ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΡΑ ΑΞΙΟΠΙΣΤΙΑΣ ΣΤΑΘΜΙΣΜΕΝΩΝ ΚΑΤΑΝΟΜΩΝ

Σταυρούλα Ν. Μπακάλη

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
Σεπτέμβριος 2011

ΑΦΙΕΡΩΣΗ

Στους γονείς μου

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις θερμές και ειλικρινείς μου ευχαριστίες στην επιβλέπουσα καθηγήτρια κα Κατερίνα Δημάκη για την άριστη συνεργασία και την πολύτιμη βοήθειά της κατά τη διάρκεια συγγραφής αυτής της εργασίας. Επίσης, ευχαριστώ τους αγαπημένους μου συμφοιτητές Ελευθερία, Μαργαρίτα, Ζαχαρία, Βαγγέλη, Δημήτρη και Νίκο για τις όμορφες στιγμές που ζήσαμε και συνεχίζουμε να ζούμε. Τέλος, ένα μεγάλο ευχαριστώ στους γονείς μου για την ηθική και υλική στήριξή τους όλα αυτά τα χρόνια.

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

Γεννήθηκα πριν από 29 περίπου χρόνια στην Αθήνα. Το 2001 και μετά από επίμονη προσπάθεια κατάφερα να εισαχθώ στο τμήμα Μαθηματικών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών από το οποίο πήρα το πτυχίο μου το καλοκαίρι του 2007. Παράλληλα με τις σπουδές μου εργάστηκα σαν μαθηματικός παραδίδοντας ιδιαίτερα μαθήματα σε μαθητές Γυμνασίου και Λυκείου. Τον Οκτώβρη του ίδιου έτους ξεκίνησε η φοίτησή μου στο ΠΜΣ Ειδίκευσης στη Στατιστική με κατεύθυνση «Στατιστικές Μέθοδοι στην Ιατρική και Φαρμακευτική» η οποία ολοκληρώνεται με την υποβολή της παρούσας εργασίας.

ABSTRACT

Stavroula Bakali

RELIABILITY MEASURES OF WEIGHTED DISTRIBUTIONS

September 2011

In life length studies the observations generated from a stochastic process cannot always be considered as a random sample from the original distribution. In many situations, the observations are recorded according to some weight function and then the resulting distribution is called weighted distribution.

The aim of this study is to present the basic properties of weighted distributions as well as a variety of relationships between the reliability measures of the original distribution and associated weighted distribution.

ΠΕΡΙΛΗΨΗ

Σταυρούλα Μπακάλη

ΜΕΤΡΑ ΑΞΙΟΠΙΣΤΙΑΣ ΣΤΑΘΜΙΣΜΕΝΩΝ ΚΑΤΑΝΟΜΩΝ

Σεπτέμβριος 2011

Στις έρευνες διάρκειας ζωής οι παρατηρήσεις που παράγονται από μια στοχαστική διαδικασία δεν αποτελούν πάντα τυχαίο δείγμα της αρχικής κατανομής. Σε πολλές περιπτώσεις, οι παρατηρήσεις καταγράφονται επηρεασμένες από κάποια συνάρτηση στάθμισης και τότε η κατανομή που προκύπτει ονομάζεται σταθμισμένη ή μεροληπτική κατανομή.

Σκοπός της εργασίας αυτής είναι η παρουσίαση των βασικών ιδιοτήτων των σταθμισμένων κατανομών καθώς επίσης και των σχέσεων που αφορούν στα μέτρα αξιοπιστίας μεταξύ αρχικής και σταθμισμένης κατανομής.

ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΚΕΦΑΛΑΙΟ 1
ΕΙΣΑΓΩΓΗ

Κύριο αντικείμενο της ανάλυσης επιβίωσης είναι οι έρευνες διάρκειας ζωής. Συχνά, σε τέτοιου είδους έρευνες, οι παρατηρήσεις που παράγονται από μια στοχαστική διαδικασία καταγράφονται επηρεασμένες από κάποια συνάρτηση στάθμισης και τότε η κατανομή που προκύπτει ονομάζεται σταθμισμένη ή μεροληπτική κατανομή.

Παρόλο που η ιδέα των σταθμισμένων κατανομών ανήκει στον Fisher (1934), ο Rao (1965) ήταν αυτός που μελέτησε πρώτος τις σταθμισμένες κατανομές επισημαίνοντας ότι υπάρχουν περιπτώσεις (μη παρατηρησιμότητα γεγονότων, μερική καταστροφή των παρατηρήσεων, δειγματοληψία με άνισες πιθανότητες) όπου οι καταγεγραμμένες παρατηρήσεις δε μπορούν να θεωρηθούν σαν τυχαίο δείγμα της αρχικής κατανομής. Υπό αυτές τις συνθήκες λοιπόν, οι σταθμισμένες κατανομές αποτελούν ένα χρήσιμο εργαλείο στατιστικού σχεδιασμού ειδικά όταν το δείγμα δεν πληροί τις προϋποθέσεις της συνήθους στατιστικής μεθοδολογίας.

Οι σταθμισμένες κατανομές βρίσκουν ευρεία εφαρμογή σε πολλούς τομείς. Η παρούσα εργασία μελετά τις σταθμισμένες κατανομές υπό το πρίσμα της θεωρίας αξιοπιστίας.

Έτσι, στο δεύτερο κεφάλαιο γίνεται μια εισαγωγή στην ανάλυση επιβίωσης. Εισάγεται η έννοια της λογοκρισίας, ενώ παρουσιάζονται οι συναρτήσεις χρόνου επιβίωσης καθώς και οι σχέσεις που τις συνδέουν.

Στο τρίτο κεφάλαιο γίνεται εκτενής παρουσίαση των σημαντικότερων κατανομών χρόνου ζωής. Δίνονται οι βασικές τους ιδιότητες καθώς και τα μοντέλα που οδηγούν σε κάθε μια από αυτές. Στην πλειοψηφία τους είναι συνεχείς κατανομές μιας και ο χρόνος επιβίωσης θεωρείται συνεχής μεταβλητή και σπανιότερα διακριτή.

Στο τέταρτο κεφάλαιο γίνεται η εισαγωγή στις σταθμισμένες κατανομές. Παρουσιάζονται οι βασικές τους ιδιότητες στην ειδική αλλά και στη γενική περίπτωση, ενώ ιδιαίτερη έμφαση δίνεται στην ιδιότητα του τύπου αμεταβλητότητας υπό μεροληπτική λόγω μεγέθους δειγματοληψία καθώς εξετάζεται αν και ποιες από τις κατανομές που αναπτύχθηκαν στο προηγούμενο κεφάλαιο χαρακτηρίζονται από την ιδιότητα αυτή.

Τέλος, το πέμπτο κεφάλαιο αφιερώνεται στη θεμελίωση των σχέσεων που αφορούν στα μέτρα αξιοπιστίας μεταξύ αρχικής και σταθμισμένης κατανομής, τόσο στη συνεχή όσο και στη διακριτή περίπτωση, ενώ ταυτόχρονα παρουσιάζονται ενδιαφέρουσες ανισοτικές σχέσεις καθώς και ενδιαφέροντα αποτελέσματα σε σχέση με τα κριτήρια γήρανσης.

ΚΕΦΑΛΑΙΟ 2
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ

Η *ανάλυση επιβίωσης (survival analysis)* είναι μια περιοχή έρευνας στη Στατιστική η οποία δημιουργήθηκε για την ανάλυση δεδομένων που αφορούν στο χρόνο μέχρι να γίνει ένα συγκεκριμένο γεγονός. Ο *χρόνος επιβίωσης* ή *χρόνος αποτυχίας (survival time ή failure time)* είναι μια μεταβλητή η οποία μετράει το χρόνο που μεσολαβεί από τη στιγμή έναρξης της παρακολούθησης ενός ατόμου μέχρι τη στιγμή που το άτομο θα αντιμετωπίσει ένα ενδεχόμενο. Το ενδεχόμενο αυτό, ανεξαρτήτως της θετικής ή αρνητικής του σημασίας, ονομάζεται «αποτυχία». Αρχικά, η ανάλυση αναφερόταν στο χρόνο από την έναρξη μιας θεραπείας μέχρι το θάνατο και για αυτό το λόγο πήρε το συγκεκριμένο όνομα. Ωστόσο, στις ιατρικές μελέτες όπου διεξάγεται ανάλυση επιβίωσης, η μελετώμενη έκβαση δεν είναι απαραίτητα ο θάνατος, αλλά η ίαση, η εμφάνιση μιας λοίμωξης, η διασπορά ενός καρκίνου, η αντίδραση ενός οργανισμού σε ένα φάρμακο κ.α.

Η ανάλυση επιβίωσης μπορεί να εφαρμοστεί σε αρκετούς τομείς πέραν της ιατρικής όπως στη μηχανολογία πχ. για την ανάλυση του χρόνου μέχρι την καταστροφή ενός μηχανήματος ή εξαρτήματος, στις οικονομικές επιστήμες πχ. για την ανάλυση του χρόνου μέχρι ένας δείκτης να ξεπεράσει ένα όριο, στις κοινωνικές επιστήμες πχ. για την ανάλυση του χρόνου εύρεσης απασχόλησης κ.τ.λ. Στην περίπτωση της μηχανολογίας, η ανάλυση αναφέρεται και ως *θεωρία αξιοπιστίας (reliability theory).*

Παρά το γεγονός ότι τα δεδομένα είναι ποσοτικά, οι κατανομές των χρόνων επιβίωσης σε ελάχιστες μόνο περιπτώσεις ακολουθούν την κανονική κατανομή τείνοντας να έχουν δεξιά ασυμμετρία.

