 0, 0, 0, 19, 0, 0, 0, 0, 0, 0, 53, 0, 57, 0,
0, 77, 0),
x=structure(.Data=c(
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
3.56, 4.43, 5.15, 4.87, 3.26, 2.40, 2.71, 2.56, 2.48, 2.48, 2.64, 3.22, 1.79, 3.83, 3.04,
3.33, 2.89, 3.04, 3.00, 2.83, 2.83, 2.64, 4.04, 2.3, 2.77, 2.71, 3.3, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5.11, 4.47, 3.50,
4.03, 3.00, 4.50, 3.87, 3.14, 2.56, 4.55, 3.53, 3.97, 2.56, 2.83, 3.00, 3.61, 3.71, 3.22,
3.69, 3.09, 2.64, 2.83, 3.66, 2.48, 3.04, 2.30, 3.04, 2.56, 3.69, 2.30, 3.61, 2.56, 2.89,
2.89, 3.33, 3.04, 2.48, 3.04), .Dim=c(4, 65))
)
INITIAL VALUES list(r=1, beta=c(0, 0, 0, 0))
```

**Πίνακας 5.14:** Κώδικας WINBUGS για το παραμετρικό μοντέλο βασισμένο στην κατανομή Weibull

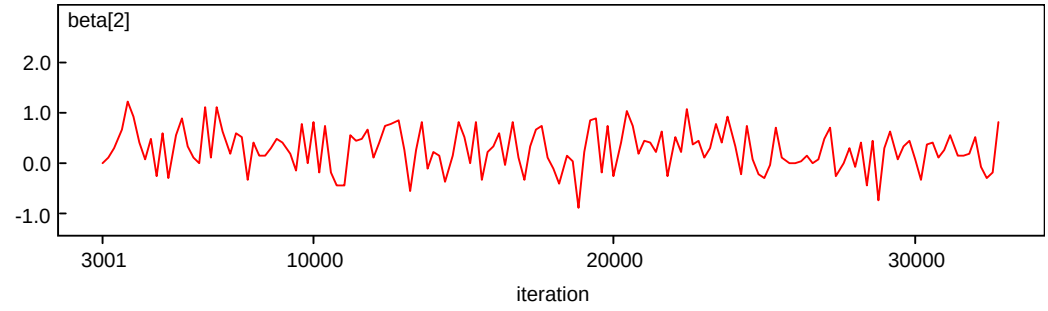
Τα παρακάτω αποτελέσματα προέκυψαν έπειτα από 30000 επαναλήψεις :

| Node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| $b_{11}$ | -4.983 | 1.45 | 0.073 | -7.906 | -4.951 | -2.24 | 3001 | 30000 |
| $b_{21}$ | 0.2556 | 0.4313 | 0.0213 | -0.607 | 0.2692 | 1.091 | 3001 | 30000 |
| $b_{12}$ | -8.006 | 1.459 | 0.07731 | -11.11 | -7.942 | -5.325 | 3001 | 30000 |
| $b_{22}$ | 1.259 | 0.3706 | 0.01896 | 0.5515 | 1.253 | 2.02 | 3001 | 30000 |
| r | 1.163 | 0.1245 | 0.0048 | 0.929 | 1.159 | 1.419 | 1001 | 30000 |

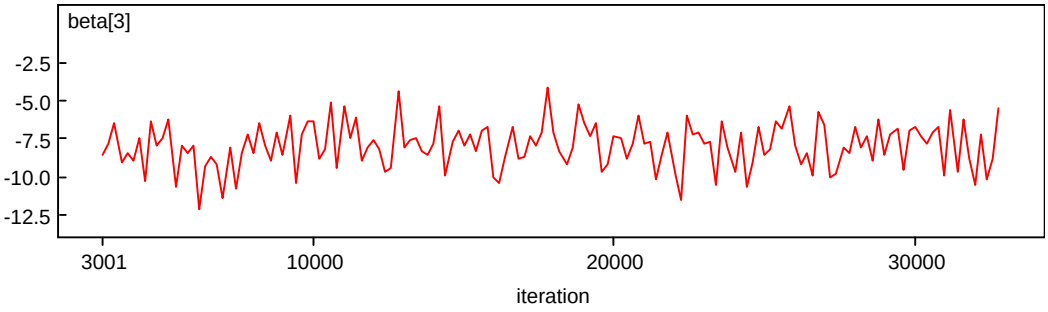
**Πίνακας 5.15:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των $b_{ij}$, $i, j = 1, 2$ και $r$



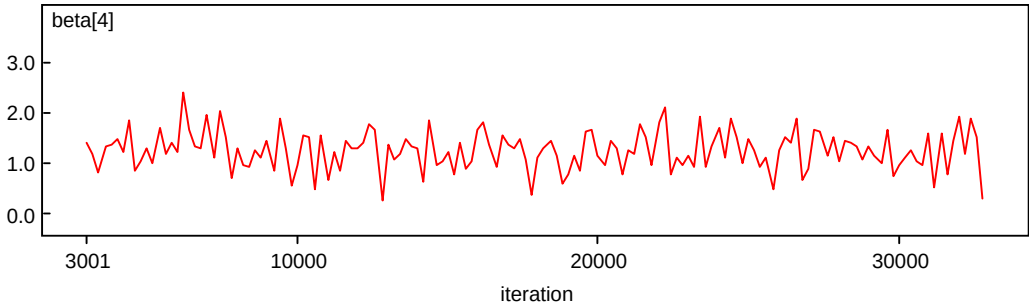

**(5.28)**



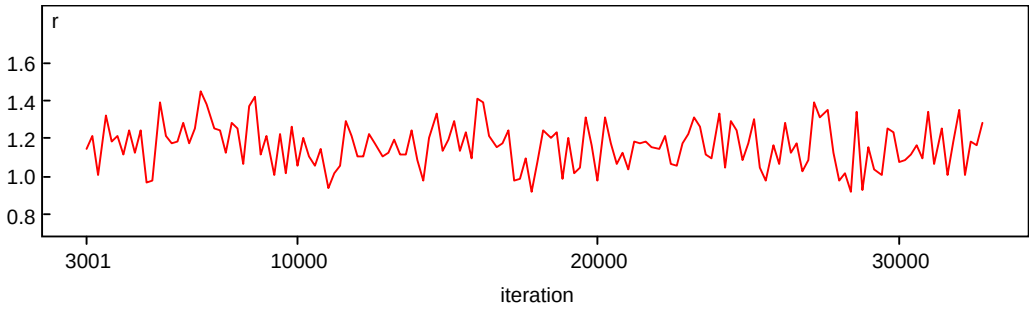

**(5.29)**




**(5.30)**



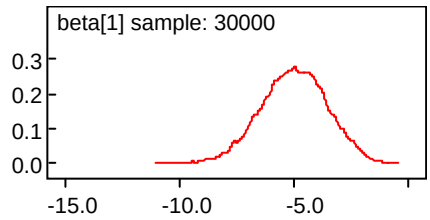

**(5.31)**



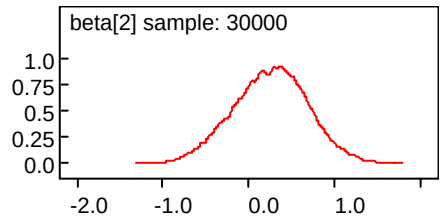

**(5.32)**

**Γραφήματα 5.28 – 5.32:** Διαγραμματική απεικόνιση του ίχνους των παραμέτρων $b_{[i]}$, $i = 1,..4$ και $r$

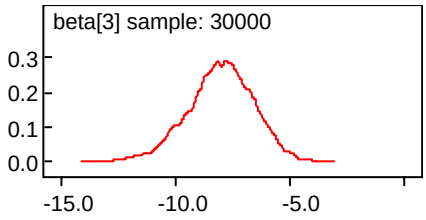
Οι προσομοιωμένες τιμές και για τις 5 παραμέτρους φαίνεται να συγκλίνουν σύμφωνα με το ίχνος τους που δίνεται στα Γραφήματα 5.28 – 5.32. Από τα γραφήματα των εκ των υστέρων κατανομών που ακολουθούν βλέπουμε ότι αυτές μπορούν να προσεγγιστούν ικανοποιητικά από την κανονική κατανομή.



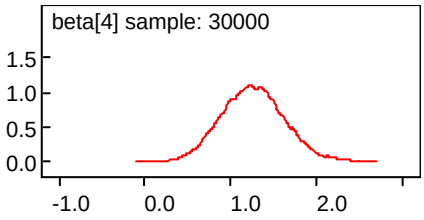

**(5.33)**



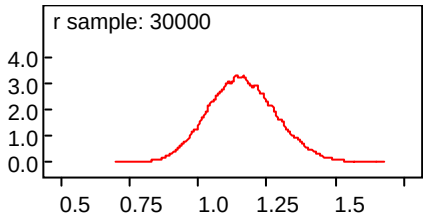

**(5.34)**




**(5.35)**




**(5.36)**




**(5.37)**

**Γραφήματα 5.33 – 5.37:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων $b_{[i]}$, $i = 1,..4$ και $r$

Από τον Πίνακα 5.15 προκύπτει ότι ο συντελεστής $b_{21} \in$ (-0.607, 1.091), που είναι το εκ των υστέρων 95% διάστημα αξιοπιστίας, το οποίο περιλαμβάνει την τιμή 0. Αυτό σημαίνει ότι η μεταβλητή *lnBUN* δεν φαίνεται να επηρεάζει το χρόνο επιβίωσης των γυναικών ασθενών. Παρόλα αυτά στην ανάλυση που ακολουθεί το μοντέλο περιλαμβάνει και τον παραπάνω συντελεστή, προκειμένου να γίνουν συγκρίσεις των χρόνων επιβίωσης μεταξύ αντρών και γυναικών και με τη χρήση της μεταβλητής *lnBUN*.

Αν χρησιμοποιήσουμε τις εκ των υστέρων μέσες τιμές, προκύπτει ότι η εκτιμώμενη συνάρτηση κινδύνου για τις γυναίκες ασθενείς είναι:

$$\hat{h}_T\left(t \mid X\right) = 1.163t^{1.163-1} \exp\left(-4.983 + 0.2556 \ln X_{20}\right)$$

και για τους άντρες ασθενείς είναι:

$$\hat{h}_T\left(t \mid X^*\right) = 1.163t^{1.163-1} \exp\left(-8.006 + 1.259 \ln X_{21}\right).$$

Συνεπώς ο αντίστοιχος λόγος στιγμιαίου κινδύνου (hazard ratio) μεταξύ γυναικών και αντρών με ίδια τιμή *BUN* ίση με $X_2$ δίδεται από τον τύπο:

$$HR = \frac{\hat{h}_T\left(t \mid X\right)}{\hat{h}_T\left(t \mid X^*\right)} = \frac{\exp\left(-4.983 + 0.2556 \ln X_2\right)}{\exp\left(-8.006 + 1.259 \ln X_2\right)} = \exp\left(3.023 - 1.0034 \ln X_2\right)$$

Το συμπέρασμα που προέκυψε από τα παραπάνω είναι ότι ο ρυθμός θανάτου για τις γυναίκες ασθενείς είναι μεγαλύτερος από το ρυθμό θανάτου των αντρών ασθενών. Αυτό ισχύει όμως για μικρές τιμές αζώτου στο αίμα (πχ $X_2 = 6$ = η μικρότερη τιμή) αφού τότε θα είναι: $HR = \exp\left(3.023 - 1.0034 \cdot \ln 6\right) =$ 3.4 το οποίο σημαίνει ότι ο ρυθμός θανάτου των γυναικών είναι 3.4 φορές μεγαλύτερος από τον ρυθμό θανάτου των αντρών. Όμως για μεγάλες τιμές του αζώτου στο αίμα (πχ $X_2 = 171$ = η μεγαλύτερη τιμή) θα είναι $HR = \exp\left(3.023 - 1.0034 \cdot \ln 172\right) =$ 0.117, ο ρυθμός θανάτου των γυναικών είναι σχεδόν τα 12% του αντίστοιχου ρυθμού θανάτου των αντρών.

Στο πρόγραμμα WINBUGS το hazard ratio (*HR*) μεταξύ γυναικών – ανδρών, ο ρυθμός επιβίωσης (Survival rate *Sr*), καθώς και το Relative Risk (*RR*), το οποίο είναι ο λόγος δύο πιθανοτήτων για δύο διαφορετικές ομάδες δεδομένων, μπορούν να θεωρηθούν ως παράμετροι. Έτσι αν "ξανατρέξουμε" το μοντέλο θα μπορούμε να υπολογίσουμε τις εκ των υστέρων κατανομές και για αυτά.

Μπορούμε να επιλέξουμε μερικές ενδεικτικές τιμές του $X_2$ = *BUN* (ξεχωριστά για γυναίκες και άντρες) για τις οποίες θα βρεθούν οι εκ των υστέρων κατανομές του hazard ratio (*HR*), του ρυθμού επιβίωσης και του σχετικού κινδύνου (*RR*). Αυτές είναι η μικρότερη ( = 6) και η μεγαλύτερη ( = 172), η διάμεσος τιμή ( = 21), όπως επίσης και τα 2 τεταρτημόρια ($Q_1$ = 14 και $Q_2$ = 37). Άρα έχουμε συνολικά 10 περιπτώσεις.

Για να γίνουν τα παραπάνω χρειάζεται η κατασκευή 4 νέων μεταβλητών *x.r[ , j] (j = 1, 2, 3, 4)*, κατά παρόμοιο τρόπο με τις *x[ , j] (j = 1, 2, 3, 4)*

Οι μεταβλητές *x.r[, 2]* και *x.r[ ,4]* θα έχουν τις παραπάνω τιμές του *lnBUN* (minimum, $Q_1$, διάμεσο, $Q_3$, maximum), ενώ οι μεταβλητές *x.r[, 1]* και *x.r[ ,3]* θα είναι δίτιμες δείκτριες μεταβλητές που θα υποδεικνύουν το φύλο των νέων παρατηρήσεων που θέλουμε να υπολογίσουμε a – posteriori. Επίσης θα χρησιμοποιηθεί η μεταβλητή *t.r[ ]* που παίρνει την τιμή 24 και για τις 10 περιπτώσεις. Δηλαδή θα υπολογιστούν οι εκ

των υστέρων κατανομές των ρυθμών θανάτων *HR,* των πιθανοτήτων επιβίωσης *Sr* και των σχετικών κινδύνων *RR* για χρονική περίοδο 24 μηνών (Woodworth – σελ. 273).

Χρησιμοποιούμε το ίδιο προηγούμενο μοντέλο (Πίνακας 5.14), με επιπλέον δεδομένα:

