Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΚΙΝΔΥΝΩΝ**

ΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ ΤΗΣ ΥΠΟΘΕΣΗΣ ΤΟΥ ΑΝΑΛΟΓΙΚΟΥ ΚΙΝΔΥΝΟΥ

Γεώργιος Π. Παπαδόπουλος

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην *Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων*

Πειραιάς
Ιανουάριος 2015

Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΚΙΝΔΥΝΩΝ**

ΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ ΤΗΣ ΥΠΟΘΕΣΗΣ ΤΟΥ ΑΝΑΛΟΓΙΚΟΥ ΚΙΝΔΥΝΟΥ

Γεώργιος Π. Παπαδόπουλος

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην *Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων*

Πειραιάς
Ιανουάριος 2015

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη Συνέλευση του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. …….. συνεδρίασή της, σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων.

Τα μέλη της Επιτροπής ήταν:

- ……………………………….. (Επιβλέπων)
- ………………………………..
- ………………………………..

Η έγκριση της Διπλωματική Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

UNIVERSITY OF PIRAEUS

School of Finance and Statistics

Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN
ACTUARIAL SCIENCE AND RISK MANAGEMENT

GRAPHICAL METHODS FOR CHECKING THE PROPORTIONAL HAZARDS ASSUMPTION

By

George P. Papadopoulos

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Actuarial Science and Risk Management

Piraeus, Greece
January 2015

Στους γονείς μου

Ιωάννη και Μαρία

Ευχαριστίες

{Αντικαταστήστε αυτό το κείμενο με τις ευχαριστίες σας}

Περίληψη

{Αντικαταστήστε αυτό το κείμενο με την δική σας περίληψη}

Abstract

{Replace this text with your own abstract}

Περιεχόμενα

Κατάλογος Πινάκων

Κατάλογος Σχημάτων

Κατάλογος Συντομογραφιών

τ.μ.	τυχαία μεταβλητή
σ.π.π.	συνάρτηση πυκνότητας πιθανότητας
DFR	Decreasing Failure Rate
	……………………….
	……………………….

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Συνάρτηση επιβίωσης και συνάρτηση κινδύνου

Έστω T μια (απολύτως) συνεχής τυχαία μεταβλητή με σύνολο τιμών $R_T = [0,\infty)$ η οποία δηλώνει το χρόνο ζωής ενός ατόμου. Η πιθανότητα $S(t)$ του ενδεχομένου $\{T \geq t\}$, $t \geq 0$, ονομάζεται συνάρτηση επιβίωσης (*survival function*) της τυχαίας μεταβλητής T και ορίζεται* από τη σχέση

$$S(t) = P(T \geq t), \quad t \geq 0.$$

Η συνάρτηση επιβίωσης $S(t)$ δηλώνει την πιθανότητα να είναι ο χρόνος ζωής ενός ατόμου μεγαλύτερος ή ίσος του χρόνου t. Στα πλαίσια της θεωρίας αξιοπιστίας η συνάρτηση $S(t)$ ονομάζεται συνάρτηση αξιοπιστίας (*reliability function*).

Η $S(t)$ συνδέεται με τη συνάρτηση κατανομής $F(t)$ της T σύμφωνα με την προφανή σχέση

$$S(t) = P(T \geq t) = 1 - P(T < t) = 1 - F(t-0) = 1 - F(t)$$

(όπου $F(t-0) = \lim_{x \to 0^-} F(t+x)$). Η συνάρτηση $S(t)$ είναι φθίνουσα συνάρτηση, συνεχής, $S(0) = 1$ και $\lim_{t\to\infty} S(t) = 0$ (αφού $F(0) = 0$ και $\lim_{t\to\infty} F(t) = 1$). Επίσης μπορούμε να γράψουμε ότι

$$S(t) = \int_t^\infty f(u)du$$

όπου $f(t)$ είναι η συνάρτηση πυκνότητας της τυχαίας μεταβλητής T.