2.1 Λογοκριμένα Δεδομένα

Το ιδιαίτερο χαρακτηριστικό που εμφανίζεται στα δεδομένα επιβίωσης και διαφοροποιεί σημαντικά την ανάλυση τους είναι ότι οι χρόνοι επιβίωσης κάποιων παρατηρήσεων δεν είναι ακριβείς. *Λογοκριμένα* ή *περικομμένα δεδομένα (censored data)* είναι τα δεδομένα για τα οποία δεν είναι γνωστός ο χρόνος που συμβαίνει το υπό μελέτη γεγονός. Λογοκριμένες παρατηρήσεις μπορούν να προκύψουν είτε λόγω λήξης της έρευνας, είτε γιατί χάθηκε η επικοινωνία με κάποιους από τους ασθενείς, είτε γιατί λόγω κακών επιδράσεων είναι απαραίτητο κάποιοι ασθενείς να σταματήσουν την έρευνα, είτε λόγω άρνησής τους να συμμετάσχουν έως την ολοκλήρωση της έρευνας.

2.1.1 Είδη Λογοκρισίας

Το πιο συνηθισμένο είδος λογοκρισίας είναι η *δεξιά λογοκρισία (right censoring)*. Στην περίπτωση αυτή, ο χρόνος επιβίωσης είναι μεγαλύτερος από την τιμή που έχει καταγραφεί. Η δεξιά λογοκρισία χωρίζεται σε 3 κατηγορίες: τη *λογοκρισία τύπου Ι*, τη *λογοκρισία τύπου ΙΙ* και τη *λογοκρισία τύπου ΙΙΙ*.

- *Λογοκρισία τύπου Ι*

  Η χρονική διάρκεια της έρευνας καθώς και ο αριθμός των συμμετεχόντων καθορίζονται από την αρχή. Έτσι, οι χρόνοι επιβίωσης των ασθενών που «απέτυχαν» κατά τη διάρκεια της έρευνας αποτελούν τις ακριβείς παρατηρήσεις, ενώ οι χρόνοι επιβίωσης των ασθενών που δεν «απέτυχαν» κατά τη διάρκεια της έρευνας, εάν δεν υπάρχουν απώλειες από ατυχήματα, θεωρούνται ίσοι με τη χρονική διάρκεια της έρευνας και αποτελούν τις λογοκριμένες παρατηρήσεις.

- *Λογοκρισία τύπου ΙΙ*

  Σε αυτό τον τύπο λογοκρισίας, η έρευνα συνεχίζεται μέχρι να «αποτύχει» ένας συγκεκριμένος αριθμός ασθενών. Έτσι, οι χρόνοι επιβίωσης του συγκεκριμένου αριθμού ασθενών αποτελούν τις ακριβείς παρατηρήσεις, ενώ οι χρόνοι επιβίωσης των υπολοίπων ασθενών, εάν δεν υπάρχουν απώλειες από ατυχήματα, θεωρούνται ίσοι με τη μεγαλύτερη μη λογοκριμένη παρατήρηση και αποτελούν τις λογοκριμένες παρατηρήσεις.

- *Λογοκρισία τύπου ΙΙΙ*

  Αυτός ο τύπος λογοκρισίας συναντάται κυρίως σε κλινικές μελέτες όπου ενώ οι χρονικές στιγμές έναρξης και λήξης της έρευνας είναι προκαθορισμένες, οι ασθενείς εισέρχονται σε αυτή σε τυχαίες χρονικές στιγμές. Οι ακριβείς παρατηρήσεις είναι οι χρόνοι επιβίωσης των ασθενών που «απέτυχαν» κατά τη διάρκεια της έρευνας. Σε αυτή την περίπτωση όμως, οι χρόνοι επιβίωσης των λογοκριμένων παρατηρήσεων είναι διαφορετικοί, αφού για κάθε ασθενή ο χρόνος επιβίωσης ξεκινά από τη στιγμή που εισέρχεται στη μελέτη και διαρκεί μέχρι την απομάκρυνσή του από αυτή ή μέχρι το τέλος της έρευνας.

Εκτός από τη δεξιά λογοκρισία, υπάρχουν δύο ακόμα είδη λογοκρισίας: η *αριστερή λογοκρισία (left censoring)* στην οποία ο χρόνος επιβίωσης είναι μικρότερος από την τιμή που έχει καταγραφεί και η *λογοκρισία διαστήματος (interval censoring)* στην οποία ο χρόνος επιβίωσης βρίσκεται σε ένα διάστημα (T_1, T_2).

2.2 Συναρτήσεις Χρόνου Επιβίωσης

Έστω X μια μη αρνητική τυχαία μεταβλητή η οποία αναφέρεται στο χρόνο επιβίωσης της συνιστώσας ενός συστήματος. Εδώ πρέπει να αναφέρουμε ότι ο όρος «συνιστώσα» χρησιμοποιείται για να υποδηλώσει το άτομο, το εξάρτημα, τον οικονομικό δείκτη κ.τ.λ. Η τυχαία μεταβλητή X μπορεί να είναι είτε συνεχής είτε διακριτή. Η πιθανή συμπεριφορά επιβίωσης της συνιστώσας προσδιορίζεται από μια συνάρτηση $f_X(t)$ που στην ουσία καθορίζει την πιθανότητα στιγμιαίας «αποτυχίας» τη χρονική στιγμή t. Επομένως, στην περίπτωση της συνεχούς τυχαίας μεταβλητής η *συνάρτηση πυκνότητας πιθανότητας (probability density function)* ή *πυκνότητα αποτυχίας (failure density)* δίνεται από τον τύπο

$$f_X(t) = \lim_{\Delta t \to 0} \frac{P(t < X < t + \Delta t)}{\Delta t}, \qquad t \geq 0$$

ενώ στην περίπτωση της διακριτής τυχαίας μεταβλητής η *συνάρτηση πιθανότητας (probability function)* δίνεται από τον τύπο

$$f_X(t) = P(X = t), \qquad t = 0, 1, 2, \dots$$

Η συνάρτηση (πυκνότητας) πιθανότητας έχει τις ακόλουθες δύο ιδιότητες:

1. Είναι μη αρνητική συνάρτηση:

   $f_X(t) \geq 0$ για κάθε $t \geq 0$ ή $t = 0, 1, 2, \dots$

   $= 0$ για $t < 0$

2. Στην περίπτωση της συνεχούς τυχαίας μεταβλητής το εμβαδό μεταξύ της καμπύλης και του άξονα των t ισούται με 1, ενώ στην περίπτωση της διακριτής τυχαίας μεταβλητής το άθροισμα των επιμέρους πιθανοτήτων για $t = 0, 1, 2, \ldots$ είναι 1.

Αντίστοιχη προς τη συνάρτηση $f_X(t)$ είναι η *αθροιστική συνάρτηση κατανομής αποτυχίας (probability of failure)* $F_X(t)$ η οποία καθορίζει την πιθανότητα «αποτυχίας» μέχρι τη χρονική στιγμή t. Επομένως, στην περίπτωση της συνεχούς τυχαίας μεταβλητής

$$F_X(t) = P(X \leq t) = \int_0^t f_X(x)dx, \qquad t \geq 0$$

ενώ στην περίπτωση της διακριτής τυχαίας μεταβλητής

$$F_X(t) = P(X \leq t) = \sum_{x \leq t} f_X(x), \qquad t = 0, 1, 2, \ldots$$

Ωστόσο, στις μελέτες όπου διεξάγεται ανάλυση επιβίωσης, εκείνο που κυρίως μας ενδιαφέρει είναι η πιθανότητα η συνιστώσα του συστήματος να μην έχει «αποτύχει» μέχρι τη χρονική στιγμή t. Έτσι, η κατανομή των χρόνων επιβίωσης περιγράφεται συνήθως από τρεις συναρτήσεις: τη *συνάρτηση επιβίωσης*, τη *συνάρτηση κινδύνου* και τη *συνάρτηση μέσης υπολειπόμενης ζωής*.

2.2.1 Συνάρτηση Επιβίωσης

Ορισμός 1 Έστω X μια μη αρνητική τυχαία μεταβλητή η οποία εκφράζει το χρόνο επιβίωσης της συνιστώσας ενός συστήματος. Η *συνάρτηση επιβίωσης (survival function)* του χρόνου επιβίωσης X συμβολίζεται με $S(t)$ ή $\bar{F}_X(t)$ και ορίζεται ως η πιθανότητα η συνιστώσα να επιβιώσει για χρόνο μεγαλύτερο του t.

$$S(t) = P(X > t) \qquad (2.1)$$

Από τον ορισμό της αθροιστικής συνάρτησης κατανομής, επειδή $F_X(t) = P(X \leq t)$, έχουμε ότι

$$S(t) = 1 - F_X(t) \qquad (2.2)$$

Η συνάρτηση επιβίωσης είναι μια φθίνουσα συνάρτηση του t και έχει τις ακόλουθες ιδιότητες:

1. $S(t) = 1$ για $t = 0$
2. $S(t) = 0$ για $t = \infty$

Δηλαδή η πιθανότητα η συνιστώσα να επιβιώσει τουλάχιστον στο χρόνο 0 είναι 1, ενώ η πιθανότητα η συνιστώσα να επιβιώσει σε άπειρο χρόνο είναι 0.