```
MODEL Weibull PHR {
# Prior distribution of baseline hazard function
r~dgamma(1, 0.001)
# Prior distribution of the regression coefficients
for (i in 1:k) {beta[i]~dnorm(0, 0.001)}
# Likelihood of the survival time data
for (j in 1:n) {
HRx[j]<-exp(inprod(x[,j], beta[ ]))
lambda[j]<-HRx[j]
t.obs[j]~dweib(r, lambda[j])I(t.cen[j], ) }

# Requested survival rates of interest in this analysis
for (j in 1:m) {
HRr[j]<-exp(inprod(x.r[ , j], beta[ ]))
 Sr[j]<-exp(-HRr[j]*pow(t.r[j], r))
}

# Contrasts of interest in this analysis
# Female/Male RRs & HRs
 for (j in 1:5) {
 RRfm[j]<-(1-Sr[2*j])/(1-Sr[2*j-1])
 HRfm[j]<-HRr[2*j]/HRr[2*j-1]}
}
DATA list(k=4,m=10,n=65,
t.obs=c(3, NA, 5, 6, 6, 7, 7, NA, NA, 11, NA, NA, 13, NA, 16, 18, 19, NA , 24, 26, NA,
41, NA, 52, 58, 88, 92, 1.25, 1.25, 2, 2, 2, NA, 5, 6, 6, 7, NA, 9, 11, 11, 11, 11, NA, 14,
15, 16, NA, 17, 17, 19, NA, 25, 32, 35, 37, 41, 51, NA, 54, NA, 66, 67, NA, 89),
t.cen=c(0, 4, 0, 0, 0, 0, 0, 7, 8, 0, 12, 12, 0, 13, 0, 0, 0, 19, 0, 0, 28, 0, 41, 0, 0, 0, 0, 0, 0,
0, 0, 0, 4, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 11, 0, 0, 0, 16, 0, 0, 0, 19, 0, 0, 0, 0, 0, 0, 53, 0, 57, 0,
0, 77, 0),
x=structure(.Data=c(
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
3.56, 4.43, 5.15, 4.87, 3.26, 2.40, 2.71, 2.56, 2.48, 2.48, 2.64, 3.22, 1.79, 3.83, 3.04,
3.33, 2.89, 3.04, 3.00, 2.83, 2.83, 2.64, 4.04, 2.3, 2.77, 2.71, 3.3, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

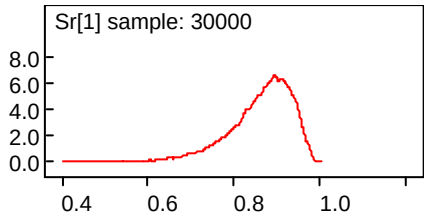
```
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5.11, 4.47, 3.50,
4.03, 3.00, 4.50, 3.87, 3.14, 2.56, 4.55, 3.53, 3.97, 2.56, 2.83, 3.00, 3.61, 3.71, 3.22,
3.69, 3.09, 2.64, 2.83, 3.66, 2.48, 3.04, 2.30, 3.04, 2.56, 3.69, 2.30, 3.61, 2.56, 2.89,
2.89, 3.33, 3.04, 2.48, 3.04), .Dim=c(4, 65))


t.r=c(24, 24, 24, 24, 24, 24, 24, 24, 24, 24),

x.r=structure(.Data=c(0, 1, 0, 1, 0, 1, 0, 1, 0, 1,
0, 1.79, 0, 2.64, 0, 3.04, 0, 3.61, 0, 5.15,
1, 0, 1, 0, 1, 0, 1, 0, 1, 0,
1.79, 0, 2.64, 0, 3.04, 0, 3.61, 0, 5.15, 0),.Dim=c(4, 10))


INITIAL VALUES list(r=1, beta=c(0, 0, 0, 0))
```

**Πίνακας 5.16:** Κώδικας WINBUGS για το παραμετρικό μοντέλο βασισμένο στην κατανομή Weibull με τις επιπλέον μεταβλητές *x.r[ ]* και *t.r[ ]*

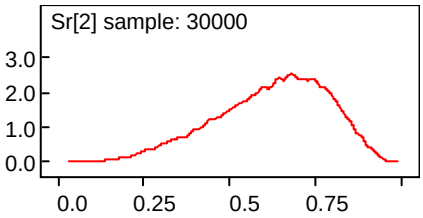
Τα αποτελέσματα για το ρυθμό επιβίωσης σε 24 μήνες, δηλαδή η $S_k(24)$, με $k = 1, 2,...10$, ξεχωριστά για γυναίκες και άντρες σε κάθε διαφορετική τιμή του *BUN* είναι:

| **Φύλο** | **BUN** | **node** | **mean** | **sd** | **MC error** | **2.5%** | **median** | **97.5%** |
|---|---|---|---|---|---|---|---|---|
| Α | 6 | $Sr_{[1]}$ | 0.86606 | 0.07428 | 0.004295 | 0.6782 | 0.8742 | 0.9642 |
| Γ | 6 | $Sr_{[2]}$ | 0.634 | 0.1611 | 0.00883 | 0.2801 | 0.6527 | 0.8916 |
| Α | 14 | $Sr_{[3]}$ | 0.6788 | 0.08034 | 0.004241 | 0.513 | 0.6824 | 0.8266 |
| Γ | 14 | $Sr_{[4]}$ | 0.58 | 0.09537 | 0.004069 | 0.3878 | 0.5812 | 0.7618 |
| Α | 21 | $Sr_{[5]}$ | 0.5356 | 0.07169 | 0.002682 | 0.3975 | 0.5355 | 0.6747 |
| Γ | 21 | $Sr_{[6]}$ | 0.5452 | 0.08076 | 0.001685 | 0.3853 | 0.5469 | 0.6972 |
| Α | 37 | $Sr_{[7]}$ | 0.2857 | 0.07588 | 0.001862 | 0.1519 | 0.28 | 0.4475 |
| Γ | 37 | $Sr_{[8]}$ | 0.4868 | 0.1127 | 0.003923 | 0.2736 | 0.4866 | 0.7079 |
| Α | 172 | $Sr_{[9]}$ | 0.01145 | 0.0357 | 0.001451 | 0 | 0 | 0.1119 |
| Γ | 172 | $Sr_{[10]}$ | 0.3471 | 0.2546 | 0.01338 | 0.002164 | 0.3149 | 0.8499 |

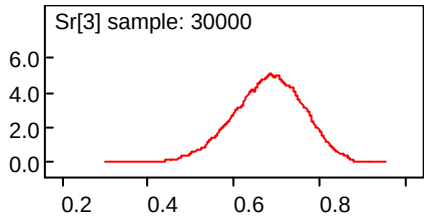
**Πίνακας 5.17:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών της πιθανότητας επιβίωσης *Sr[i], i = 1, 2,...10*



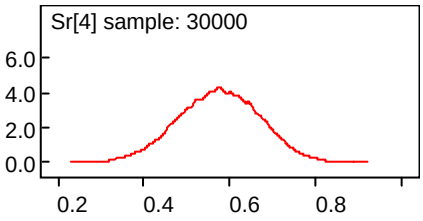

**(5.38)**



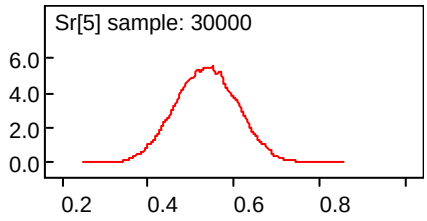

**(5.39)**



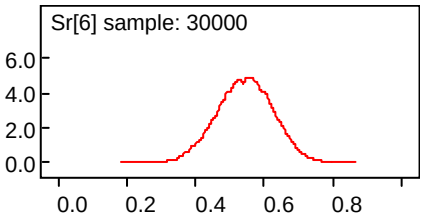

**(5.40)**



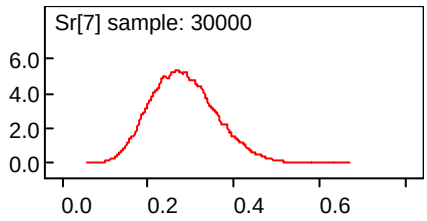

**(5.41)**



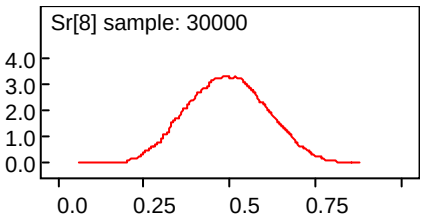

**(5.42)**



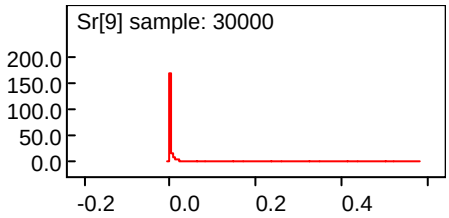

**(5.43)**

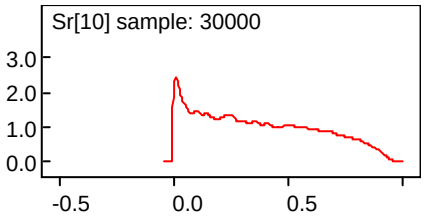



**(5.44)**



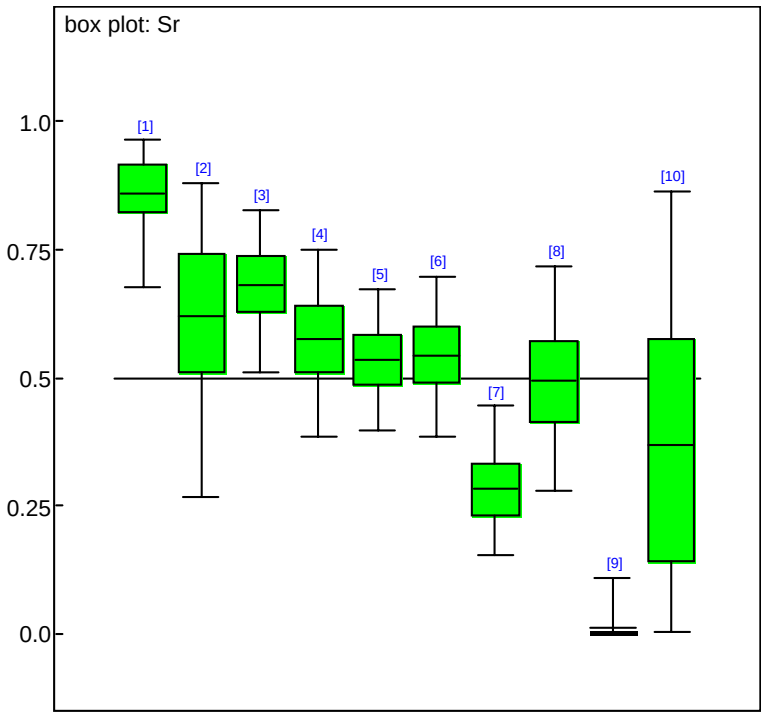

**(5.45)**




**(5.46)** **(5.47)**

**Γραφήματα 5.38 – 5.47:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων *Sr[i] i = 1,2,...10,*
δηλαδή των πιθανοτήτων ένας ασθενής να επιζήσει περισσότερο από 24 μήνες




**Γράφημα 5.48:** Γραφική παράσταση των παραμέτρων *Sr[i], i = 1,2,...10,*
δηλαδή των πιθανοτήτων ένας ασθενής να επιζήσει περισσότερο από 24 μήνες

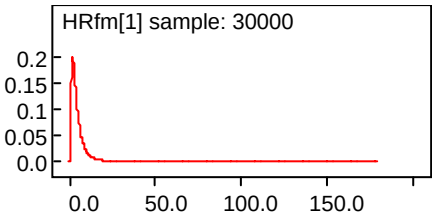
Από τα παραπάνω προκύπτει το συμπέρασμα ότι ο παράγοντας *BUN* είναι σημαντικός για τους άντρες ασθενείς, καθώς αύξηση του *BUN* από την ελάχιστη στη μέγιστη τιμή, συνεπάγεται μείωση του a – posteriori αναμενόμενου ρυθμού επιβίωσης από 86.6% σε 1.1%. Για τις γυναίκες υπάρχει μείωση, αλλά μικρότερη (από 63.4% σε 34.7%).