Η $S(t)$ συνδέεται με την $f(t)$ σύμφωνα με τη σχέση $f(t) = -S'(t)$, αφού

$$f(t) = \frac{dF(t)}{dt} = \frac{d[1-S(t)]}{dt} = -\frac{dS(t)}{dt}.$$

* Ορισμένοι συγγραφείς ορίζουν τη συνάρτηση επιβίωσης ως $S(t) = P(T > t)$.

Μια άλλη βασική ποσότητα στην ανάλυση επιβίωσης είναι η συνάρτηση κινδύνου (*hazard function* ή *hazard rate*) της τυχαίας μεταβλητής T η οποία συμβολίζεται με $h(t)$ και ορίζεται με τη σχέση

$$h(t) = \lim_{\Delta t \to 0^+} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}, \quad t \ge 0.$$

Η $h(t)$ δηλώνει τη στιγμιαία πιθανότητα "θανάτου" ενός ατόμου το χρόνο t δοθέντος ότι αυτό επέζησε μέχρι τη χρονική στιγμή t. Η ποσότητα $h(t)\Delta t$ είναι προσεγγιστικά η πιθανότητα θανάτου ενός ατόμου στο διάστημα $[t, t + \Delta t)$ γνωρίζοντας ότι το άτομο έχει επιβιώσει μέχρι τη χρονική στιγμή t, δηλαδή

$$h(t)\Delta t \cong P(t \le T < t + \Delta t \mid T \ge t).$$

Η $h(t)$ εμφανίζεται στην θεωρία αξιοπιστίας ως (δεσμευμένη**)** βαθμίδα αποτυχίας ((*conditional*) *failure rate*), στη δημογραφία ως ένταση θνησιμότητας (*force of mortality*), στα οικονομικά ως αντίστροφος λόγος του Mill (*inverse Mill's ratio*) κτλ. Η $h(t)$ ικανοποιεί τη σχέση

$$h(t) = \frac{f(t)}{S(t)}$$

αφού

$$h(t) = \lim_{\Delta t \to 0^+} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} = \frac{1}{P(T \ge t)} \lim_{\Delta t \to 0^+} \frac{P(t \le T < t + \Delta t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

Επειδή $f(t) = -S'(t)$ προκύπτει άμεσα ότι η $h(t)$ συνδέεται με την $S(t)$ σύμφωνα με τη σχέση

$$h(t) = \frac{-S'(t)}{S(t)} = -\frac{d \log S(t)}{dt}. \qquad (1.1)$$

Ολοκληρώνοντας τα δύο μέλη της σχέσης (1.1) ως προς t και χρησιμοποιώντας τη συνθήκη $S(0) = 1$ προκύπτει ότι

$$\int_0^t h(u)du = \int_0^t -\frac{d \log S(u)}{du} du = -\log S(t)$$

οπότε

$$S(t) = \exp\left(-\int_0^t h(u)du\right).$$

Επίσης, αφού $f(t) = h(t)S(t)$, μπορούμε να γράψουμε ότι

$$f(t) = h(t)\exp\left(-\int_0^t h(u)du\right).$$

Η συνάρτηση

$$H(t) = \int_0^t h(u)du, \quad t \geq 0$$

ονομάζεται αθροιστική συνάρτηση κινδύνου (*cumulative hazard function*), και μπορούμε γράψουμε ότι

$$S(t) = \exp(-H(t)), \quad H(t) = -\log S(t).$$

Επειδή $\lim_{t\to\infty} S(t) = 0$ προκύπτει ότι $\lim_{t\to\infty} H(t) = \infty$. Συνεπώς για τη συνάρτηση $h(t)$ προκύπτουν οι ιδιότητες

$$h(t) \geq 0, \quad \int_0^\infty h(t)dt = \lim_{t\to\infty} H(t) = \infty.$$

Συνδυάζοντας όλα τα παραπάνω αποτελέσματα προκύπτει ο ακόλουθος πίνακας στον οποίο δίνονται οι σχέσεις με τις οποίες συνδέονται οι ποσότητες $f(t)$, $S(t)$, $h(t)$, $H(t)$.

ΠΙΝΑΚΑΣ 1-1

Σχέσεις μεταξύ των ποσοτήτων $f(t)$, $S(t)$, $h(t)$, $H(t)$.