Στην πράξη, όταν δεν υπάρχουν λογοκριμένες παρατηρήσεις, η συνάρτηση επιβίωσης εκτιμάται ως ο λόγος των συνιστωσών που επιβιώνουν για χρόνο μεγαλύτερο του t προς τον συνολικό αριθμό των συνιστωσών:

$$\hat{S}(t) = \frac{\text{αρ. συνιστωσών που επιβιώνουν για χρόνο μεγαλύτερο του } t}{\text{συνολικός αριθμός συνιστωσών}}$$

Όταν στα δεδομένα υπάρχουν λογοκριμένες παρατηρήσεις, τότε η παραπάνω σχέση δε μπορεί να εφαρμοστεί και καταφεύγουμε σε μη παραμετρικές μεθόδους για την εκτίμηση του $S(t)$ όπως είναι η μέθοδος γινομένου ορίου ή αλλιώς μέθοδος Kaplan-Meier.

2.2.2 Συνάρτηση Κινδύνου

Ορισμός 2 Έστω X μια μη αρνητική τυχαία μεταβλητή η οποία εκφράζει το χρόνο επιβίωσης της συνιστώσας ενός συστήματος. Η *συνάρτηση κινδύνου (hazard function)* του χρόνου επιβίωσης X συμβολίζεται με $h_X(t)$ και ορίζεται ως το όριο της πιθανότητας η συνιστώσα να «αποτύχει» στο διάστημα $(t, t + \Delta t)$ δοθέντος ότι επιβίωσε μέχρι τη χρονική στιγμή t ανά μονάδα πλάτους Δt. Μπορεί να εκφραστεί ως εξής:

$$h_X(t) = \lim_{\Delta t \to 0} \frac{P(t < X \leq t + \Delta t | X > t)}{\Delta t}$$

ή

$$h_X(t) = \frac{1}{\bar{F}_X(t)} \lim_{\Delta t \to 0} \frac{P(t < X \leq t + \Delta t)}{\Delta t}$$

Επομένως, η συνάρτηση κινδύνου για την περίπτωση της συνεχούς τυχαίας μεταβλητής δίνεται από τη σχέση

$$h_X(t) = \frac{f_X(t)}{\bar{F}_X(t)}, \qquad t \geq 0 \tag{2.3}$$

ενώ για την περίπτωση της διακριτής τυχαίας μεταβλητής από τη σχέση

$$h_X(t) = \frac{P(X = t)}{P(X \geq t)}, \qquad t = 0, 1, 2, \ldots \tag{2.4}$$

Ορισμός 3 (Barlow & Proschan (1975)) Μια τυχαία μεταβλητή διάρκειας ζωής $X \geq 0$ λέμε ότι έχει *αύξοντα ρυθμό αποτυχίας (increasing failure rate (IFR))* ή *φθίνοντα ρυθμό αποτυχίας (decreasing failure rate (DFR))* εάν η $h_X(t)$ είναι αύξουσα συνάρτηση του t ή φθίνουσα συνάρτηση του t αντιστοίχως.

Γενικότερα, η συνάρτηση κινδύνου μπορεί να αυξάνεται, να μειώνεται, να μένει σταθερή ή να δηλώνει μια πιο περίπλοκη διαδικασία. Παρόλα αυτά, είναι εξαιρετικά χρήσιμη για τη μελέτη του χρόνου επιβίωσης δεδομένου ότι στις περισσότερες περιπτώσεις υπάρχουν πληροφορίες σχετικά με τη διαχρονική εξέλιξή της.

2.2.3 Συνάρτηση Μέσης Υπολειπόμενης Ζωής

Ορισμός 4 Έστω X μια μη αρνητική τυχαία μεταβλητή που εκφράζει το χρόνο επιβίωσης της συνιστώσας ενός συστήματος. Η *συνάρτηση μέσης υπολειπόμενης ζωής (mean residual life function)* κατά τη χρονική στιγμή t συμβολίζεται με $\mu^X(t)$ και ορίζεται ως ο αναμενόμενος χρόνος ζωής της συνιστώσας δεδομένου ότι επιβίωσε μέχρι τη χρονική στιγμή t.

$$\mu^X(t) = E(X - t | X > t) \tag{2.5}$$

Προφανώς, η $\mu^X(0)$ εκφράζει την αναμενόμενη ζωή μιας καινούριας συνιστώσας, ενώ η $\mu^X(t)$ εκφράζει την αναμενόμενη ζωή που υπολείπεται σε μια συνιστώσα ηλικίας t. Με τον όρο «ηλικία t» εννοούμε ότι η συνιστώσα έζησε ως τη χρονική στιγμή t και εξακολουθεί να ζει στο χρόνο t.

Ορισμός 5 (Bryson & Siddiqui (1969)) Μια τυχαία μεταβλητή διάρκειας ζωής $X \geq 0$ λέμε ότι έχει *αύξουσα μέση υπολειπόμενη ζωή (increasing mean residual life (IMRL))* ή *φθίνουσα μέση υπολειπόμενη ζωή (decreasing mean residual life (DMRL))* εάν η $\mu^X(t)$ είναι αύξουσα συνάρτηση του t ή φθίνουσα συνάρτηση του t αντιστοίχως.

Οι Bryson & Siddiqui (1969) έδειξαν ότι IFR $\Rightarrow$ DMRL και DFR $\Rightarrow$ IMRL δίνοντας ταυτόχρονα αντιπαράδειγμα για να δείξουν ότι το αντίστροφο δεν ισχύει.

Η συνάρτηση μέσης υπολειπόμενης ζωής δεν έχει χρησιμοποιηθεί αρκετά σε εφαρμογές. Παρόλα αυτά, είναι καταλληλότερος δείκτης φθοράς από τη συνάρτηση κινδύνου.

2.3 Σχέσεις μεταξύ των Συναρτήσεων Χρόνου Επιβίωσης

Οι τρεις συναρτήσεις $\bar{F}_X(t)$, $h_X(t)$ και $\mu^X(t)$ είναι μαθηματικά ισοδύναμες με την έννοια ότι αν δοθεί μία από αυτές, τότε μπορούμε να βρούμε τις άλλες δύο. Με τα τέσσερα θεωρήματα που ακολουθούν δίνονται οι σχέσεις που τις συνδέουν τόσο στη συνεχή όσο και στη διακριτή περίπτωση.

Θεώρημα 1 Έστω X μια μη αρνητική τυχαία μεταβλητή η οποία έχει συνάρτηση επιβίωσης $\bar{F}_X(t)$. Η κατανομή της X ορίζεται μονοσήμαντα από τη συνάρτηση κινδύνου $h_X(t)$.

Συγκεκριμένα, για την περίπτωση της συνεχούς τυχαίας μεταβλητής

$$\bar{F}_X(t) = exp\left[-\int_0^t h_X(x)dx\right], \qquad t \geq 0 \tag{2.6}$$

ενώ για την περίπτωση της διακριτής τυχαίας μεταβλητής

$$\bar{F}_X(t) = \prod_{i=0}^{t-1}[1 - h_X(i)], \qquad t = 0, 1, 2, \dots \tag{2.7}$$

Θεώρημα 2 Έστω X μια μη αρνητική τυχαία μεταβλητή η οποία έχει συνάρτηση επιβίωσης $\bar{F}_X(t)$. Αποδεικνύεται ότι για την περίπτωση της συνεχούς τυχαίας μεταβλητής

$$\mu^X(t) = \frac{1}{\bar{F}_X(t)}\int_t^\infty \bar{F}_X(x)dx\,, \qquad t \geq 0 \tag{2.8}$$

ενώ για την περίπτωση της διακριτής τυχαίας μεταβλητής

$$\mu^X(t) = \frac{1}{\bar{F}_X(t)}\sum_{x=t}^{\infty} \bar{F}_X(x), \qquad t = 0, 1, 2, \dots \tag{2.9}$$

Θεώρημα 3 Έστω X μια μη αρνητική τυχαία μεταβλητή η οποία έχει συνάρτηση επιβίωσης $\bar{F}_X(t)$. Η κατανομή της X ορίζεται μονοσήμαντα από τη συνάρτηση μέσης υπολειπόμενης ζωής $\mu^X(t)$.

Συγκεκριμένα, για την περίπτωση της συνεχούς τυχαίας μεταβλητής

$$\bar{F}_X(t) = \frac{\mu^X(0)}{\mu^X(t)} exp\left[-\int_0^t \frac{dx}{\mu^X(x)}\right], \qquad t \geq 0 \tag{2.10}$$

ενώ για την περίπτωση της διακριτής τυχαίας μεταβλητής

$$\bar{F}_X(t) = \prod_{i=0}^{t-1} \frac{\mu^X(i) - 1}{\mu^X(i+1)}, \qquad t = 0, 1, 2, \dots \tag{2.11}$$

Θεώρημα 4 Έστω X μια μη αρνητική τυχαία μεταβλητή η οποία έχει συνάρτηση επιβίωσης $\bar{F}_X(t)$. Αποδεικνύεται ότι:

για την περίπτωση της συνεχούς τυχαίας μεταβλητής

$$h_X(t) = \frac{1 + \mu'^X(t)}{\mu^X(t)}, \qquad t \geq 0 \tag{2.12}$$

ενώ για την περίπτωση της διακριτής τυχαίας μεταβλητής

$$h_X(t) = \frac{\mu^X(t) - \mu^X(t-1) + 1}{\mu^X(t)}, \qquad t = 0, 1, 2, \dots \tag{2.13}$$

ΚΕΦΑΛΑΙΟ 3
ΚΑΤΑΝΟΜΕΣ ΧΡΟΝΟΥ ΖΩΗΣ

Η κατανομή «αποτυχίας» περιγράφει μαθηματικά τη διάρκεια ζωής ενός ατόμου, ενός εξαρτήματος, ενός οργανισμού κ.τ.λ. Κατά συνέπεια, η επιλογή της κατάλληλης κατανομής παίζει σημαντικό ρόλο στην ανάλυση επιβίωσης. Σε αυτό το κεφάλαιο παρουσιάζονται οι σημαντικότερες κατανομές χρόνου ζωής καθώς και οι ιδιότητές τους.