Για την διάμεσο τιμή του *BUN* = 21, οι ρυθμοί επιβίωσης για άντρες και γυναίκες είναι σχεδόν ίσοι (53.6% και 54.5% αντίστοιχα).

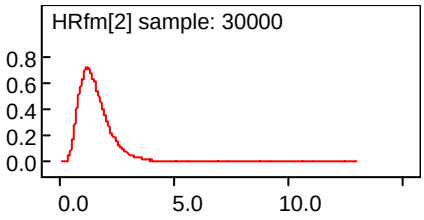
Τα αποτελέσματα των λόγων στιγμιαίων κινδύνων (*HR*) και των σχετικών κινδύνων (*RR*) για σύγκριση γυναικών – αντρών μετά από 20000 επαναλήψεις είναι:

| **BUN** | **node** | **mean** | **sd** | **MC error** | **2.5%** | **median** | **97.5%** |
|---|---|---|---|---|---|---|---|
| 6 | $HR_{[1]}$ | 4.49 | 4.715 | 0.2238 | 0.6468 | 3.201 | 17.15 |
| 14 | $HR_{[2]}$ | 1.55 | 0.6917 | 0.0291 | 0.6141 | 1.419 | 3.289 |
| 21 | $HR_{[3]}$ | 1.018 | 0.3309 | 0.0072 | 0.5158 | 0.969 | 1.802 |
| 37 | $HR_{[4]}$ | 0.608 | 0.2453 | 0.0077 | 0.2481 | 0.569 | 1.200 |
| 172 | $HR_{[5]}$ | 0.24 | 0.341 | 0.0161 | 0.018 | 0.131 | 1.095 |
| 6 | $RR_{[1]}$ | 3.52 | 3.093 | 0.1555 | 0.6728 | 2.742 | 11.44 |
| 14 | $RR_{[2]}$ | 1.392 | 0.490 | 0.0215 | 0.6762 | 1.315 | 2.586 |
| 21 | $RR_{[3]}$ | 1.002 | 0.234 | 0.0052 | 0.6136 | 0.978 | 1.527 |
| 37 | $RR_{[4]}$ | 0.727 | 0.181 | 0.0059 | 0.3991 | 0.719 | 1.106 |
| 172 | $RR_{[5]}$ | 0.661 | 0.259 | 0.0135 | 0.1516 | 0.693 | 1.002 |

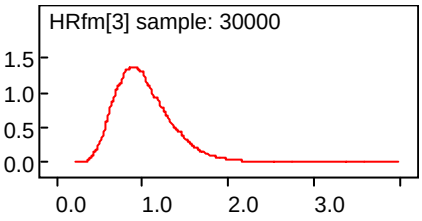
**Πίνακας 5.18:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των ρυθμών θανάτων *HR[i],* και των σχετικών κινδύνων *RR[i], i = 1, 2,…5*



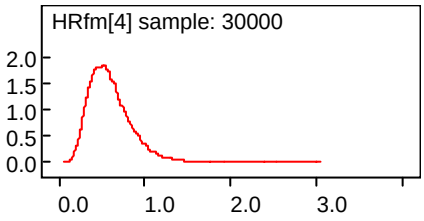

**(5.49)**



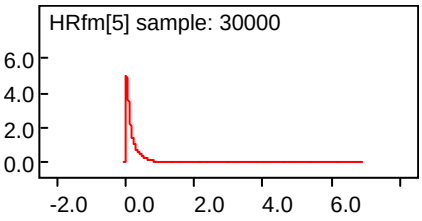

**(5.50)**



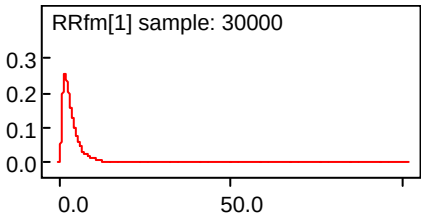

**(5.51)**



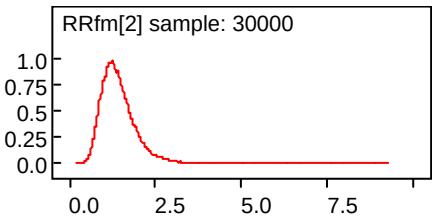

**(5.52)**



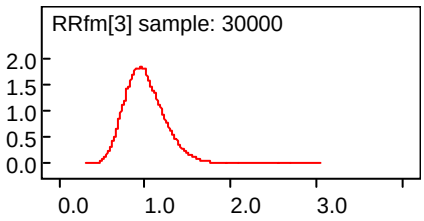

**(5.53)**



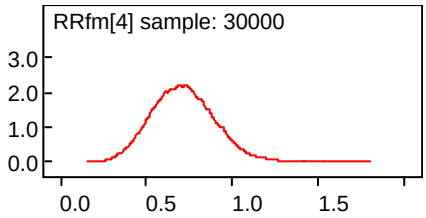

**(5.54)**

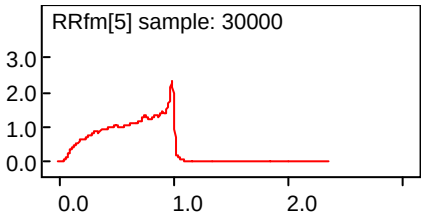



**(5.55)**



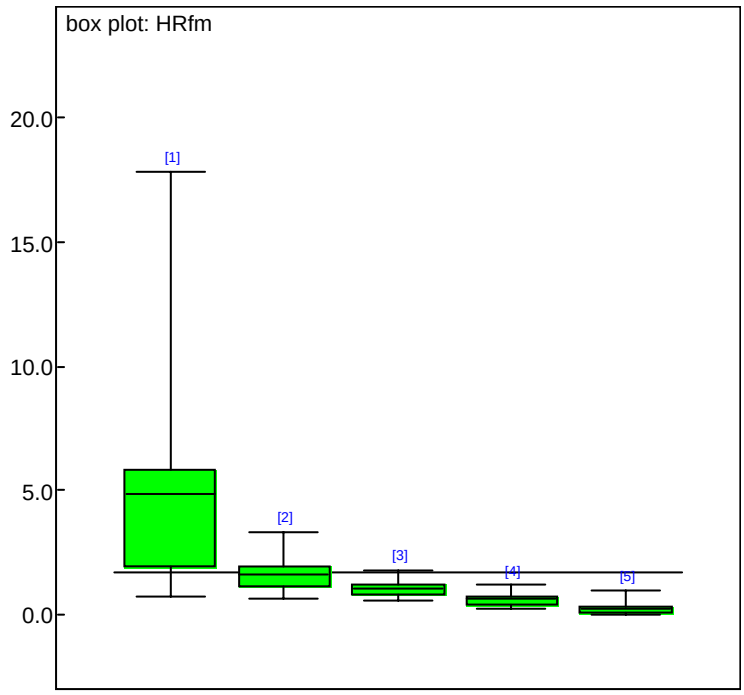

**(5.56)**



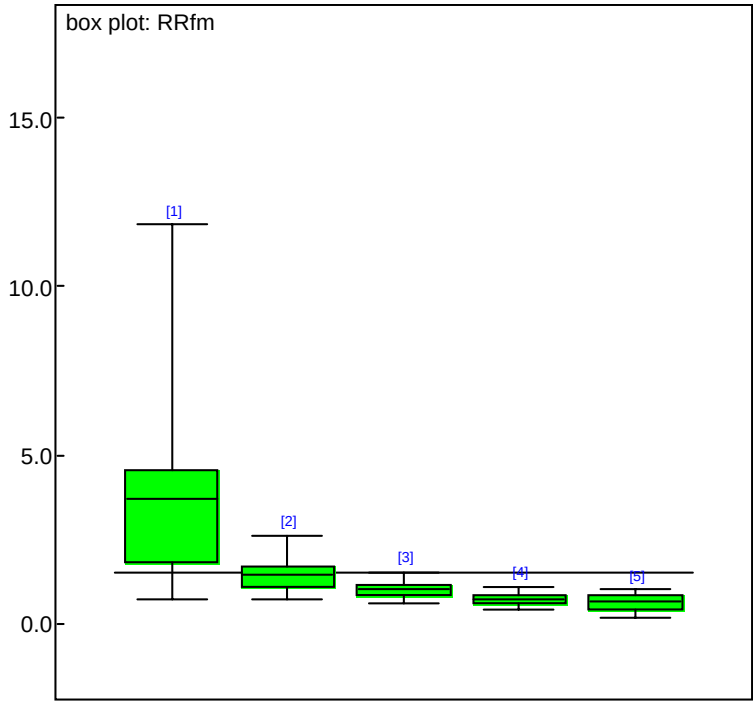




**(5.57)** **(5.58)**

**Γραφήματα 5.49 – 5.58:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων *HR[i]* και *RR[i], i = 1,2,...5,* δηλαδή των ρυθμών θανάτου και των σχετικών κινδύνων




**(5.59):** Hazard ratios μεταξύ γυναικών και αντρών




**(5.60):** Relative risk μεταξύ γυναικών και αντρών

**Γραφήματα 5.59 & 5.60:** Γραφικές παραστάσεις των παραμέτρων *HR[i]* και *RR[i], i = 1,2,...5,* δηλαδή των ρυθμών θανάτου και των σχετικών κινδύνων

Από τον Πίνακα 5.17 έχουμε ότι για *BUN* = 6, ο ρυθμός θανάτου για τις γυναίκες σε διάστημα 24 μηνών είναι 4.49 φορές μεγαλύτερος από τον αντίστοιχο κίνδυνο που διατρέχουν οι άντρες ($HR_{[1]}$), ενώ για τη διάμεσο τιμή του *BUN* = 21, οι γυναίκες και οι άντρες διατρέχουν σχεδόν τον ίδιο κίνδυνο ($HR_{[3]} = 1.018$). Αντιθέτως, όταν η τιμή του *BUN* γίνεται μέγιστη (172) ο ρυθμός θανάτου των γυναικών ίσος με το ένα τέταρτο του αντίστοιχου ρυθμού θανάτου των αντρών.

Οι σχετικοί κίνδυνοι (*RR*) έχουν παρόμοια ερμηνεία με τους λόγους στιγμιαίου κινδύνου (*HR*). Το *RR* ορίζεται ως ο λόγος $\frac{1-S_f(24)}{1-S_m(24)}$. Είναι η πιθανότητα οι γυναίκες ασθενείς να ζήσουν λιγότερο από 24 μήνες προς την αντίστοιχη πιθανότητα για τους άντρες ασθενείς.

Από τα αποτελέσματα προκύπτει ότι για την μικρότερη τιμή του *BUN* = 6, το *RR* θανάτου μέσα σε διάστημα 24 μηνών μεταξύ γυναικών και αντρών είναι 3.52, άρα

για τις γυναίκες υπάρχει 2.5 φορές μεγαλύτερος κίνδυνος απ' ότι για τους άντρες. Όμως για τιμές του *BUN* μεγαλύτερες από τη διάμεσο ( > 21) για τις γυναίκες υπάρχει μικρότερος κίνδυνος απ' ότι για τους άντρες (28% και 34% μικρότερος κίνδυνος κατά μέσο όρο για *BUN* 37 και 172 αντίστοιχα).

Γενικά λοιπόν, οι γυναίκες κινδυνεύουν περισσότερο από τους άντρες για μικρές τιμές του αζώτου στο αίμα (κάτω της διαμέσου $BUN = 21$)

Παράδειγμα 5.8β

Το τμήμα πυρηνικής ιατρικής ενός νοσηλευτικού ιδρύματος έκανε μία έρευνα σε 24 ασθενείς, οι οποίοι είναι υποθυρεοειδικοί. Μέτρησε την ποσότητα της ορμόνης *TSH* σε κάθε έναν από αυτούς και τους χορήγησε από μία θεραπεία, η οποία διήρκησε 10 ημέρες. Υπήρχαν 3 είδη θεραπείας, όπου η κάθε μία εφαρμόστηκε σε 8 ασθενείς.