	$f(t)$	$S(t)$	$h(t)$	$H(t)$
$f(t)$	$f(t)$	$-S'(t)$	$h(t)\exp\left(-\int_0^t h(u)du\right)$	$H'(t)\exp(-H(t))$
$S(t)$	$\int_t^\infty f(u)du$	$S(t)$	$\exp\left(-\int_0^t h(u)du\right)$	$\exp(-H(t))$
$h(t)$	$\dfrac{f(t)}{\int_t^\infty f(u)du}$	$-(\log S(t))'$	$h(t)$	$H'(t)$
$H(t)$	$-\log\left(\int_t^\infty f(u)du\right)$	$-\log S(t)$	$\int_0^t h(u)du$	$H(t)$

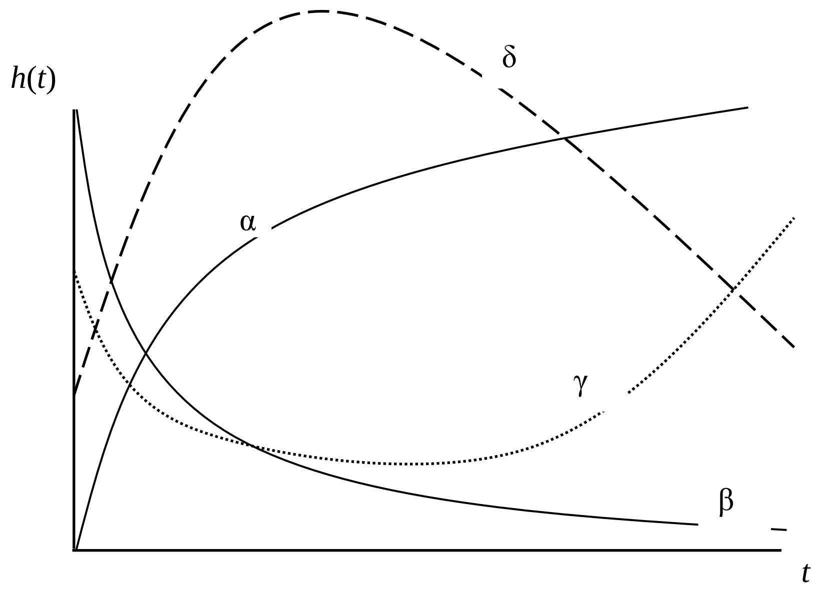
1.2 Μερικά σχόλια για τη συνάρτηση κινδύνου

Για την περιγραφή μιας τυχαίας μεταβλητής T χρησιμοποιούμε συνήθως τη συνάρτηση πυκνότητάς της ή τη συνάρτηση πιθανότητάς της $f(t)$ (ανάλογα με το αν είναι συνεχής ή διακριτή) ή ακόμη τη συνάρτηση επιβίωσής της $S(t)$. Η συνάρτηση κινδύνου $h(t)$ της *T,* αν και δε χρησιμοποιείται συχνά, είναι ιδιαιτέρως χρήσιμη για την περιγραφή κατανομών χρόνου ζωής (*lifetime distributions*) αφού δηλώνει τον τρόπο με τον οποίο μεταβάλλεται η στιγμιαία πιθανότητα θανάτου ενός ατόμου συναρτήσει του χρόνου. Σε πολλές εφαρμογές μπορεί να υπάρχουν ποιοτικές πληροφορίες για τη συνάρτηση κινδύνου οι οποίες μας βοηθούν στην επιλογή κατάλληλου παραμετρικού μοντέλου για την περιγραφή της κατανομής χρόνου ζωής.

Στο ακόλουθο διάγραμμα δίνονται τέσσερα βασικά είδη μορφών συναρτήσεων κινδύνου

ΣΧΗΜΑ 1-1

Τυπικές συναρτήσεις κινδύνου



Η κατανομή (α) έχει αύξουσα συνάρτηση κινδύνου (*increasing failure rate* ή *IFR*), η (β) φθίνουσα (*decreasing failure rate* ή *DFR*), η (γ) έχει συνάρτηση κινδύνου μορφής λεκάνης (*bathtub-shaped*), ενώ η (δ) έχει μορφή καμπούρας (*hump-shaped*). Κατανομές χρόνου ζωής με συναρτήσεις κινδύνου της μορφής (α)-(δ) είναι ιδιαίτερα χρήσιμες στην πράξη.