3.1 Εκθετική Κατανομή

Η πιο απλή αλλά και πιο σημαντική κατανομή στις μελέτες επιβίωσης είναι η εκθετική κατανομή. Στα τέλη της δεκαετίας του 1940, οι ερευνητές άρχισαν να επιλέγουν την εκθετική κατανομή για να περιγράψουν το πρότυπο ζωής των ηλεκτρονικών συστημάτων. Η εκθετική κατανομή παίζει τόσο σημαντικό ρόλο στις έρευνες διάρκειας ζωής όσο η κανονική κατανομή σε άλλες περιοχές έρευνας της Στατιστικής. Είναι γνωστή για την ιδιότητα «έλλειψης μνήμης» και παρά το γεγονός ότι αρκετά δεδομένα επιβίωσης δε μπορούν να περιγραφούν επαρκώς από την εκθετική κατανομή, η μελέτη της είναι απαραίτητη καθώς διευκολύνει το χειρισμό γενικότερων καταστάσεων.

Ορισμός 6 Μια συνεχής τυχαία μεταβλητή X με συνάρτηση επιβίωσης

$$\bar{F}_X(t) = e^{-\lambda t}, \qquad t \geq 0, \qquad \lambda > 0 \tag{3.1}$$

ακολουθεί την *εκθετική κατανομή* με παράμετρο $\lambda > 0$.
Συμβολικά, $X \sim exp\ (\lambda)$.

Η αθροιστική συνάρτηση κατανομής αποτυχίας της X δίνεται από τη σχέση

$$F_X(t) = 1 - \bar{F}_X(t) = 1 - e^{-\lambda t}, \qquad t \geq 0, \qquad \lambda > 0 \tag{3.2}$$

ενώ η συνάρτηση πυκνότητας πιθανότητας από τη σχέση

$$f_X(t) = -\frac{d}{dt}\bar{F}_X(t) = \lambda e^{-\lambda t}, \qquad t \geq 0, \qquad \lambda > 0 \tag{3.3}$$

Βασικές Ιδιότητες

Έστω $X \sim exp(\lambda)$, $t \geq 0$, $\lambda > 0$. Τότε

1. $E(X) = {}^{1}/_{\lambda}$
2. $V(X) = {}^{1}/_{\lambda^2}$

Πρόταση 1 Έστω X συνεχής τυχαία μεταβλητή που ορίζεται στο $[0, +\infty)$. Η X ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda > 0$ αν και μόνο αν η συνάρτηση κινδύνου κατά τη χρονική στιγμή t είναι σταθερή και ίση με λ.

Απόδειξη Με βάση το θεώρημα 1 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση κινδύνου. Επομένως, αρκεί να υπολογίσουμε τη συνάρτηση κινδύνου που στην περίπτωση της εκθετικής κατανομής είναι:

$$h_X(t) = \frac{f_X(t)}{\bar{F}_X(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

Πρόταση 2 Έστω X συνεχής τυχαία μεταβλητή που ορίζεται στο $[0, +\infty)$. Η X ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda > 0$ αν και μόνο αν η μέση υπολειπόμενη ζωή κατά τη χρονική στιγμή t είναι σταθερή και ίση με ${}^{1}/_{\lambda}$.

Απόδειξη Με βάση το θεώρημα 3 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση μέσης υπολειπόμενης ζωής. Επομένως, αρκεί να υπολογίσουμε τη μέση υπολειπόμενη ζωή που στην περίπτωση της εκθετικής κατανομής είναι:

$$\mu^X(t) = \frac{1}{\bar{F}_X(t)} \int_t^\infty \bar{F}_X(x) dx = \frac{\frac{1}{\lambda} e^{-\lambda t}}{e^{-\lambda t}} = \frac{1}{\lambda}$$

Από τις δύο προηγούμενες προτάσεις παρατηρούμε ότι αν X είναι μια συνεχής τυχαία μεταβλητή που περιγράφει το χρόνο ζωής της συνιστώσας ενός συστήματος και $X \sim exp(\lambda)$, $\lambda > 0$, τότε το γινόμενο της συνάρτησης κινδύνου επί τη μέση υπολειπόμενη ζωή είναι σταθερό και ίσο με τη μονάδα.

Μοντέλα που οδηγούν στην εκθετική κατανομή

1. Ο χρόνος αναμονής έως ότου πραγματοποιηθεί το πρώτο γεγονός σε μια διαδικασία Poisson με παράμετρο λ ακολουθεί την εκθετική κατανομή.
2. Σταθερή συνάρτηση κινδύνου οδηγεί σε εκθετική κατανομή. Αν $h_X(t) = \lambda,\ \lambda > 0$, τότε $X \sim exp\ (\lambda)$.
3. Σταθερή μέση υπολειπόμενη ζωή οδηγεί σε εκθετική κατανομή. Αν $\mu^X(t) = {}^1/_\lambda$, τότε $X \sim exp\ (\lambda)$.
4. Έλλειψη αθροιστικής μνήμης της εκθετικής κατανομής. Αν η τυχαία μεταβλητή X ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda > 0$, τότε η δεσμευμένη πιθανότητα ο χρόνος επιβίωσης X να υπερβεί το $t + y$ δοθέντος ότι έχει υπερβεί το t ισούται με την πιθανότητα το X να υπερβεί το y. Με άλλα λόγια, ο χρόνος που έχει ήδη διαρρεύσει δεν επηρεάζει το χρόνο που απαιτείται για την πραγματοποίηση ενός γεγονότος σε μια διαδικασία Poisson.

3.2 Κατανομή Weibull

Η κατανομή Weibull αποτελεί γενίκευση της εκθετικής κατανομής. Ωστόσο, αντίθετα από την εκθετική κατανομή, η συνάρτηση κινδύνου της Weibull δεν είναι σταθερά κι έτσι η εφαρμογή της είναι ευρύτερη. Η κατανομή προτάθηκε από τον Weibull (1939) και η εφαρμογή της σε διάφορες καταστάσεις «αποτυχίας» συζητήθηκε πάλι από τον Weibull (1951).

Ορισμός 7 Μια συνεχής τυχαία μεταβλητή X με συνάρτηση επιβίωσης

$$\bar{F}_X(t) = exp[-(\lambda t)^p], \qquad t \geq 0, \qquad \lambda, p > 0 \tag{3.4}$$

ακολουθεί την *κατανομή Weibull* με παραμέτρους $\lambda, p > 0$.
Συμβολικά, $X \sim Weibull(\lambda, p)$.

Η αθροιστική συνάρτηση κατανομής της X δίνεται από τη σχέση

$$F_X(t) = 1 - \bar{F}_X(t) = 1 - exp[-(\lambda t)^p], \qquad t \geq 0, \qquad \lambda, p > 0 \tag{3.5}$$

ενώ η συνάρτηση πυκνότητας πιθανότητας από τη σχέση

$$f_X(t) = -\frac{d}{dt}\bar{F}_X(t) = p\lambda^p t^{p-1}\, exp[-(\lambda t)^p], \qquad t \geq 0, \qquad \lambda, p > 0 \tag{3.6}$$

Η κατανομή Weibull αν και είναι δημοφιλής παρουσιάζει κάποιες τεχνικές δυσκολίες στη μελέτη της. Οι δυσκολίες αυτές αντιμετωπίζονται ικανοποιητικά εξαιτίας της συναρτησιακής σχέσης που υπάρχει μεταξύ των τυχαίων μεταβλητών που ακολουθούν την εκθετική κατανομή και την κατανομή Weibull.

Πρόταση 3 Αν $X \sim exp\,(\theta)$, τότε $Y = X^{1/p} \sim Weibull(\lambda = \theta^{1/p}, p)$.

Βασικές Ιδιότητες

Έστω $X \sim Weibull(\lambda, p),\ t \geq 0,\ \lambda, p > 0$. Τότε

1. $E(X) = \lambda^{-1} \Gamma\left(\frac{p+1}{p}\right)$
2. $V(X) = \lambda^{-2}\left[\Gamma\left(\frac{p+1}{p}\right) - \Gamma^2\left(\frac{p+1}{p}\right)\right]$

Πρόταση 4 Έστω X συνεχής τυχαία μεταβλητή που ορίζεται στο $[0, +\infty)$**.** Η X ακολουθεί την κατανομή Weibull με παραμέτρους $\lambda, p > 0$ αν και μόνο αν η συνάρτηση κινδύνου δίνεται από τη σχέση $h_X(t) = p\lambda^p t^{p-1}$.