Μετά το πέρας των 10 ημερών σε κάθε ασθενή άρχισε να μετριέται καθημερινά για μία εβδομάδα η ποσότητα της ορμόνης *TSH*. Ο χρόνος (σε ημέρες) που καταγράφηκε για κάθε ασθενή είναι ο χρόνος από την έναρξη της θεραπείας έως την ημέρα που τα επίπεδα της συγκεκριμένης ορμόνης "έπεσαν" κάτω από τις 5.4 μονάδες μU/ml, άρα η θεραπεία κρίθηκε αποτελεσματική για αυτούς τους ασθενείς. Υπήρχαν όμως και 6 ασθενείς, στους οποίους η θεραπεία που τους χορηγήθηκε δεν έδωσε ικανοποιητικά αποτελέσματα κι έτσι δεν μετρήθηκαν φυσιολογικές τιμές της ορμόνης *TSH*, εντός του διαστήματος της μίας εβδομάδας από τη λήξη της θεραπείας (περικομμένες παρατηρήσεις – αυτοί οι πραγματικοί χρόνοι είναι > 17 ημέρες, αλλά άγνωστοι).

Σκοπός της έρευνας ήταν να γίνουν συγκρίσεις για την αποτελεσματικότητα του κάθε είδους θεραπείας. Χρησιμοποιήθηκε το γενικό μοντέλο $h_T(t \mid X) = h_0(t)\exp\left(b_1X_1 + b_2X_2\right)$, όπου η μεταβλητή $X_1$ είναι το είδος της θεραπείας και έχει 3 επίπεδα (με τιμές 0, 1 και 2) και η $X_2$ είναι η ποσότητα της ορμόνης *TSH* στον κάθε ασθενή πριν την έναρξη της θεραπείας.

Το επίπεδο αναφοράς της συνάρτησης κινδύνου, δηλαδή η $h_0(t)$, ακολουθεί την κατανομή Weibull με παραμέτρους $r$ και $\lambda = 1$.

Έτσι έχουμε ότι: $h_0(t) = rt^{r-1}$ με *r > 0* και η εκ των προτέρων κατανομή για την παράμετρο *r* είναι η Γάμμα *G(0.1, 0.001).*

Υπολογίστηκαν εκ των υστέρων τιμές και για τους συντελεστές *bi (i = 1, 2).* Η ανάλυση έγινε για 3 διαφορετικά μοντέλα, ένα για κάθε είδος θεραπείας, τα οποία είναι:
$h_T(t \mid X)_1 = rt^{r-1}\exp(b_{11}X_1 + b_{21}X_2)$, όπου ο $2^{ος}$ δείκτης (1) αντιπροσωπεύει το $1^{ο}$ είδος θεραπείας,
$h_T(t \mid X)_2 = rt^{r-1}\exp(b_{12}X_1 + b_{22}X_2)$, όπου ο $2^{ος}$ δείκτης (2) αντιπροσωπεύει το $2^{ο}$ είδος θεραπείας και
$h_T(t \mid X)_3 = rt^{r-1}\exp(b_{13}X_1 + b_{23}X_2)$, όπου ο $2^{ος}$ δείκτης (3) αντιπροσωπεύει το $3^{ο}$ είδος θεραπείας.

Για την ανάλυση κατασκευάστηκαν 6 επεξηγηματικές μεταβλητές *x[ , i], i = 1, 2, 3, 4,5, 6* ξεχωριστά για κάθε θεραπεία.

Ισχύει ότι: $x[,1] = \begin{cases} 1, & \theta\varepsilon\rho\alpha\pi\varepsilon\acute{\iota}\alpha\ 1 \\ 0, & \acute{\alpha}\lambda\lambda\eta \end{cases}$ $x[,2] = \begin{cases} TSH, & \mu\varepsilon\ \theta\varepsilon\rho\alpha\pi\varepsilon\acute{\iota}\alpha\ 1 \\ 0, & \mu\varepsilon\ \acute{\alpha}\lambda\lambda\eta \end{cases}$

$x[,3] = \begin{cases} 1, & \theta\varepsilon\rho\alpha\pi\varepsilon\acute{\iota}\alpha\ 2 \\ 0, & \acute{\alpha}\lambda\lambda\eta \end{cases}$ $x[,4] = \begin{cases} TSH, & \mu\varepsilon\ \theta\varepsilon\rho\alpha\pi\varepsilon\acute{\iota}\alpha\ 2 \\ 0, & \mu\varepsilon\ \acute{\alpha}\lambda\lambda\eta \end{cases}$

$x[,5] = \begin{cases} 1, & \theta\varepsilon\rho\alpha\pi\varepsilon\acute{\iota}\alpha\ 3 \\ 0, & \acute{\alpha}\lambda\lambda\eta \end{cases}$ και $x[,6] = \begin{cases} TSH, & \mu\varepsilon\ \theta\varepsilon\rho\alpha\pi\varepsilon\acute{\iota}\alpha\ 3 \\ 0, & \mu\varepsilon\ \acute{\alpha}\lambda\lambda\eta \end{cases}$

Το μοντέλο στο WINBUGS είναι:

```
MODEL Weibull PHR {
# Prior distribution of baseline hazard function
r~dgamma(0.1, 0.001)
# Prior distribution of the regression coefficients
for (i in 1:k) {beta[i]~dnorm(1, 0.001)}
# Likelihood of the survival time data
for (j in 1:n) {
HRx[j]<-exp(inprod(x[,j], beta[ ]))
lambda[j]<-HRx[j]
```

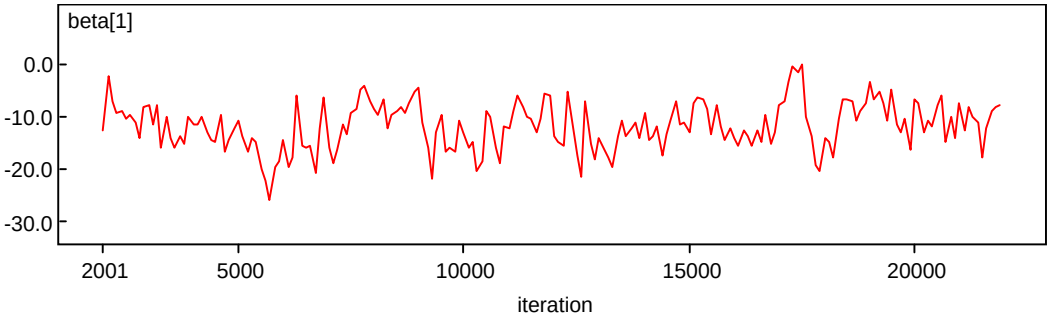
```
t.obs[j]~dweib(r, lambda[j])I(t.cen[j], ) }

# Requested survival rates of interest in this analysis
for (j in 1:m) {
HRr[j]<-exp(inprod(x.r[ , j], beta[ ]))
 Sr[j]<-exp(-HRr[j]*pow(t.r[j], r))
}

# Contrasts of interest in this analysis
# RRs & HRs
 for (j in 1:3) {
 RR12[j]<-(1-Sr[3*j-2])/(1-Sr[3*j-1])
 RR13[j]<-(1-Sr[3*j-2])/(1-Sr[3*j])
 RR23[j]<-(1-Sr[3*j-1])/(1-Sr[3*j])
 HR12[j]<-HRr[3*j-2]/HRr[3*j-1]
 HR13[j]<-HRr[3*j-2]/HRr[3*j]
 HR23[j]<-HRr[3*j-1]/HRr[3*j]
}
}
DATA list(k=6,m=9,n=24,
t.obs=c(NA, 13, NA, 13, 17, 15, NA, 14, 12, NA, 14, 13, 13, 15, 3, NA, 16, 15, 15, NA,
15, 12, 14, 12),
t.cen=c(18, 0, 19, 0, 0, 0, 18, 0, 0, 19, 0, 0, 0, 0, 0, 20, 0, 0, 0, 18, 0, 0, 0, 0),

t.r=c(17, 17, 17, 17, 17, 17, 17, 17, 17),

x.r=structure(.Data=c(1, 0, 0, 1, 0, 0, 1, 0, 0,
5.5, 0, 0, 6.05, 0, 0, 8.1, 0, 0,
0, 1, 0, 0, 1, 0, 0, 1, 0,
0, 5.5, 0, 0, 6.05, 0, 0, 8.1, 0,
0, 0, 1, 0, 0, 1, 0, 0, 1,
0, 0, 5.5, 0, 0, 6.05, 0, 0, 8.1),.Dim=c(6, 9)),
x=structure(.Data=c(1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0,
5.8, 5.6, 6, 5.5, 8, 7, 7.1, 7.3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 6.2, 5.9, 6.4, 6, 5.7, 8.1, 5.9, 6.1, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5.6, 6, 6, 7, 7.1, 5.8, 8, 6.7),.Dim=c(6, 24))
)

INITIAL VALUES list(r=0.1, beta=c(0, 0, 0, 0, 0, 0))
```

**Πίνακας 5.19:** Κώδικας WINBUGS για το παραμετρικό μοντέλο βασισμένο στην κατανομή Weibull (β)

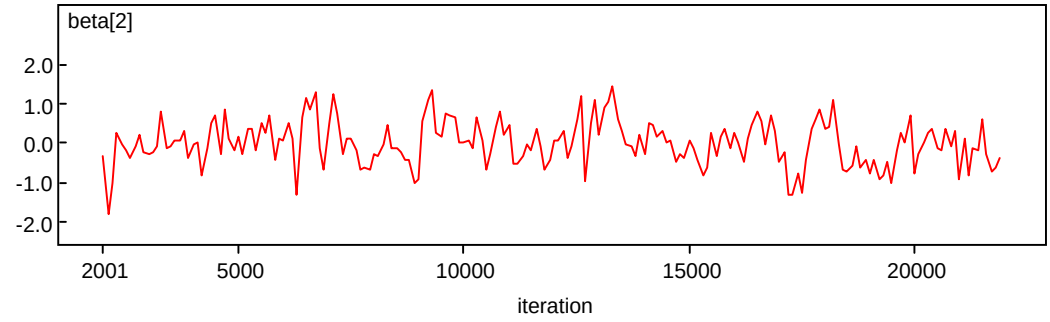
Τα αποτελέσματα για τους συντελεστές $b_i$ και $r$ δίνονται παρακάτω και προέκυψαν έπειτα από 20000 επαναλήψεις

| Node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| $b_{11}$ | -11.85 | 4.507 | 0.335 | -20.88 | -11.94 | -2.805 | 2001 | 20000 |
| $b_{21}$ | 0.006 | 0.567 | 0.039 | -1.223 | 0.038 | 1.058 | 2001 | 20000 |
| $b_{12}$ | -12.07 | 4.588 | 0.332 | -20.91 | -12.02 | -3.5 | 2001 | 20000 |
| $b_{22}$ | 0.109 | 0.559 | 0.028 | -1.065 | 0.153 | 1.059 | 2001 | 20000 |
| $b_{13}$ | -9.541 | 4.243 | 0.306 | -18.32 | -9.48 | -1.728 | 2001 | 20000 |
| $b_{23}$ | -0.237 | 0.547 | 0.036 | -1.344 | -0.221 | 0.738 | 2001 | 20000 |
| $r$ | 3.989 | 0.887 | 0.065 | 2.502 | 3.913 | 5.865 | 2001 | 20000 |

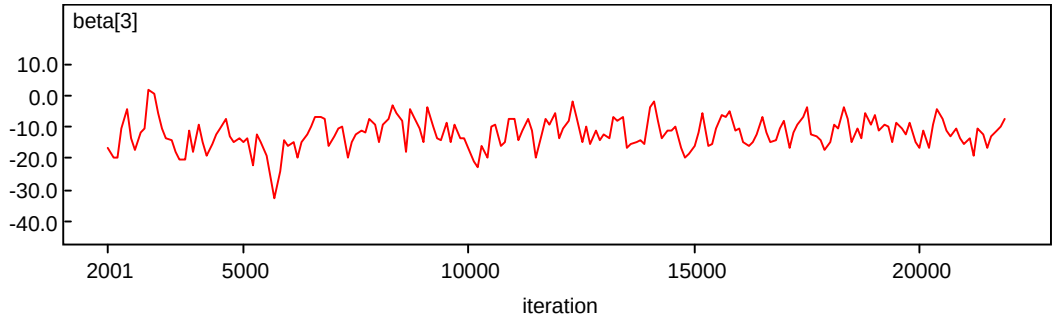
**Πίνακας 5.20:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των $b_{ij}$, $i, = 1,2,\ j = 1, 2, 3$ και $r$