1.3 Ανασκόπηση παραμετρικών μοντέλων

1.3.1 Εκθετική κατανομή

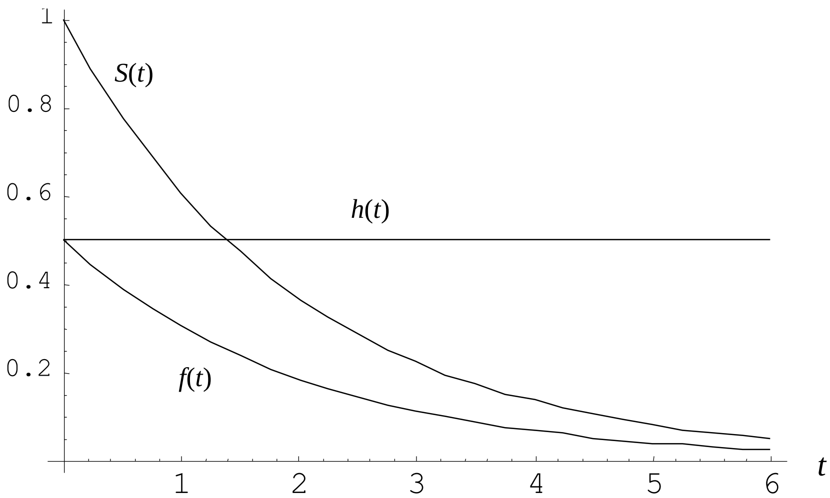
Η συνάρτηση πυκνότητας $f(t)$, η συνάρτηση επιβίωσης $S(t)$ και η συνάρτηση κινδύνου $h(t)$ της εκθετικής (*exponential*) κατανομής με παράμετρο λ δίνονται από τις σχέσεις

$$f(t) = \lambda e^{-\lambda t}, \qquad S(t) = e^{-\lambda t}, \quad h(t) = \lambda \tag{1.2}$$

Το κύριο χαρακτηριστικό της εκθετικής κατανομής είναι ότι έχει σταθερή συνάρτηση κινδύνου που οφείλεται στην ιδιότητα έλλειψης μνήμης της εκθετικής κατανομής. Χρησιμοποιώντας τις σχέσεις (1.2) για $\lambda = 0.5$ προκύπτει το ακόλουθο σχήμα.

ΣΧΗΜΑ 1-2

Εκθετική κατανομή με παράμετρο $\lambda = 0.5$



ΚΕΦΑΛΑΙΟ 2

Πίνακες επιβίωσης

2.1 Εισαγωγή

Οι πίνακες επιβίωσης (*life tables*) είναι μια από τις πρώτες και πιο διαδεδομένες μεθόδους στη δημογραφία και στον αναλογισμό για την περιγραφή δεδομένων που αφορούν χρόνους ζωής. Οι πίνακες επιβίωσης είναι μια "επέκταση" των συνηθισμένων πινάκων συχνοτήτων (*frequency tables*) στην περίπτωση που υπάρχουν λογοκριμένα δεδομένα. Στο παρόνν κεφάλαιο θα παρουσιαστεί η αναλογιστική μέθοδος κατασκευής πινάκων επιβίωσης (*actuarial method*)

2.2 Πίνακες επιβίωσης

Έστω ένα τυχαίο δείγμα χρόνων ζωής μεγέθους *n* από ένα συγκεκριμένο πληθυσμό το οποίο περιέχει πλήρη και λογοκριμένα δεδομένα. Τα δεδομένα παρουσιάζονται σε ένα "πίνακα συχνοτήτων" με $k+1$ διαστήματα (τάξεις) της μορφής $I_j=[a_{j-1},a_j)$, $j=1,2,...,k+1$ όπου $a_0=0$, $a_k=M$ (*M* είναι μια προσδιορισμένη σταθερά) και $a_{k+1}=\infty$. Για το διάστημα I_j, $j=1,2,...,k+1$, έχουμε τις ακόλουθες πληροφορίες:

- d_j : αριθμός θανάτων (*failures*) στο διάστημα $I_j=[a_{j-1},a_j)$ (δηλαδή πλήρης χρόνοι ζωής),
- c_j : αριθμός διαφυγών (*lost, withdrawn*) στο διάστημα $I_j=[a_{j-1},a_j)$ (δηλαδή λογοκριμένοι χρόνοι ζωής),
- r_j : αριθμός ατόμων που βρίσκονται σε κίνδυνο (*at risk*) το χρόνο a_{j-1}.

Ο αριθμός r_j παριστάνει τον αριθμό των ατόμων που βρίσκονται σε ζωή το χρόνο a_{j-1}, δηλαδή ο αριθμός των ατόμων για τα οποία είμαστε σίγουροι ότι δεν έχουν πεθάνει σε χρόνο προγενέστερο της χρονικής στιγμής a_{j-1}. Συνεπώς, ο αριθμός r_j παριστάνει τον αριθμό των ατόμων των οποίων οι χρόνοι ζωής δύνανται να αποτύχουν ή να λογοκριθούν το χρόνο a_{j-1} ή μεταγενέστερο. Ο πραγματικός αριθμός των ατόμων που βρίσκονται σε ζωή το χρόνο a_{j-1} είναι άγνωστος και μπορεί να είναι μεγαλύτερος από τον αριθμό r_j γιατί στον πραγματικό αριθμό πιθανόν να συμπεριλαμβάνονται άτομα των οποίων οι χρόνοι ζωής έχουν λογοκριθεί στο παρελθόν πριν το χρόνο a_{j-1}. Από τα παραπάνω προκύπτει ότι

$$r_1 = n \quad \text{και} \quad r_j = r_{j-1} - d_{j-1} - c_{j-1} \quad \text{για} \quad j = 2,3,\ldots,k+1.$$

ΠΑΡΑΡΤΗΜΑΤΑ

Π1 Ακολουθιακοί έλεγχοι

Π2 …………………………………………………..

Π1 Ακολουθιακοί έλεγχοι

Έστω $X_1, X_2,... X_n$ ένα τυχαίο δείγμα από τη (συνεχή) κατανομή $f(x;\vartheta)$. Σύμφωνα με το Θεώρημα *Neyman-Pearson* για τον έλεγχο της υπόθεσης

$$H_0 : \vartheta = \vartheta_0 \quad — \quad H_1 : \vartheta = \vartheta_1$$

σε επίπεδο σημαντικότητας a έχουμε ότι η μηδενική υπόθεση απορρίπτεται όταν

$$\lambda(\mathbf{x}) = \frac{L(\mathbf{x};\vartheta_1)}{L(\mathbf{x};\vartheta_0)} \geq k$$

όπου

$$L(\mathbf{x};\vartheta_1) = \prod_{i=1}^{n} f(x_i;\vartheta_1), \quad L(\mathbf{x};\vartheta_0) = \prod_{i=1}^{n} f(x_i;\vartheta_0)$$

και η κρίσιμη τιμή k υπολογίζεται από τη σχέση $E_{\vartheta_0}[\lambda(\mathbf{X})] = a$. Οι “σταθερές” ποσότητες του άνω ελέγχου είναι το μέγεθος του δείγματος n και το σφάλμα τύπου I, a. Δοθέντος των δύο σταθερών ποσοτήτων η επιλογή της ισχυρότατης ελεγχοσυνάρτησης $\lambda(\mathbf{x})$ σύμφωνα με το Θεώρημα *Neyman-Pearson*, γίνεται με τέτοιο τρόπο έτσι ώστε να ελαχιστοποιηθεί το σφάλμα τύπου II, β.