Απόδειξη Με βάση το θεώρημα 1 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση κινδύνου. Επομένως, αρκεί να υπολογίσουμε τη συνάρτηση κινδύνου που στην περίπτωση της κατανομής Weibull είναι:

$$h_X(t) = \frac{f_X(t)}{\bar{F}_X(t)} = \frac{p\lambda^p t^{p-1}\, exp[-(\lambda t)^p]}{exp[-(\lambda t)^p]} = p\lambda^p t^{p-1}$$

Η κατανομή Weibull για $p > 1$ είναι IFR ⇒ DMRL, ενώ για $0 < p < 1$ είναι DFR ⇒ IMRL. Για $p = 1$ προκύπτει η εκθετική κατανομή η οποία έχει σταθερό ρυθμό αποτυχίας.

Μοντέλα που οδηγούν στην κατανομή Weibull

1. Αν η συνάρτηση κινδύνου κατά τη χρονική στιγμή t είναι δυναμοσυνάρτηση του χρόνου, τότε η κατανομή της τυχαίας μεταβλητής X είναι η Weibull με παραμέτρους $\lambda, p > 0$.

2. Η ασυμπτωτική κατανομή της μικρότερης διατεταγμένης στατιστικής συνάρτησης από μια προκαθορισμένη κατανομή πιθανότητας αποδεικνύεται ότι είναι κατανομή Weibull.

3.3 Κατανομή Γάμμα

Η κατανομή Γάμμα, η οποία εμπεριέχει την εκθετική και τη χ^2 κατανομή, είναι άλλη μια κατανομή που έχει ευρεία εφαρμογή στο πλαίσιο της ανάλυσης επιβίωσης.

Ορισμός 8 Μια συνεχής τυχαία μεταβλητή X με συνάρτηση επιβίωσης

$$\bar{F}_X(t) = \int_t^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta y} y^{\alpha-1} dy, \qquad t, y > 0 \tag{3.7}$$

ακολουθεί την *κατανομή Γάμμα* με παραμέτρους $\alpha, \beta > 0$, όπου $\Gamma(\alpha)$ είναι η συνάρτηση Γάμμα η οποία ορίζεται από τη σχέση

$$\Gamma(t) = \int_0^{\infty} e^{-y} y^{t-1} dy, \qquad t > 0.$$

Συμβολικά, $X \sim Γάμμα(\alpha, \beta)$.

Η αθροιστική συνάρτηση κατανομής της X δίνεται από τη σχέση

$$F_X(t) = \int_0^{t} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta y} y^{\alpha-1} dy, \qquad t, y > 0 \tag{3.8}$$

ενώ η συνάρτηση πυκνότητας πιθανότητας από τη σχέση

$$f_X(t) = -\frac{d}{dt}\bar{F}_X(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta t} t^{a-1}, \qquad t > 0 \tag{3.9}$$

Η κατανομή Γάμμα με παραμέτρους α και $\beta = 1$ ονομάζεται τυποποιημένη Γάμμα και η συνάρτηση πυκνότητας πιθανότητας δίνεται από τη σχέση

$$f_X(t) = \frac{1}{\Gamma(\alpha)} e^{-t} t^{a-1}, \ \ t > 0, a > 0$$

ενώ η αθροιστική συνάρτηση κατανομής από τη σχέση

$$I_a(t) \equiv F_X(t) = P(X \leq t) = \int_0^{t} \frac{1}{\Gamma(\alpha)} e^{-y} y^{a-1} dy, \qquad t, y > 0.$$

Βασικές Ιδιότητες

Έστω $X \sim Γάμμα(\alpha,\beta),\ t>0,\ \alpha,\beta>0$. Τότε

1. $E(X) = {}^{\alpha}/_{\beta}$
2. $V(X) = {}^{\alpha}/_{\beta^2}$

Πρόταση 5 Έστω X συνεχής τυχαία μεταβλητή που ορίζεται στο $[0,+\infty)$. Η X ακολουθεί την κατανομή Γάμμα με παραμέτρους $\alpha,\beta>0$ αν και μόνο αν η συνάρτηση κινδύνου δίνεται από τη σχέση

$$h_X(t) = \frac{\beta(\beta t)^{a-1} exp\,(-\beta t)}{\Gamma(\alpha)[1-I_a(\beta t)]} \tag{3.10}$$

Απόδειξη Με βάση το θεώρημα 1 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση κινδύνου. Επομένως, αρκεί να υπολογίσουμε τη συνάρτηση κινδύνου που στην περίπτωση της κατανομής Γάμμα είναι:

$$h_X(t) = \frac{f_X(t)}{\bar{F}_X(t)} = \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta t} t^{a-1}}{\int_t^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta y} y^{\alpha-1} dy} = \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta t} t^{a-1}}{1-\int_0^t \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta y} y^{a-1} dy}$$

$$= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta t} t^{a-1}}{1-\int_0^t \frac{\beta}{\Gamma(\alpha)} e^{-\beta y} (\beta y)^{\alpha-1} dy} = \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta t} t^{a-1}}{1-\int_0^{\beta t} \frac{1}{\Gamma(\alpha)} e^{-u} u^{a-1} du}$$

$$= \frac{\beta^\alpha e^{-\beta t} t^{a-1}}{\Gamma(\alpha)[1-I_a(\beta t)]} = \frac{\beta(\beta t)^{a-1} exp\,(-\beta t)}{\Gamma(\alpha)[1-I_a(\beta t)]}$$

Η κατανομή Γάμμα για $\alpha>1$ είναι IFR $\Rightarrow$ DMRL, ενώ για $0<\alpha<1$ είναι DFR $\Rightarrow$ IMRL. Για $\alpha=1$ προκύπτει η εκθετική κατανομή η οποία έχει σταθερό ρυθμό αποτυχίας.

Μοντέλα που οδηγούν στην κατανομή Γάμμα

1. Ο χρόνος αναμονής έως ότου πραγματοποιηθεί το n-οστό γεγονός σε μια διαδικασία Poisson με παράμετρο λ ακολουθεί την κατανομή Γάμμα με παραμέτρους $\alpha=n$ και $\beta=\lambda$.
2. Αν $X_1, X_2, \dots, X_n$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την εκθετική κατανομή με παράμετρο λ, τότε η τυχαία

μεταβλητή $X = X_1 + X_2 + \ldots + X_n$ ακολουθεί την κατανομή Γάμμα με παραμέτρους $\alpha = n$ και $\beta = \lambda$.

3.4 Λογαριθμοκανονική Κατανομή

Η κανονική κατανομή, η οποία χρησιμοποιείται ευρέως σε πολλές περιοχές έρευνας της Στατιστικής, έχει ένα σοβαρό μειονέκτημα ως κατανομή χρόνου ζωής. Αν $X \sim N(\mu, \sigma^2)$, τότε η X μπορεί θεωρητικά να πάρει και αρνητικές τιμές κάτι που δεν είναι αποδεκτό για χρόνο ζωής. Έτσι, συχνότερα από την κανονική, ως κατανομή χρόνου ζωής μπορεί να θεωρηθεί ένας μετασχηματισμός της η λογαριθμοκανονική κατανομή. Ο Mc Alister (1879) φαίνεται να είναι ο πρώτος που ασχολήθηκε λεπτομερώς με τη λογαριθμοκανονική κατανομή.

Ορισμός 9 Έστω X ο χρόνος ζωής της συνιστώσας ενός συστήματος κι έστω $Y = lnX \sim N(\mu, \sigma^2)$. Τότε η συνεχής τυχαία μεταβλητή X με συνάρτηση επιβίωσης

$$\bar{F}_X(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2}(\frac{lnx - \mu}{\sigma})^2\right] dx, t > 0, \mu \geq 0, \sigma > 0 \qquad (3.11)$$

ακολουθεί τη *λογαριθμοκανονική κατανομή* με παραμέτρους μ και σ^2. Συμβολικά, $X \sim \Lambda(\mu, \sigma^2)$.

Η αθροιστική συνάρτηση κατανομής της X δίνεται από τη σχέση

$$F_X(t) = \int_0^{t} \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2}(\frac{lnx - \mu}{\sigma})^2\right] dx, t > 0, \mu \geq 0, \sigma > 0 \qquad (3.12)$$

ενώ η συνάρτηση πυκνότητας πιθανότητας από τη σχέση

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma t} exp\left[-\frac{1}{2}(\frac{lnt - \mu}{\sigma})^2\right], t > 0, \mu \geq 0, \sigma > 0 \qquad (3.13)$$

Βασικές Ιδιότητες

Έστω $X \sim \Lambda(\mu, \sigma^2)$. Τότε

1. $E(X) = exp\left(\mu + \frac{\sigma^2}{2}\right)$
2. $V(X) = [exp(2\mu + \sigma^2)]\,[exp(\sigma^2) - 1]$

Πρόταση 6 Αν $X \sim N(\mu, \sigma^2)$, τότε $Y = e^X \sim \Lambda(\mu, \sigma^2)$.

Οι Gertsbach και Kordonskiy (1969) έδειξαν ότι η λογαριθμοκανονική κατανομή είναι IDFR ⇒ DIMRL.