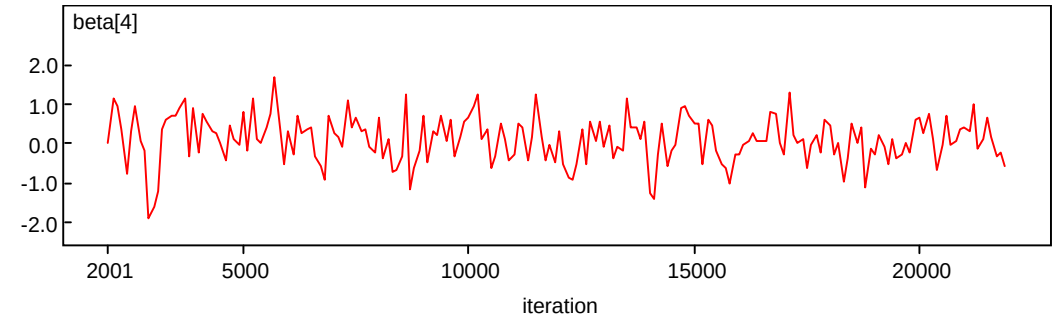

**(5.61)**



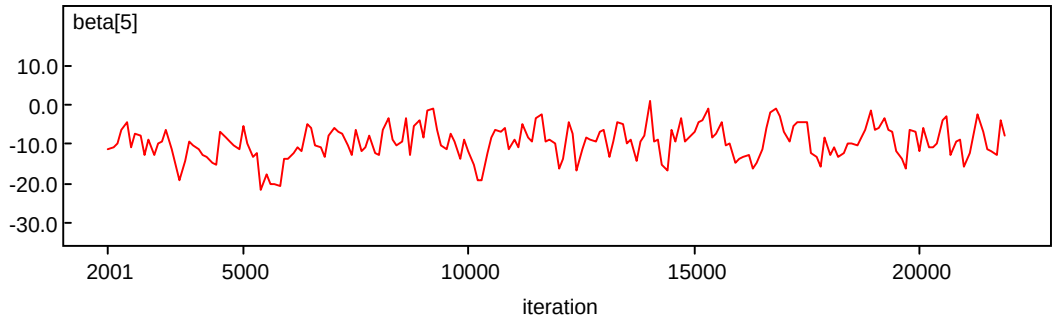

**(5.62)**



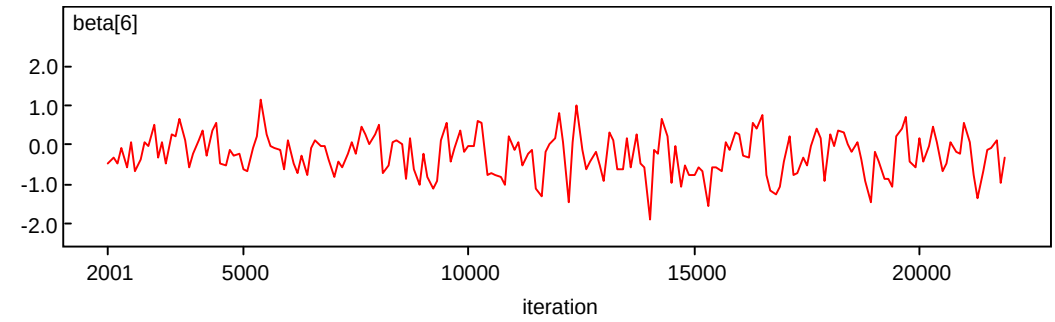

**(5.63)**



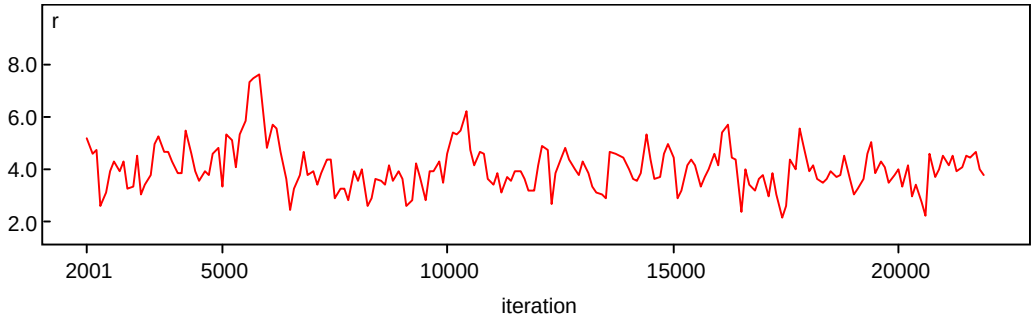

**(5.64)**




**(5.65)**

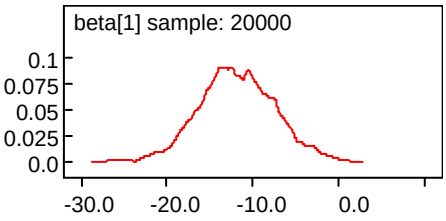



**(5.66)**

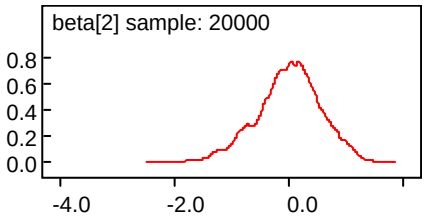



**(5.67)**

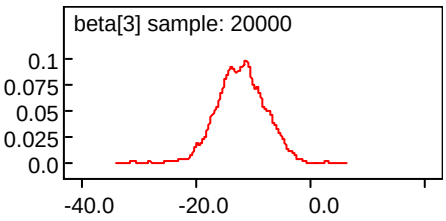
**Γραφήματα 5.61 – 5.67:** Διαγραμματική απεικόνιση του ίχνους των παραμέτρων $b_{[ij]}$, $i = 1, 2,..6$ και $r$ (β)



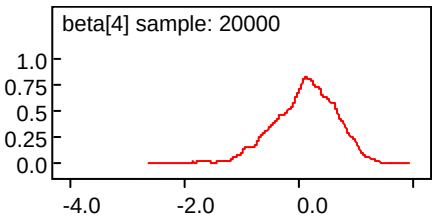

**(5.68)**



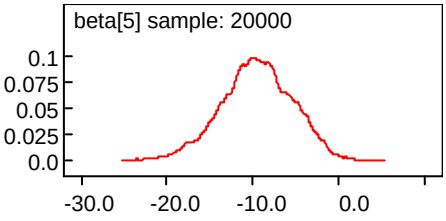

**(5.69)**



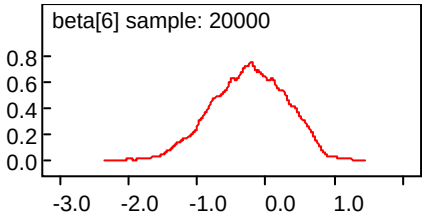

**(5.70)**



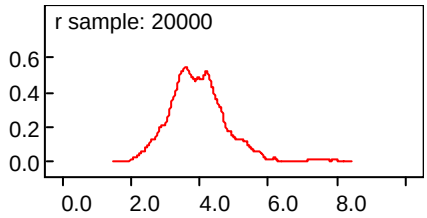

**(5.71)**




**(5.72)**




**(5.73)**




**(5.74)**

**Γραφήματα 5.68 – 5.74:** Εκτίμηση των εκ των υστέρων συναρτήσεων πυκνότητας πιθανότητας των παραμέτρων $b_{[i]}$, $i = 1, 2,,..6$ και $r$ (β)

Το *DIC* για το μοντέλο είναι 128.714

Από τον Πίνακα 5.20 προκύπτει ότι οι συντελεστές $b_{21}$, $b_{22}$ και $b_{23}$ μπορούν να πάρουν την τιμή 0. Δηλαδή η ποσότητα της ορμόνης *TSH* που μετρήθηκε σε κάθε ασθενή δεν φαίνεται να επηρεάζει το χρόνο ανάρρωσης των ασθενών και αυτό ισχύει και για τα τρία είδη θεραπείας. Δηλαδή ο ρυθμός ανάρρωσης των ασθενών ουσιαστικά εξαρτάται μόνο από το είδος της θεραπείας που εφαρμόστηκε.

Το μοντέλο που ακολουθεί δεν περιλαμβάνει τις μεταβλητές $X_{21}$, $X_{22}$, $X_{23}$:

```
MODEL Weibull PHR {
# Prior distribution of baseline hazard function
r~dgamma(0.1, 0.001)
# Prior distribution of the regression coefficients
for (i in 1:k) {beta[i]~dnorm(1, 0.001)}
# Likelihood of the survival time data
for (j in 1:n) {
HRx[j]<-exp(x[1,j]*beta[1]+x[3,j]*beta[3]+x[5,j]*beta[5])
x[2,j]~dnorm(0,1)
x[4,j] ~dnorm(0,1)
x[6,j]~dnorm(0,1)
lambda[j]<-HRx[j]
t.obs[j]~dweib(r, lambda[j])I(t.cen[j], ) }
DATA list(k=6,m=9,n=24,
t.obs=c(NA, 13, NA, 13, 17, 15, NA, 14, 12, NA, 14, 13, 13, 15, 3, NA, 16, 15, 15, NA,
15, 12, 14, 12),
```

```
t.cen=c(18, 0, 19, 0, 0, 0, 18, 0, 0, 19, 0, 0, 0, 0, 0, 20, 0, 0, 0, 18, 0, 0, 0, 0),
x=structure(.Data=c(1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0,
5.8, 5.6, 6, 5.5, 8, 7, 7.1, 7.3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 6.2, 5.9, 6.4, 6, 5.7, 8.1, 5.9, 6.1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5.6, 6, 6, 7, 7.1, 5.8, 8, 6.7),.Dim=c(6, 24))
)
INITIAL VALUES list(r=0.1, beta=c(0, 0, 0, 0, 0, 0))
```

**Πίνακας 5.21:** Κώδικας WINBUGS για το παραμετρικό μοντέλο βασισμένο στην κατανομή Weibull (β), χωρίς τη χρήση της μεταβλητής $X_2$ = ορμόνη *TSH*

Το παραπάνω μοντέλο έχει *DIC* = 1269.34, δηλαδή μεγαλύτερο από το προηγούμενο πλήρες μοντέλο, το οποίο τελικά είναι καλύτερο (*DIC* = 128.714) κι έτσι έχουμε τα ακόλουθα συμπεράσματα, όπου παίζει ρόλο η μέση τιμή της ορμόνης *TSH* (6.45 μU/ml).

Σύμφωνα με τα Γραφήματα 5.61 – 5.67 οι προσομοιωμένες τιμές συγκλίνουν.

Από τον Πίνακα 5.20 προκύπτει ότι η εκτιμώμενη συνάρτηση κινδύνου με βάση τους εκ των υστέρων μέσους είναι:

- Για τους ασθενείς που τους χορηγήθηκε η $1^{\eta}$ θεραπεία:

$$\hat{h}_T\left(t \mid X\right)_1 = 3.989t^{3.989-1}\exp\left(-11.85 + 0.006175X_2\right)$$

- Για τους ασθενείς που τους χορηγήθηκε η $2^{\eta}$ θεραπεία:

$$\hat{h}_T\left(t \mid X\right)_2 = 3.989t^{3.989-1}\exp\left(-12.07 + 0.1093X_2\right)$$

- Για τους ασθενείς που τους χορηγήθηκε η $3^{\eta}$ θεραπεία:

$$\hat{h}_T\left(t \mid X\right)_3 = 3.989t^{3.989-1}\exp\left(-9.541 - 0.2369X_2\right)$$

Συνεπώς προκύπτουν τρεις λόγοι στιγμιαίων κινδύνων:

- Για $1^{\eta}$ -$2^{\eta}$ θεραπεία:

$$HR_{12} = \frac{\hat{h}_T\left(t \mid X\right)_1}{\hat{h}_T\left(t \mid X\right)_2} = \exp(0.22 - 0.1026X_2)$$

- Για $1^{η}$ – $3^{η}$ θεραπεία:

$$HR_{13} = \frac{\hat{h}_T(t \mid X)_1}{\hat{h}_T(t \mid X)_3} = \exp(-2.309 + 0.243X_2)$$

- Για $2^{η}$ -$3^{η}$ θεραπεία:

$$HR_{23} = \frac{\hat{h}_T(t \mid X)_2}{\hat{h}_T(t \mid X)_3} = \exp(-2.529 + 0.3462X_2)$$

Για την μέση τιμή της ορμόνης *TSH* (6.45 μU/ml) τα hazard ratios είναι: $HR_{12} = 0.64$, $HR_{13} = 0.48$ και $HR_{23} = 0.74$ .

Άρα ο ρυθμός ανάρρωσης των ασθενών που ακολούθησαν την $1^{η}$ θεραπεία είναι κατά 36% μικρότερος του ρυθμού ανάρρωσης για αυτούς που ακολούθησαν την $2^{η}$ θεραπεία (καλύτερη η $2^{η}$) και 52% μικρότερος του ρυθμού ανάρρωσης για αυτούς που ακολούθησαν την $3^{η}$ θεραπεία (καλύτερη η $3^{η}$). Ο ρυθμός ανάρρωσης των ασθενών που ακολούθησαν την $2^{η}$ θεραπεία είναι 26% μικρότερος του ρυθμού ανάρρωσης για αυτούς που ακολούθησαν την $3^{η}$ θεραπεία (καλύτερη η $3^{η}$). Έτσι προκύπτει πως καλύτερη θεραπεία είναι η $3^{η}$.