Μια εναλλακτική προσέγγιση είναι να σταθεροποιήσουμε τα σφάλματα τύπου I και τύπου II (a και β) και να εργαστούμε με μεταβλητό μέγεθος δείγματος. Αν οι πληροφορίες που έχουμε από το δείγμα $X_1, X_2,... X_n$ (βήμα ελέγχου *n*) είναι επαρκείς για να αποφασίσουμε για την αποδοχή ή την μη αποδοχή της H_0 τότε η δειγματοληψία σταματά και λαμβάνεται η απόφαση. Αν οι πληροφορίες δεν είναι επαρκείς τότε μια επιπρόσθετη παρατήρηση λαμβάνεται και αποφασίζουμε για την αποδοχή ή την μη αποδοχή της H_0 με τη βοήθεια του δείγματος $X_1, X_2,... X_n, X_{n+1}$ (βήμα ελέγχου *n*+1)). Η διαδικασία αυτή επαναλαμβάνεται έως ότου μπορούμε να αποφασίσουμε για την αποδοχή ή την μη αποδοχή της H_0. Οι έλεγχοι αυτού του είδους ονομάζονται ακολουθιακοί έλεγχοι (*sequential tests*).

Ο ακολουθιακός έλεγχος λόγου πιθανοφάνειας (*sequential probability ratio test*) που πρότεινε ο Wald (1947) έχει ως ακολούθως: Για τον έλεγχο της υπόθεσης

$$H_0 : \vartheta = \vartheta_0 \quad — \quad H_1 : \vartheta = \vartheta_1$$

όπου τα σφάλματα a και β είναι προκαθορισμένα, σχηματίζουμε σε κάθε βήμα ελέγχου το λόγο των πιθανοφανειών

$$\lambda_n(\mathbf{x}) \equiv \lambda_n = \frac{L_n(\mathbf{x};\vartheta_1)}{L_n(\mathbf{x};\vartheta_0)} = \frac{\prod_{i=1}^{n} f(x_i;\vartheta_1)}{\prod_{i=1}^{n} f(x_i;\vartheta_0)}$$

και χρησιμοποιούμε δύο κρίσιμες τιμές A και B με $A<1<B$ με τις οποίες αποφασίζουμε για την αποδοχή ή την μη αποδοχή της H_0 σύμφωνα με το ακόλουθο πλαίσιο

Κατάσταση στο *n*-οστό βήμα ελέγχου	**Απόφαση**
$A < \lambda_n < B$	Η δειγματοληψία συνεχίζεται
$\lambda_n \le A$	Αποδοχή της H_0
$\lambda_n \ge B$	Αποδοχή της H_1

Οι κρίσιμες τιμές A και B μπορεί να αποδειχθεί ότι ικανοποιούν τις σχέσεις

$$A \ge \frac{\beta}{1-a}, \quad B \le \frac{1-\beta}{a}$$

και στην πράξη χρησιμοποιούνται οι τιμές

$$A = \frac{\beta}{1-a}, \quad B = \frac{1-\beta}{a}$$

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Παπαϊωάννου, Τ. και Φερεντίνος, Κ. (1999). *Ιατρική Στατιστική και Στοιχεία Βιομαθηματικών*, Εκδόσεις Πανεπιστημίου Ιωαννίνων, Ιωάννινα.

Ξένη

Andersen, P. K., Borgan, ∅., Gill, R. D. and Keiding, N. (1993). *Statistical models based on counting processes*, Springer-Verlag, New York.

Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model, *Journal of the American Statistical Association*, **83**, 203-212.

Epstein, B. (1958). The exponential distribution and its role in life testing, *Industrial Quality Control*, **15**, 2-7.

Epstein, B. and Sobel, M. (1954). Some theorems relevant to life testing from an exponential distribution, *Annals of Mathematical Statistics*, **25**, 373-381.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Discrete Univariate Distributions*, 2nd ed., John Wiley, New York.

Mahibbur, M. and Covindarajulu, Z. (1997). Nonparametric estimation of the ratio of variance components, in *Advances in Combinatorial Methods and Applications to Probability and Statistics*, Edited by Balakrishnan, N., Birkhauser, Boston, 363-384.

Wald, A. (1947). *Sequential analysis*, John Wiley, New York.