<u>Μοντέλο που οδηγεί στην λογαριθμοκανονική κατανομή</u>

1. Η μορφή της συνάρτησης κινδύνου της λογαριθμοκανονικής κατανομής είναι ιδιαίτερα πολύπλοκη. Για το λόγο αυτό είναι πολύ δύσκολο να προκύψει η συνάρτηση επιβίωσης μέσω της συνάρτησης κινδύνου. Αποδεικνύεται όμως ότι η λογαριθμοκανονική κατανομή μπορεί να προκύψει μέσω φυσικής διαδικασίας κατά την οποία η «αποτυχία» συμβαίνει όταν η συνιστώσα του συστήματος υπερβεί κάποιο όριο αντοχής. Συγκεκριμένα, ας υποθέσουμε ότι $X_1 < X_2 < \cdots < X_n$ είναι μια ακολουθία τυχαίων μεταβλητών που περιγράφουν το βαθμό παραμόρφωσης σε διαδοχικά στάδια της εξέλιξής της. Η μεταβολή του βαθμού παραμόρφωσης $X_i - X_{\iota-1}$ στο στάδιο i είναι ανάλογη προς το μέγεθος του βαθμού παραμόρφωσης X_{i-1} και η συνιστώσα «αποτυγχάνει» όταν το μέγεθος του βαθμού παραμόρφωσης γίνει X_n. Αποδεικνύεται ότι το lnX_n για μεγάλο n ακολουθεί ασυμπτωτικά την κανονική κατανομή και κατά συνέπεια το X_n ακολουθεί τη λογαριθμοκανονική κατανομή.

3.5 Γεωμετρική Κατανομή

Η μεταβλητή που μας ενδιαφέρει στις μελέτες επιβίωσης είναι ο χρόνος ως την «αποτυχία». Πολύ συχνά λοιπόν, ο χρόνος μετριέται ή θεωρείται ως συνεχής τυχαία μεταβλητή όταν καταγράφεται σε ώρες ή μέρες κι έτσι η υπεροχή των συνεχών κατανομών έναντι των διακριτών είναι συντριπτική. Ωστόσο, κάποιοι συγγραφείς έχουν αναγνωρίσει την ανάγκη των διακριτών κατανομών για να μοντελοποιήσουν καταστάσεις όπως για παράδειγμα την επιθεώρηση ενός μηχανήματος ανά αραιά χρονικά διαστήματα. Η πιο απλή διακριτή κατανομή χρόνου ζωής, αντίστοιχη της εκθετικής, είναι η γεωμετρική κατανομή.

Ορισμός 10 Μια διακριτή τυχαία μεταβλητή X με συνάρτηση επιβίωσης

$$\bar{F}_X(t) = P(X > t) = q^{t+1}, \qquad t = 0,1,2,\dots, 0 < p < 1, q = 1 - p \qquad (3.14)$$

ακολουθεί τη *γεωμετρική κατανομή* με παράμετρο $p > 0$.

Συμβολικά, $X \sim G\ (p)$.

Η αθροιστική συνάρτηση κατανομής της X δίνεται από τη σχέση

$$F_X(t) = P(X \leq t) = 1 - q^{t+1} \qquad (3.15)$$

ενώ η συνάρτηση πιθανότητας από τη σχέση

$$P(X = t) = pq^t, \qquad t = 0,1,2,\dots, 0 < p < 1, q = 1 - p \qquad (3.16)$$

<u>Βασικές Ιδιότητες</u>

Έστω $X \sim G(p), t = 0,1,2,\dots,\ 0 < p < 1,\ q = 1 - p$. Τότε

1. $E(X) = {}^{q}/_{p}$
2. $V(X) = {}^{q}/_{p^2}$

Πρόταση 7 Έστω X διακριτή τυχαία μεταβλητή που παίρνει ακέραιες και μη αρνητικές τιμές. Η X ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p > 0$ αν και μόνο αν η συνάρτηση κινδύνου κατά τη χρονική στιγμή t είναι σταθερή και ίση με p.

Απόδειξη Με βάση το θεώρημα 1 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση κινδύνου. Επομένως, αρκεί να υπολογίσουμε τη συνάρτηση κινδύνου που στην περίπτωση της γεωμετρικής κατανομής είναι:

$$h_X(t) = \frac{P(X = t)}{P(X \geq t)} = \frac{P(X = t)}{P(X > t) + P(X = t)} = \frac{pq^t}{q^{t+1} + pq^t} = \frac{pq^t}{q^t(q + p)} = p$$

Πρόταση 8 Έστω X διακριτή τυχαία μεταβλητή που παίρνει ακέραιες και μη αρνητικές τιμές. Η X ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p > 0$ αν και μόνο αν η μέση υπολειπόμενη ζωή κατά τη χρονική στιγμή t είναι σταθερή και ίση με ${}^{1}/_{p}$.

Απόδειξη Με βάση το θεώρημα 3 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση μέσης υπολειπόμενης ζωής.

Επομένως, αρκεί να υπολογίσουμε τη μέση υπολειπόμενη ζωή που στην περίπτωση της γεωμετρικής κατανομής είναι:

$$\mu^X(t) = \frac{1}{\bar{F}_X(t)}\sum_{x=t}^{\infty}\bar{F}_X(x) = \frac{1}{q^{t+1}}\sum_{x=t}^{\infty}q^{x+1} = \frac{q}{q^{t+1}}\sum_{x=t}^{\infty}q^{x} = \frac{1}{q^t}\left(\frac{q^t}{1-q}\right) = \frac{1}{p}$$

Από τις δύο προηγούμενες προτάσεις παρατηρούμε ότι αν X είναι μια διακριτή τυχαία μεταβλητή που περιγράφει το χρόνο ζωής της συνιστώσας ενός συστήματος και $X \sim G(p)$, $\lambda > 0$, τότε το γινόμενο της συνάρτησης κινδύνου επί τη μέση υπολειπόμενη ζωή είναι σταθερό και ίσο με τη μονάδα.

<u>Μοντέλα που οδηγούν στη γεωμετρική κατανομή</u>

1. Ο αριθμός των αποτυχιών μέχρι να εμφανιστεί η πρώτη επιτυχία σε μια ακολουθία δοκιμών Bernoulli ακολουθεί τη γεωμετρική κατανομή.
2. Σταθερή συνάρτηση κινδύνου οδηγεί σε γεωμετρική κατανομή. Αν X διακριτή τυχαία μεταβλητή και $h_X(t) = p,\ p > 0$, τότε $X \sim G(p)$.
3. Σταθερή μέση υπολειπόμενη ζωή οδηγεί σε γεωμετρική κατανομή. Αν X διακριτή τυχαία μεταβλητή και $\mu^X(t) = {}^{1}\!/_{p}$, τότε $X \sim G(p)$.
4. Έλλειψη αθροιστικής μνήμης της γεωμετρικής κατανομής. Αν η τυχαία μεταβλητή X ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p > 0$, τότε η δεσμευμένη πιθανότητα ο χρόνος επιβίωσης X να υπερβεί το $t + y$ δοθέντος ότι έχει υπερβεί το t ισούται με την πιθανότητα το X να υπερβεί το y.

ΚΕΦΑΛΑΙΟ 4
ΣΤΑΘΜΙΣΜΕΝΕΣ ΚΑΤΑΝΟΜΕΣ

Οι *σταθμισμένες κατανομές (weighted distributions)* αναπτύχθηκαν στο πλαίσιο των τεχνικών στατιστικού σχεδιασμού. Εμφανίζονται κυρίως όταν οι παρατηρήσεις που παράγονται από μια στοχαστική διαδικασία καταγράφονται επηρεασμένες από κάποια συνάρτηση στάθμισης και βρίσκουν εφαρμογή σε έρευνες διάρκειας ζωής, σε έρευνες που αφορούν πληθυσμούς άγριων ζώων, μεγέθη ανθρώπινων οικογενειών κ.τ.λ. Σε αυτό το κεφάλαιο δίνεται ο ορισμός των σταθμισμένων κατανομών και μελετώνται κάποιες από τις βασικές τους ιδιότητες.

4.1 Ορισμοί

Έστω X μια τυχαία μεταβλητή με συνάρτηση (πυκνότητας) πιθανότητας $f_X(x;\theta)$ κι ας υποθέσουμε ότι θέλουμε να καταλήξουμε σε κάποια συμπεράσματα για την παράμετρο θ. Η συνήθης στατιστική μεθοδολογία προϋποθέτει ένα τυχαίο δείγμα από την $f_X(x)$. Στην πράξη όμως, υπάρχουν περιπτώσεις όπου είναι αδύνατο ή έχει υψηλό κόστος να έχουμε ένα τέτοιο δείγμα. Ακόμα και στην περίπτωση που είναι εφικτό ένα τυχαίο δείγμα, ο ερευνητής μπορεί να αποφασίσει να δουλέψει με ένα μεροληπτικό δείγμα δεδομένου ότι ένα μεροληπτικό δείγμα, προσεκτικά επιλεγμένο, μπορεί να δώσει περισσότερη πληροφορία αν σχετιστεί η κατάλληλη συνάρτηση στάθμισης με την συνάρτηση (πυκνότητας) πιθανότητας.

Ορισμός 11 Έστω X μια τυχαία μεταβλητή με συνάρτηση (πυκνότητας) πιθανότητας $f_X(x)$. Έστω επίσης $w(x)$ μια μη αρνητική συνάρτηση για την οποία υποθέτουμε ότι η $E[w(X)]$ υπάρχει. Θεωρούμε την τυχαία μεταβλητή X^w με συνάρτηση (πυκνότητας) πιθανότητας

$$f_X^w(x) = \frac{w(x)f_X(x)}{E[w(X)]} \tag{4.1}$$

Τότε η τυχαία μεταβλητή X^w ονομάζεται *σταθμισμένη τυχαία μεταβλητή* με συνάρτηση στάθμισης $w(x)$ και η $f_X(x)$ *σταθμισμένη συνάρτηση (πυκνότητας) πιθανότητας* που αντιστοιχεί στην αρχική συνάρτηση $f_X(x)$.