Τα αποτελέσματα για τις παραμέτρους *HR, RR* και *Sr,* προκύπτουν αν κατασκευαστούν 6 νέες μεταβλητές *x.r[, j]* με *j = 1, 2,…6*, κατά παρόμοιο τρόπο με αυτές του Παραδείγματος 5.8α

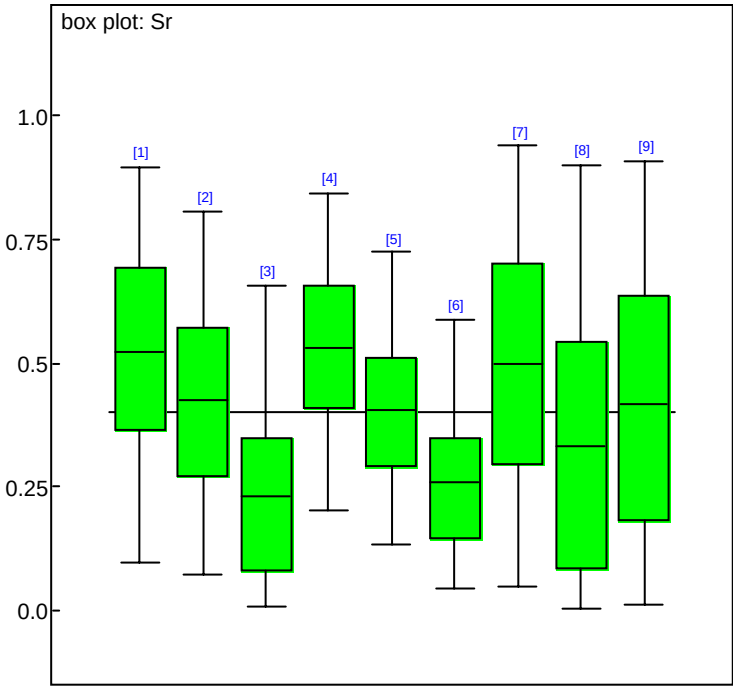
Για τιμές του *TSH* στις μεταβλητές *x.r[, 2], x.r[ ,4]* και *x.r[, 6]* λαμβάνουμε την ελάχιστη τιμή = 5.5, την διάμεσο = 6.05 και την μέγιστη τιμή = 8.1. Επίσης χρησιμοποιείται και η μεταβλητή *t.r[ ],* που παίρνει την τιμή 17 και για τις 9 περιπτώσεις. Χρησιμοποιείται η τιμή 17, διότι οι παρατηρήσεις με χρόνους πάνω από 17 ημέρες είναι περικομμένες. Έτσι θα υπολογιστούν εκ των υστέρων κατανομές των ρυθμών ανάρρωσης *HR,* των πιθανοτήτων μη ανάρρωσης *Sr* και των *RR* για χρονική περίοδο 17 ημερών.

Προέκυψαν τα εξής αποτελέσματα:

| **Θερ.** | **TSH** | **node** | **mean** | **sd** | **MC error** | **2.5%** | **median** | **97.5%** | **start** | **sample** |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.5 | *Sr[1]* | 0.5201 | 0.22 | 0.013 | 0.095 | 0.5297 | 0.896 | 2001 | 20000 |
| 2 | 5.5 | *Sr[2]* | 0.4214 | 0.2 | 0.01 | 0.068 | 0.4166 | 0.8031 | 2001 | 20000 |
| 3 | 5.5 | *Sr[3]* | 0.236 | 0.19 | 0.009 | 0.007 | 0.1938 | 0.6626 | 2001 | 20000 |
| 1 | 6.05 | *Sr[4]* | 0.5302 | 0.17 | 0.007 | 0.202 | 0.5331 | 0.8413 | 2001 | 20000 |
| 2 | 6.05 | *Sr[5]* | 0.4032 | 0.16 | 0.004 | 0.132 | 0.3957 | 0.7204 | 2001 | 20000 |
| 3 | 6.05 | *Sr[6]* | 0.2611 | 0.15 | 0.005 | 0.044 | 0.2396 | 0.5891 | 2001 | 20000 |
| 1 | 8.1 | *Sr[7]* | 0.5009 | 0.25 | 0.015 | 0.048 | 0.505 | 0.9414 | 2001 | 20000 |
| 2 | 8.1 | *Sr[8]* | 0.3371 | 0.27 | 0.016 | 0.002 | 0.2769 | 0.904 | 2001 | 20000 |
| 3 | 8.1 | *Sr[9]* | 0.3119 | 0.27 | 0.016 | 0,.009 | 0.3927 | 0.9091 | 2001 | 20000 |

**Πίνακας 5.22:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των πιθανοτήτων να μην αναρρώσει ένας ασθενής, δηλ. τα *Sr[i], i = 1, 2,…9*

Από τη συνάρτηση επιβίωσης σε κάθε μία από τις τιμές της ορμόνης *TSH,* η πιθανότητα κάποιος ασθενής να μην αναρρώσει εντός των 17 ημερών από την έναρξη της θεραπείας, είναι μεγαλύτερη για την $1^{η}$ θεραπεία και μικρότερη για την $3^{η}$ θεραπεία, αφού *Sr[1] > Sr [3], Sr[4] > Sr[6]* και *Sr[7] > Sr[9],* κάτι που βλέπουμε και στο Γράφημα 5.75 που ακολουθεί.

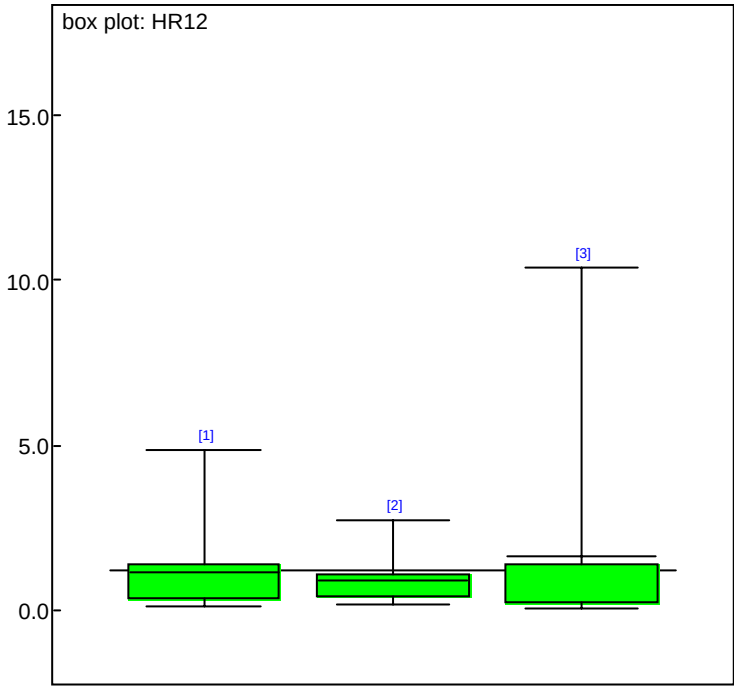



**Γράφημα 5.75:** Γραφική παράσταση των παραμέτρων *Sr[i], i = 1,2,…9* δηλαδή των πιθανοτήτων ένας ασθενής να μην αναρρώσει εντός 17 ημερών

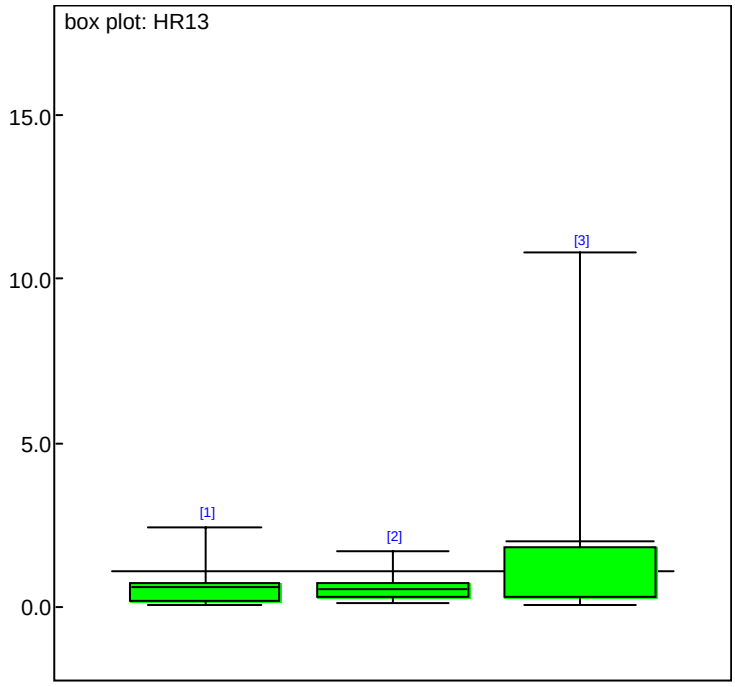
Ακολουθούν οι εκ των υστέρων κατανομές των παραμέτρων *HR* και *RR*, όπως και οι αντίστοιχες γραφικές παραστάσεις:

| node | mean | sd | MC error | 2.5% | median | 97.5% | start | sample |
|---|---|---|---|---|---|---|---|---|
| $HR_{12[1]}$ | 1.167 | 1.493 | 0.06857 | 0.09 | 0.724 | 4.868 | 2001 | 20000 |
| $HR_{12[2]}$ | 0.8718 | 0.7179 | 0.02276 | 0.1469 | 0.6806 | 2.753 | 2001 | 20000 |
| $HR_{12[3]}$ | 1.675 | 5.252 | 0.2523 | 0.0312 | 0.528 | 10.49 | 2001 | 20000 |
| $HR_{13[1]}$ | 0.6032 | 0.6898 | 0.03128 | 0.048 | 0.3948 | 2.451 | 2001 | 20000 |
| $HR_{13[2]}$ | 0.5574 | 0.4435 | 0.01482 | 0.0977 | 0.4417 | 1.726 | 2001 | 20000 |
| $HR_{13[3]}$ | 1.982 | 6.272 | 0.333 | 0.04712 | 0.7079 | 10.62 | 2001 | 20000 |
| $HR_{23[1]}$ | 0.8125 | 0.9841 | 0.0465 | 0.09008 | 0.5375 | 3.155 | 2001 | 20000 |
| $HR_{23[2]}$ | 0.7994 | 0.5869 | 0.01776 | 0.1786 | 0.6491 | 2.314 | 2001 | 20000 |
| $HR_{23[3]}$ | 3.97 | 13.04 | 0.6873 | 0.07732 | 1.333 | 22.65 | 2001 | 20000 |
| $RR_{12[1]}$ | 0.9891 | 0.7651 | 0.03658 | 0.1652 | 0.8205 | 2.891 | 2001 | 20000 |
| $RR_{12[2]}$ | 0.8615 | 0.4597 | 0.01584 | 0.2448 | 0.7848 | 1.978 | 2001 | 20000 |
| $RR_{12[3]}$ | 1.287 | 3.373 | 0.1753 | 0.0843 | 0.7414 | 5.859 | 2001 | 20000 |
| $RR_{13[1]}$ | 0.6827 | 0.4146 | 0.0207 | 0.1282 | 0.6226 | 1.677 | 2001 | 20000 |
| $RR_{13[2]}$ | 0.6682 | 0.3072 | 0.01141 | 0.2039 | 0.6329 | 1.384 | 2001 | 20000 |
| $RR_{13[3]}$ | 1.374 | 2.656 | 0.1508 | 0.1048 | 0.8367 | 5.805 | 2001 | 20000 |
| $RR_{23[1]}$ | 0.8284 | 0.4567 | 0.02164 | 0.2462 | 0.7546 | 1.94 | 2001 | 20000 |
| $RR_{23[2]}$ | 0.85 | 0.3276 | 0.01022 | 0.3601 | 0.8084 | 1.612 | 2001 | 20000 |
| $RR_{23[3]}$ | 1.821 | 3.296 | 0.1893 | 0.1688 | 1.112 | 7.686 | 2001 | 20000 |

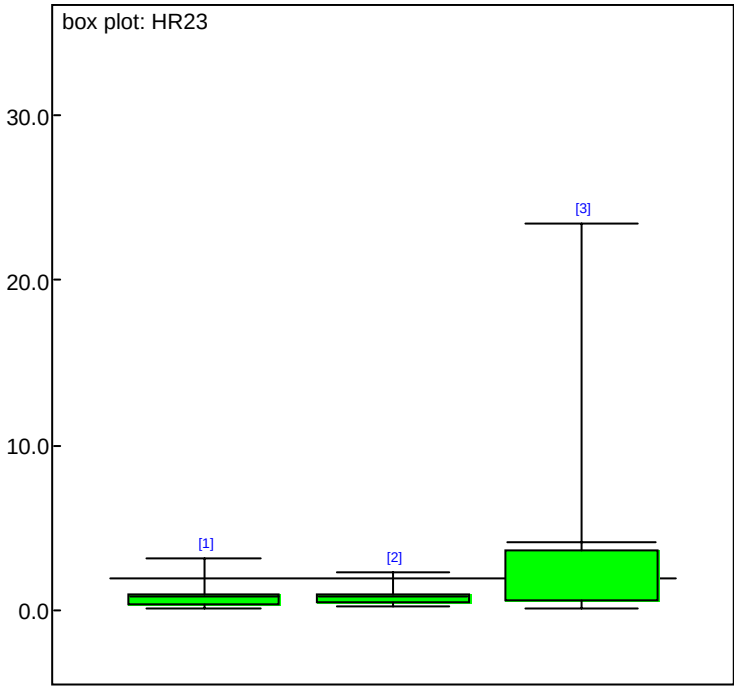
**Πίνακας 5.23:** Περιγραφικοί δείκτες των εκ των υστέρων κατανομών των $HR_{12[i]}$, $HR_{13[i]}$, $HR_{23[i]}$, $RR_{12[i]}$, $RR_{13[i]}$ και $RR_{23[i]}$, $i = 1, 2, 3$ δηλαδή των ρυθμών ανάρρωσης και των σχετικών κινδύνων



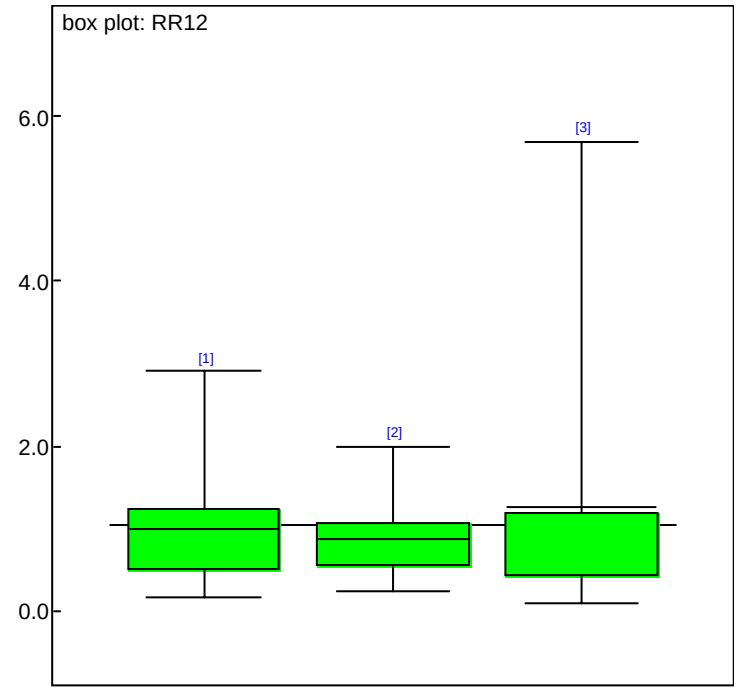

**(5.76):** Hazard ratios μεταξύ θεραπειών 1 και 2



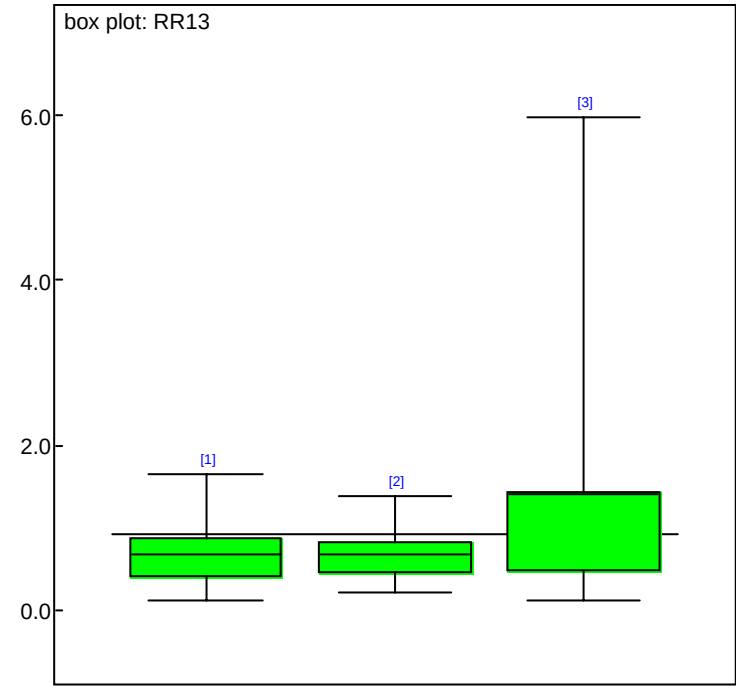

**(5.77):** Hazard ratios μεταξύ θεραπειών 1 και 3



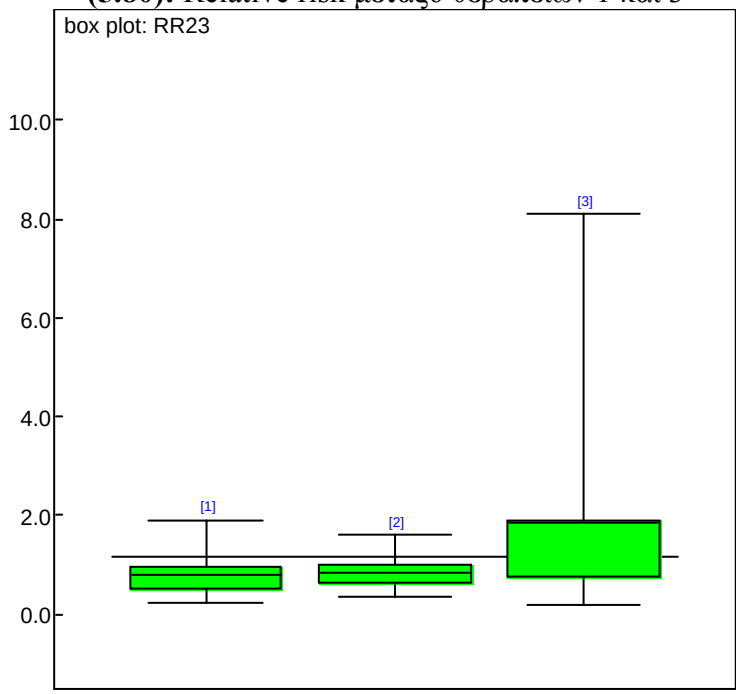

**(5.78):** Hazard ratios μεταξύ θεραπειών 2 και 3




**(5.79):** Relative risk μεταξύ θεραπειών 1 και 2




**(5.80):** Relative risk μεταξύ θεραπειών 1 και 3




**(5.81):** Relative risk μεταξύ θεραπειών 2 και 3

**Γραφήματα 5.76 – 5.81:** Γραφικές παραστάσεις των παραμέτρων $HR_{12}$, $HR_{13}$, $HR_{23}$, $RR_{12}$, $RR_{13}$ και $RR_{23}$ αντίστοιχα, δηλαδή των ρυθμών ανάρρωσης και των σχετικών κινδύνων για τις 3 τιμές της ορμόνης *TSH*

Από τον Πίνακα 5.23 βλέπουμε πως γενικά οι ρυθμοί ανάρρωσης δεν διαφέρουν σημαντικά, για σχεδόν όλες τις τιμές της ορμόνης *TSH*, παρά μόνο για τη μέγιστη τιμή της, αφού ο ρυθμός ανάρρωσης με την $1^{η}$ θεραπεία είναι 67% μεγαλύτερος απ' ότι με τη $2^{η}$ και σχεδόν διπλάσιος από την $3^{η}$ (1.982). Ειδικά ο ρυθμός ανάρρωσης με την $2^{η}$ θεραπεία είναι σχεδόν τετραπλάσιος από τον αντίστοιχο ρυθμό ανάρρωσης των ασθενών της $3^{ης}$ θεραπείας.

Το *RR* είναι ο λόγος $\frac{1-S_k(17)}{1-S_m(17)}$, για 2 θεραπείες *k, m = 1, 2, 3* και δείχνει την αναλογία της πιθανότητας ένας ασθενής με την θεραπεία *k*, να αναρρώσει σε λιγότερο από 17 ημέρες σε σχέση με έναν ασθενή με την θεραπεία *m*.

Από τα *RR* προκύπτει πως η $3^{η}$ θεραπεία είναι καλύτερη για κάθε τιμή της ορμόνης, αφού $RR_{13} < 1$ και $RR_{23} < 1$. Η τελευταία σχέση δεν ισχύει, μόνο για τη μέγιστη τιμή της ορμόνης ($RR_{23[3]} = 1.821$) Δηλαδή υπάρχει διαφορά μόνο για $2^{η} - 3^{η}$ θεραπεία για τη μέγιστη τιμή 8.1μU/ml, όπως εξάλλου φαίνεται και από τα Γραφήματα 5.79 – 5.81.

# ΕΠΙΛΟΓΟΣ - ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα εργασία αναπτύχθηκαν οι προσεγγίσεις κατά Bayes εφαρμοσμένες σε δεδομένα επιβίωσης. Αρχικά λοιπόν, αναλύθηκαν οι θεωρητικές κατανομές, οι οποίες προσεγγίζουν καλύτερα τέτοιου είδους δεδομένα κι έπειτα ακολούθησε μία πλήρη ανάπτυξη του θεωρήματος Bayes και των εφαρμογών του. Στη συνέχεια, σκοπός ήταν να αναλυθούν οι μέθοδοι που περιγράφηκαν, με όσο το δυνατόν περισσότερο κατανοητό τρόπο, μέσα από παραδείγματα πάνω σε ιατρικά δεδομένα. Βασικός στόχος της ανάλυσης ήταν η εκτίμηση της συνάρτησης επιβίωσης και της συνάρτησης κινδύνου, αλλά και η εκτίμηση των παραμέτρων που εμφανίζονται σε κάθε κατανομή, καθώς και η συμπερασματολογία σχετικά με τις μεταβλητές που τις επηρεάζουν και τις καθορίζουν.

Για τις παραπάνω αναλύσεις χρησιμότατο εργαλείο αποτέλεσε το στατιστικό πρόγραμμα WINBUGS, στο οποίο γίνεται χρήση προσομοιωμένων τιμών από τις εκ των υστέρων κατανομές των αγνώστων παραμέτρων του κάθε μοντέλου. Αρχικά το WINBUGS χρησιμοποιήθηκε για τη σύγκριση των 5 κατανομών που παρουσιάστηκαν αναλυτικά στο 1$^{ο}$ Κεφάλαιο. Η σύγκριση των κατανομών αυτών υλοποιήθηκε με τη χρήση του Deviance Information Criterion (*DIC*) και ως καταλληλότερη κρίθηκε η κατανομή Weibull, για το σετ δεδομένων που χρησιμοποιήθηκε.

Στη συνέχεια αναλύθηκαν οι μέθοδοι εκτίμησης της συνάρτησης επιβίωσης $S(t)$, καθώς και της συνάρτησης κινδύνου $h_T(t)$. Για την εκτίμηση της $h_T(t)$ χρησιμοποιήθηκε η προσέγγιση του Cox με το παλινδρομικό μοντέλο αναλογικού κινδύνου: $h_T(t \mid \underset{\sim}{X}) = h_0(t)\exp\left(b_1X_1 + b_2X_2 + \ldots b_pX_p\right)$. Με το υπόδειγμα αυτό μπορούμε να μελετήσουμε τις επιδράσεις ανεξάρτητων μεταβλητών $X_i$ $(i = 1, 2, \ldots p)$ πάνω στη συνάρτηση κινδύνου. Παρουσιάστηκαν δύο στατιστικές τεχνικές για την ανάλυση δεδομένων με το παραπάνω μοντέλο: η μη παραμετρική και η παραμετρική με τη χρήση της κατανομής Weibull, η οποία προτιμάται πιο συχνά από τις υπόλοιπες σε δεδομένα επιβίωσης.

Το μη παραμετρικό μοντέλο εφαρμόστηκε σε 42 χρόνους επιβίωσης ασθενών, στους οποίους χορηγήθηκε μία θεραπεία και το placebo και μετρήθηκε ο χρόνος μέχρι το θάνατό τους. Σκοπός ήταν να βρεθούν εκτιμήσεις για τις αυξήσεις των αριθμών των

θανάτων μέσα σε πολύ μικρά διαστήματα της μορφής *[t, t + dt),* καθώς και για τις παραμέτρους *b* και $\Lambda_0(t) = \int_0^t h_0(u)du$, δηλαδή το συντελεστή της μεταβλητής *X* και την αθροιστική συνάρτηση κατανομής του επιπέδου αναφοράς της συνάρτησης κινδύνου $h_0(t)$. Τα αποτελέσματα στο WINBUGS έδωσαν ότι η αναμενόμενη αύξηση των θανάτων με το placebo είναι μεγαλύτερη από την αντίστοιχη αναμενόμενη αύξηση θανάτων με την πραγματική θεραπεία, όπως επίσης και ότι η συνάρτηση επιβίωσης *S(t)* έχει καλύτερες τιμές για την θεραπεία απ' ότι για το placebo.