Η $E[w(X)]$ είναι ένας όρος που χρησιμοποιείται ούτως ώστε η συνολική πιθανότητα να είναι 1. Παρατηρούμε ότι $f = f^w$ αν και μόνο αν η $w(x)$ είναι σταθερά.

Για να κατανοήσουμε καλύτερα το νόημα και τη σημασία των σταθμισμένων κατανομών παραθέτουμε ένα παράδειγμα των Bayarri και DeGroot. Έστω X ο αριθμός των ομιλιών που παρακολουθούν οι συμμετέχοντες ενός στατιστικού συνεδρίου. Ας υποθέσουμε ότι η μεταβλητή X έχει συνάρτηση πιθανότητας $f_X(x;\theta)$ κι ότι μας ενδιαφέρει η συμπερασματολογία για το θ. Ένας προφανής τρόπος να πάρουμε ένα τυχαίο δείγμα $X_1, \ldots, X_\nu$ από την $f_X(x)$ είναι να επιλέξουμε τυχαία ν συμμετέχοντες από τη λίστα του συνεδρίου και να τους ρωτήσουμε πόσες ομιλίες παρακολούθησαν. Αυτή η διαδικασία φαίνεται βολική και εύκολη στην πραγματικότητα όμως δεν είναι. Ένας πιο απλός και εύκολος τρόπος για να καταγράψουμε τις παρατηρήσεις X_i θα ήταν να επιλέξουμε τυχαία μια ή περισσότερες ομιλίες και να πάρουμε συνέντευξη από ν συμμετέχοντες από αυτούς που παρακολούθησαν την ομιλία (ή τις ομιλίες). Βέβαια σε αυτή την περίπτωση, όσες περισσότερες ομιλίες παρακολουθεί κάποιος τόσο πιο πιθανό είναι να συμπεριληφθεί στο δείγμα. Επομένως, οι παρατηρήσεις δεν προέρχονται από την $f_X(x)$ αλλά από μια άλλη κατανομή που είναι η αρχική σταθμισμένη από μια αύξουσα συνάρτηση του x.

Ο πίνακας 4.1 περιέχει κάποιες από τις συναρτήσεις στάθμισης που χρησιμοποιούνται συνήθως στην πράξη. Οι συναρτήσεις αυτές είναι μονότονες, αύξουσες ή φθίνουσες. Σε κάθε περίπτωση, η επιλογή της κατάλληλης συνάρτησης στάθμισης εξαρτάται από τη φύση του προβλήματος αλλά και από την εμπειρία του ερευνητή.

Συναρτήσεις Στάθμισης
$w(x) = x, \qquad x \geq 0$
$w(x) = x^a \; για \; α = 1, 2, 3, \frac{1}{2}, \frac{2}{3} \; και \; για \; 0 < α < 1$
$w(x) = \binom{x}{2}^a \; για \; α = 1, \frac{1}{2} \; και \; για \; 0 < α < 1$
$w(x) = x^{(a)} = x(x-1)\dots(x-a+1)$
$w(x) = 1 - (1-β)^x \; για \; 0 < β < 1$
$w(x) = x + 1, \qquad ax + β$
$w(x) = e^{ax}, \qquad θ^x$
$w(x) = (αx + β)/(δx + γ)$
$w(x) = P(Y \leq x) \; για \; κάποια \; τ.μ. Y$
$w(x) = P(Y > x) \; για \; κάποια \; τ.μ. Y$
$w(x) = r(x), όπου \; r(x) είναι \; η \; πιθανότητα \; επιβίωσης \; της \; παρατήρησης \; x$

Πίνακας 4.1: Τύποι των Συναρτήσεων Στάθμισης

Στην περίπτωση που η συνάρτηση στάθμισης έχει τύπο $w(x) = x^a$ η τυχαία μεταβλητή X^w ονομάζεται *μεροληπτική ως προς το μέγεθος εκδοχή της X τάξης α (size-biased version of X of order α)* και συμβολίζεται με X^{*a}. Η συνάρτηση πυκνότητας πιθανότητας της X^{*a} δίνεται από τη σχέση:

$$f_X^{*a}(x) = \frac{x^a f_X(x)}{E(X^a)} \tag{4.2}$$

Στην ειδική περίπτωση όπου $\alpha = 1$, δηλαδή όταν η συνάρτηση στάθμισης εξαρτάται από το μήκος της μονάδας, τότε η τυχαία μεταβλητή X^w ονομάζεται *μεροληπτική ως προς το μήκος εκδοχή της X (length-biased version of X)* και συμβολίζεται με X^*. Η συνάρτηση πυκνότητας πιθανότητας της X^* δίνεται από τη σχέση:

$$f_X^*(x) = \frac{x f_X(x)}{E(X)} \tag{4.3}$$

Η συνάρτηση στάθμισης $w(x) = x$ εμφανίζεται συχνά στη δειγματοληψία. Σύμφωνα με τον Cox, η $f_X^*(x)$ έχει την ακόλουθη ερμηνεία: Ας υποθέσουμε ότι θέλουμε να πάρουμε ένα δείγμα από έναν πληθυσμό μετρώντας τον χρόνο αποτυχίας που έχει συνάρτηση πυκνότητας πιθανότητας $f_X(x)$ κι ότι η πιθανότητα επιλογής οποιασδήποτε μονάδας είναι ανάλογη του μήκους της x. Τότε οι επιλεγμένοι χρόνοι αποτυχίας έχουν συνάρτηση πυκνότητας πιθανότητας $f_X^*(x)$.

Η μεροληπτική ως προς το μήκος κατανομή βρίσκει ευρεία εφαρμογή σε βιομετρικές έρευνες όπως η έγκαιρη ανίχνευση μιας ασθένειας π.χ. καρκίνος του μαστού (Zelen & Feinleib (1969)). O Rao (1965) χρησιμοποίησε τη μεροληπτική ως προς το μήκος κατανομή σε μελέτες που αφορούσαν πληθυσμούς άγριων ζώων και μεγέθη ανθρώπινων οικογενειών.

Η συνάρτηση πιθανότητας $f_X^*(x)$ συχνά αναφέρεται και ως κατανομή παραγοντικής ροπής πρώτης τάξης. Παρακάτω δίνεται ο ορισμός της κατανομής παραγοντικής ροπής r τάξης.

Ορισμός 12 Έστω X μια μη αρνητική ακέραια τυχαία μεταβλητή με συνάρτηση πιθανότητας $f_X(x)$. Τότε η αντίστοιχη *κατανομή παραγοντικής ροπής r τάξης (r-th factorial moment distribution)* δίνεται από τη σχέση

$$f_X^{*(r)}(x) = \frac{x^{(r)} f_X(x)}{E(X^{(r)})} \tag{4.4}$$

με $E\left(X^{(r)}\right) < \infty$ και $x^{(r)}$ το καθοδικό παραγοντικό r τάξης.

4.2 Βασικές Ιδιότητες

Σε αυτή την υποενότητα δίνονται οι βασικές ιδιότητες των σταθμισμένων κατανομών. Αρχικά παρουσιάζονται οι ιδιότητες που αφορούν στις μεροληπτικές ως προς το μήκος κατανομές μιας και αυτές εμφανίζονται συχνότερα στην πράξη και εν συνεχεία δίνονται με τη μορφή θεωρημάτων οι ιδιότητες που αφορούν στη γενική περίπτωση όπου η συνάρτηση στάθμισης είναι $w(x) > 0$. Κάποιες από αυτές τις ιδιότητες δόθηκαν από τους Mahfoud και Patil (1982), άλλες από τους Patil και Rao (1978).

Έστω X μια μη αρνητική τυχαία μεταβλητή με συνάρτηση (πυκνότητας) πιθανότητας $f_X(x)$. Έστω επίσης X^* η μεροληπτική ως προς το μήκος εκδοχή της X. Υπό την προϋπόθεση ότι οι τρεις πρώτες ροπές υπάρχουν, ισχύουν οι παρακάτω ιδιότητες:

1. Η διαφορά της αναμενόμενης τιμής της X από την αναμενόμενη τιμή της X^* ισούται με τον λόγο της διακύμανσης προς την αναμενόμενη τιμή της X. Αν λοιπόν $E(X^*) = \mu^*$, τότε
$$\mu^* - \mu = \frac{\mu_2}{\mu}$$
2. Η διαφορά της διακύμανσης της X από τη διακύμανση της X^* ισούται με το τετράγωνο του λόγου της διακύμανσης προς την αναμενόμενη τιμή της X επί την ποσότητα $\frac{\gamma_1}{CV} - 1$, όπου CV είναι ο συντελεστής μεταβλητότητας που δίνεται από τη σχέση $CV = \frac{\mu_2^{1/2}}{\mu}$ και γ_1 ο συντελεστής ασυμμετρίας που δίνεται από τη σχέση $\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$. Αν λοιπόν $V(X^*) = \mu_2^*$, τότε
$$\mu_2^* - \mu_2 = \left(\frac{\mu_2}{\mu}\right)^2 \left(\frac{\gamma_1}{CV} - 1\right)$$
3. Ο αρμονικός μέσος της X^* ισούται με την αναμενόμενη τιμή της X όταν η X είναι θετική. Αν λοιπόν $H(X^*)$ ο αρμονικός μέσος της X^*, τότε
$$H(X^*) = \mu = \left[E\left(\frac{1}{X^*}\right)\right]^{-1}$$

4. Ο λόγος της διακύμανσης προς την αναμενόμενη τιμή της X, όταν η X είναι θετική, ισούται με τη διαφορά του αρμονικού μέσου της X^* από την αναμενόμενη τιμή της X^*.

$$\frac{\mu_2}{\mu} = \mu^* - H(X^*)$$

Το επόμενο θεώρημα που δόθηκε από τους Patil et al. (1986) μας επιτρέπει τη σύγκριση των X και X^w.