Στο παραμετρικό μοντέλο το επίπεδο αναφοράς της συνάρτησης κινδύνου $h_0(t)$ είδαμε ότι ακολουθεί την κατανομή Weibull με παραμέτρους *α* και *r = 1*. Κι στα δύο παραδείγματα που αναλύθηκαν για την παρουσίαση του μοντέλου, δημιουργήθηκαν λόγοι στιγμιαίου κινδύνου (hazard ratios) για τις συγκρίσεις μεταξύ των επιπέδων των ανεξάρτητων μεταβλητών $X_{i,,}$ όπως επίσης δόθηκαν εκτιμήσεις και για τους συντελεστές $b_i$, τη συνάρτηση επιβίωσης *S(t)* και το σχετικό κίνδυνο (relative risk).

Στο πρώτο παράδειγμα αναλύθηκαν χρόνοι επιβίωσης ασθενών με πολλαπλό μελάνωμα στους οποίους εκτός του χρόνου επιβίωσης μετρήθηκαν οι τιμές αζώτου στο αίμα τους και χωρίστηκαν σε δύο ομάδες, ανάλογα με το φύλο τους. Η ανάλυση υπέδειξε ότι οι τιμές του αζώτου ήταν τελικά σημαντικές για τους άντρες ασθενείς. Όσο αυτές μεγάλωναν, μειωνόταν ο ρυθμός επιβίωσης των αντρών. Η σύγκριση μεταξύ γυναικών - αντρών έδωσε ότι γενικά ο ρυθμός θανάτου των γυναικών είναι μεγαλύτερός από το ρυθμό θανάτου των αντρών, όμως για μεγάλες τιμές του αζώτου στο αίμα (πάνω από 21 που είναι η διάμεσος) οι γυναίκες διατρέχουν μικρότερο κίνδυνο.

Τέλος, στο δεύτερο παράδειγμα αναλύθηκαν χρόνοι μέχρι την εξάλειψη των συμπτωμάτων 24 υποθυρεοειδικών ασθενών, στους οποίους εφαρμόστηκαν τρεις διαφορετικές θεραπείες, αφού μετρήθηκε η τιμή της ορμόνης *TSH*. Αρχικά είδαμε ότι η ποσότητα της ορμόνης που μετρήθηκε ουσιαστικά δεν είναι σημαντική μεταβλητή (οι συντελεστές $b_i$ μπορούν να θεωρηθούν ότι είναι μηδενικοί) και ότι ο ρυθμός ανάρρωσης εξαρτάται τελικά από το είδος της θεραπείας που εφαρμόστηκε. Έτσι από τα αποτελέσματα των ρυθμών ανάρρωσης των ασθενών είδαμε ότι καλύτερη ήταν η τρίτη θεραπεία, αφού ο ρυθμός ανάρρωσης με την πρώτη και τη δεύτερη θεραπεία ήταν

52% και 26% αντίστοιχα μικρότερος από τον αντίστοιχο ρυθμό ανάρρωσης των ασθενών της τρίτης θεραπείας.

Από όλα τα παραπάνω, ουσιαστικά καταλήγουμε στο συμπέρασμα ότι οι μη Μπεϋζιανές μέθοδοι, όπως είναι η μέθοδος των ελαχίστων τετραγώνων ή η μέθοδος μεγιστοποίησης της συνάρτησης πιθανοφάνειας, ενώ είναι πολύ πιο εύκολες, ως προς την εκτίμηση, δεν προσφέρουν το βασικό πλεονέκτημα, το οποίο προσφέρει η Μπεϋζιανή προσέγγιση: Ότι δηλαδή, σε μία μελέτη μπορούν να συμπεριληφθούν πληροφορίες από παλαιότερες έρευνες, κάτι το οποίο, ειδικά όσον αφορά ιατρικά δεδομένα όπως αυτά που αναλύθηκαν στην παρούσα εργασία, είναι πολύ χρήσιμο. Και είδαμε πως με τη χρήση του στατιστικού προγράμματος WINBUGS και μέσω της προσομοίωσης τιμών από τις εκ των υστέρων κατανομές των αγνώστων παραμέτρων του κάθε μοντέλου, τα υπολογιστικά προβλήματα που παρουσιάζονται βρίσκουν λύση. Γι' αυτό λοιπόν, η Μπεϋζιανές μέθοδοι κρίνονται πλέον καταλληλότερες και το ενδιαφέρον των αναλυτών στρέφεται ολοένα και περισσότερο γύρω από αυτές.

## ΒΙΒΛΙΟΓΡΑΦΙΑ – ΑΝΑΦΟΡΕΣ


- **Abrams K., Ashby D. & Errington D. (1996)** A Bayesian approach to Weibull survival models*, Lifetime Data Analysis*, **2**, p. 159 – 174.
- **Ahrens C. & Dieter U. (1974)** Computer methods for sampling from Gamma, Beta, Poisson and Binomial distributions*, Springer-Verlag,* **12**, p. 223 – 246.
- **Aitkin M. (1991)** Posterior Bayes factors*, Journal of the Royal Statistical Society*, *B,* 53, p. 111 – 142.
- **Aitkin M. (1981)** A note on the regression analysis of censored data*, Technometrics*, Vol. **23**, No. 2, p. 161-163.
- **Altman DG. (1991)** Practical Statistics for Medical Research, *London: Chapman & Hall.*
- **Bernardinelli L., Clayton D.G & Montomoli C. (1995)** Bayesian estimates of Disease maps; how important are priors?*, John Wiley & Sons, Ltd*
- **Bernardo J.M. & Smith F. M. (1994)** Bayesian Theory*, John Wiley & Sons Inc, New York.*
- **Berry D. & Stangl D. (1996)** Bayesian Biostatistics (Statistics, a Series of Textbooks and Monographs) *Institute of Statistics and Decision Sciences*
- **Bhat N. (1972)** Elements of Applied Stochastic Processes*, J. Wiley & Sons, New York.*
- **Box G. & Tiao C. (1973)** Bayesian Inference in Statistical Analysis, *Reading, MA: Addison-Wesley Publishing Company.*
- **Breslow N. E. (1975)** Analysis of survival data under the proportional hazards model, *Biometrics*, **30**, p. 89-99.
- **Carlin B. & Louis Th. (1996)** Bayes and Empirical Bayes methods for data analysis. *Chapman & Hall, London*
- **Chen M., Ibrahim J. & Sinha D. (2001)** Bayesian Survival Analysis, *International Journal of Epidemiology, Oxford Journals*, **31**, p.505
- **Chen M., Ibrahim J. & Shao Q. (2000)** Monte Carlo methods in Bayesian computation, *Springer - Verlag*

- **Christensen R. (1997)** Log-linear models and logistic regression*, Springer Texts in Statistics, New York: Springer*
- **Clayton D. (1994)** Bayesian analysis of frailty models*, Technical report, Medical Research Council Biostatistics Unit, Cambridge.*
- **Clayton D. (1991)** A Monte Carlo method for Bayesian inference in frailty models*, Biometrics*, **47**, p. 467 – 485.
- **Collett D. (1993)** Modelling Survival Data in Medical Research, *Chapman & Hall, London.*
- **Congdon P. (2001)** Bayesian Statistical Modelling, *John Wiley & Sons Ltd*.
- **Cox D. R. (1975)** Partial likelihood, *Biometrika, Department of Mathematics, Imperial College, London,* **62**, p. 2 - 10.
- **Cox D.R. & Miller H.D. (1965)** The theory of Stochastic processes*, Mathuen & Co, London.*
- **Cox DR & Oakes D. (1984)** Analysis of Survival data, *Chapman & Hall.*
- **Efron B. (1977)** The efficiency of Cox's likelihood function for censored data, *Journal of the Royal Statistical Society*, *B,* **34**, p. 187 – 220.
- **Fryback D.G., Stout N.K. & Rosenberg M.A. (2001)** An elementary introduction to Bayesian computing using Winbugs, *International Journal of Technology Assessment in Health Care,* **17**, p. 98-113.
- **Gatsonis C., Kass R., Carlin B., Carriquiry A., Gelman A., Verdinelli I. & West M. (2000)** Case Studies in Bayesian Statistics*, Springer-Verlag, New York,* **4***,* p. 133 - 204.
- **Gelman A., Carlin B., Rubin D. & Stern H. (1995)** Bayesian Data Analysis*, Chapman & Hall, London*
- **Gilks W.R., Richardson S. & Spiegelhalter D.J. (1996)** Markov Chain Monte Carlo in practice, *Chapman & Hall, London.*
- **Gradshteyn I.S. & Ryzhik I. M. (2000)** Table of Integrals, Series, and Products, *A. Jeffrey, Ed. New York: Academic Press.*
- **Gross A.J. & Clark V.A. (1975)** Survival distributions: Reliability applications in the Biomedical Science, *New York – Wiley.*
- **Johansen, S. (1983)** An extension of Cox's regression model, *International Statistical Revie,. Kalbfleisch,* **51**, 165-174

- **Johnson N.L. & Kotz S. (1970)** Distributions in Statistics: Continuous univariate distributions (2 vols), *Boston – Houghton Mifflin.*
- **Kalbfleisch, J. (1978)** Non-parametric Bayesian analysis of survival time data*, Journal of the Royal Statistical Society*, *B,* **40**, p. 214-221.
- **Kalbfleisch J. and Prentice R. (1980)** The statistical analysis of failure time data*, Wiley, New York.*
- **Karlin S. & Taylor H. (1975)** A first course in Stochastic Processes*, Academic Press, New York.*
- **Kaplan E.L. & Meier P. (1958)** Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, p. 457 – 481.
- **Laird N. (1982)** Empirical Bayes estimates using the nonparametric maximum likelihood estimate for the prior, *Journal of the Statistical Computation and Simulation,* **6**, p. 211 – 220.
- **Lawson A., Browne W. & Vidal Rodeiro C. (2003)** Disease Mapping with Winbugs and MLWin, *John Wiley & Sons, New York.*
- **Martz H. & Waller R. (1982)** Bayesian Reliability Analysis, *John Wiley & Sons, New York*
- **Owen A. (1988)** Empirical Likelihood Ratio Confidence Intervals for a Single Functional, *Chapman & Hall, London,* **75**, p. 237-249
- **Owen A. (2001)** Empirical likelihood*, Chapman & Hall, London.*
- **Pan X.R. (1997)** Empirical Likelihood Ratio Method for Censored Data*, Ph.D. Thesis*, *Department of Statistics, University of Kentucky, USA.*
- **Pan X.R. and Zhou M. (1999)** Empirical likelihood in terms of cumulative hazard function for censored data, *Department of Statistics, University of Kentucky, USA.*
- **Press J. (1989)** Bayesian Statistics, *Wiley New York.*
- **Raftery A. (1995)** Bayesian model selection in Social Research*, Sociological Methodology,* **25**, p. 111- 163.
- **Robert C.P. & Gasella G. (1999)** Monte Carlo statistical methods*, Springer Verlang, New York.*

- **Spigelhalter D., Thomas A., Best N. & Lunn D. (2002)** Winbugs User Manual – Version 1.4 *Winbugs 1.4 / Help menu / Examples / Vol.1, MRC Biostatistics Unit.*
- **Spiegelhalter D, Best NG, Carlin BP & Van der Linde A. (2002)** Bayesian measures of model complexity and fit (with discussion)*, Journal of the Royal Statistical Society*, *B,* **64**, p. 583-640.
- **Susarla V. & Van Ryzin J. (1976)** Nonparametric Bayesian estimation of survival curves from incomplete observations, *Journal of the American Statistical Association,* **71**, p. 897 – 902.
- **Tierney L. (1983)** A space-efficient recursive procedure for estimating a quantile of an unknown distribution, *Journal of the Statistical Computation*, **4**, p. 706-711.
- **Volinsky T. & Raftery A. (2000)** Bayesian Information Criterion for censored survival models, *Biometrics*, **56,** p. 256 – 262.
- **Woodworth George (2005)** Biostatistics – A Bayesian Introduction, *Published by Wiley-Interscience, an imprint of John Wiley & Sons, Hoboken,* **47,** p. 267 - 328.
- **Δημάκη Κ. (2006)** Ανάλυση Επιβίωσης*, Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών.*
- **Κάκουλος Θ.** (1978) Στοχαστικές Ανελίξεις*, Αθήνα.*
- **Λάγκαρης Χρ. (1988)** Θεωρία Στοχαστικών Διαδικασιών*, Εκδόσεις Πανεπιστημίου Ιωαννίνων.*
- **Παπαϊωάννου Τ. (1997)** Θεωρία Πιθανοτήτων και Στατιστικής, *Εκδόσεις Σταμούλη, Αθήνα.*
- **Παπαϊωάννου Τ. & Φερεντίνος Κ. (1983)** Μαθηματική Στατιστική*, Εκδόσεις Πανεπιστημίου Ιωαννίνων*.


**ΙΣΤΟΣΕΛΙΔΕΣ:**


- home.uchicago.edu
- www.columbia.edu
- www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml

- www.statsoft.com
- www.weibull.com
- www.wikipedia.com