Θεώρημα 5 Η τυχαία μεταβλητή X^w είναι στοχαστικά μεγαλύτερη (μικρότερη) από την X αν η συνάρτηση στάθμισης $w(x)$ είναι αύξουσα (φθίνουσα) συνάρτηση του x.

Η συνδιακύμανση των X και $w(X)$ είναι ένα μέτρο γραμμικής σχέσης μεταξύ τους. Η σχέση αυτή εξαρτάται από το πρόσημο της συνδιακύμανσης και δείχνει τη σχέση μεταξύ των αναμενόμενων τιμών των X και X^w.

Θεώρημα 6 Έστω X τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$. Έστω επίσης η συνάρτηση στάθμισης $w(x) > 0$ με $E[w(X)] < \infty$ και X^w η σταθμισμένη εκδοχή της X με συνάρτηση πυκνότητας πιθανότητας $f_X^w(x)$. Τότε $E(X^w) > E(X)$ αν $cov[X, w(X)] > 0$ και $E(X^w) < E(X)$ αν $cov[X, w(X)] < 0$.

Η μονοτονία της συνάρτησης στάθμισης δείχνει τη σχέση μεταξύ των αναμενόμενων τιμών των X και X^w. Η κατανομή η οποία σταθμίζεται από αύξουσα (φθίνουσα) συνάρτηση δίνει κατά μέσο όρο μεγαλύτερες (μικρότερες) τιμές από την αρχική.

Θεώρημα 7 Έστω X τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$. Έστω επίσης η συνάρτηση στάθμισης $w(x) > 0$ με $E[w(X)] < \infty$ και X^w η σταθμισμένη εκδοχή της X με συνάρτηση πυκνότητας πιθανότητας $f_X^w(x)$. Τότε $E(X^w) > E(X)$ αν η $w(x)$ είναι αύξουσα συνάρτηση του x και $E(X^w) < E(X)$ αν η $w(x)$ είναι φθίνουσα συνάρτηση του x.

Ενδιαφέρον παρουσιάζει το επόμενο θεώρημα αναφορικά με τις αναμενόμενες τιμές δύο σταθμισμένων τυχαίων μεταβλητών, αφού το κριτήριο για το είδος της μεταξύ τους σχέσης είναι η μονοτονία του λόγου των συναρτήσεων στάθμισης.

Θεώρημα 8 Έστω X τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$. Έστω οι συναρτήσεις στάθμισης $w_i(x) > 0$ για τις οποίες $E[w_i(X)] < \infty$ και X^{w_i} οι σταθμισμένες εκδοχές της X για $i = 1, 2$. Τότε $E(X^{w_2}) > E(X^{w_1})$ αν η $\frac{w_2(x)}{w_1(x)}$ είναι αύξουσα συνάρτηση του x και $E(X^{w_2}) < E(X^{w_1})$ αν η $\frac{w_2(x)}{w_1(x)}$ είναι φθίνουσα συνάρτηση του x.

Αυτό που έχει καθοριστική σημασία είναι η μονοτονία του λόγου των συναρτήσεων στάθμισης κι όχι η ανισοτική σχέση μεταξύ τους. Για παράδειγμα, έστω $w_1(x) = x(x-1)$ και $w_2 = x^2$. Ο λόγος $\frac{w_1}{w_2}$ είναι ίσος με $\frac{x}{x-1} = \frac{1}{1-\frac{1}{x}}$ που είναι φθίνουσα συνάρτηση του x. Έτσι έχουμε $E(X^{w_2}) < E(X^{w_1})$ κι όχι το αντίθετο όπως θα περίμενε κανείς μιας και $w_1(x) < w_2(x)$.

Θεώρημα 9 (Kirmani & Ahsanullah (1987)) Έστω X μια μη αρνητική τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$ τέτοια ώστε $P(X > x) > 0$ για κάθε x κι έστω $w(x)$ μια μη αρνητική και αύξουσα συνάρτηση στο $[0, +\infty)$ με $E[w(X)] < \infty$ και $w(x) \to +\infty$ καθώς $x \to +\infty$. Έστω επίσης X^w η σταθμισμένη εκδοχή της X με συνάρτηση πυκνότητας πιθανότητας $f_X^w(x)$. Τότε υπάρχει μια μη αρνητική τυχαία μεταβλητή Y ανεξάρτητη της X^w τέτοια ώστε $X =^d \min(X^w, Y)$.

4.3 Τύπος Αμεταβλητότητας υπό Μεροληπτική Λόγω Μεγέθους Δειγματοληψία

Το κύριο ενδιαφέρον στις έρευνες που απαιτούν στατιστική ανάλυση είναι να καθορίσουμε τα χαρακτηριστικά της αρχικής κατανομής μέσω της παρατηρούμενης κατανομής. Η παρατηρούμενη κατανομή χρησιμοποιείται σαν εργαλείο έτσι ώστε να καθοριστεί η αρχική κατανομή και να γίνει συμπερασματολογία για την παράμετρο που μας ενδιαφέρει. Για το λόγο αυτό, θα πρέπει να μελετηθεί η επίδραση της συνάρτησης στάθμισης στην αρχική κατανομή. Ιδιαίτερο λοιπόν ενδιαφέρον παρουσιάζει το αν και ποιες κατανομές χαρακτηρίζονται από την ιδιότητα του *τύπου αμεταβλητότητας υπό μεροληπτική λόγω μεγέθους δειγματοληψία (form-invariance under size-biased sampling).*

Ορισμός 13 Έστω X μια τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(x;\theta)$. Θα λέμε ότι η $f_X(x;\theta)$ διατηρεί αμετάβλητο τον τύπο της υπό μεροληπτική λόγω μεγέθους δειγματοληψία τάξης α αν η τυχαία μεταβλητή X^{*a} έχει συνάρτηση πυκνότητας πιθανότητας $f_X^{*a}(x;\eta)$, όπου η μια νέα παράμετρος που εξαρτάται από το θ και τη σταθερά α.
Συμβολικά, $f_X^{*a}(x;\eta) \equiv f_X(x;\theta)$.

Σύμφωνα με τον παραπάνω ορισμό, για μια δοσμένη f, η η είναι μια συνάρτηση του α κι όταν $\alpha = 0$ τότε $\theta = \eta$. Έτσι, καθίσταται άμεση η σύνδεση μεταξύ αρχικής και σταθμισμένης κατανομής. Εδώ θα πρέπει να αναφέρουμε ότι εκτεταμένη μελέτη της ιδιότητας του τύπου αμεταβλητότητας έγινε από τους Patil και Ord (1976), οι οποίοι απέδειξαν ότι υπό συγκεκριμένες συνθήκες η ιδιότητα αυτή χαρακτηρίζει τη λογαριθμο-εκθετική οικογένεια κατανομών.

Στη συνέχεια εξετάζεται ποιες από τις κατανομές που αναπτύχθηκαν στο προηγούμενο κεφάλαιο διατηρούν τον τύπο τους αμετάβλητο υπό μεροληπτική λόγω μεγέθους δειγματοληψία τάξης α.

<u>Εκθετική Κατανομή</u>

Η εκθετική κατανομή δεν διατηρεί αμετάβλητο τον τύπο της υπό μεροληπτική λόγω μεγέθους δειγματοληψία τάξης α.

Θεώρημα 10 Έστω ότι η τυχαία μεταβλητή X ακολουθεί την εκθετική κατανομή με παράμετρο λ. Η σταθμισμένη τυχαία μεταβλητή $X^{*\alpha}$ ακολουθεί την κατανομή Γάμμα με παραμέτρους $\alpha + 1$ και λ.

Απόδειξη Με βάση τη σχέση (4.1) για την εκθετική κατανομή και για $w(x) = x^a$ έχουμε:

$$f_X^w(x) = \frac{w(x)f_X(x)}{E[w(X)]} = \frac{x^a \lambda\, exp\,(-\lambda x)}{E(X^a)} = \frac{x^a \lambda\, exp\,(-\lambda x)}{\frac{\Gamma(\alpha+1)}{\lambda^\alpha}}$$

$$= \frac{x^{(a+1)-1}\lambda^{\alpha+1} exp\,(-\lambda x)}{\Gamma(\alpha+1)}$$

Στην ειδική περίπτωση όπου $\alpha = 1$, $X^* \sim \Gamma\acute{\alpha}\mu\mu\alpha(2, \lambda)$.

Κατανομή Weibull

Η κατανομή Weibull δεν διατηρεί αμετάβλητο τον τύπο της υπό μεροληπτική λόγω μεγέθους δειγματοληψία τάξης α.

Θεώρημα 11 Έστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Weibull με παραμέτρους λ και p. Η σταθμισμένη τυχαία μεταβλητή $X^{*\alpha}$ ακολουθεί τη γενικευμένη κατανομή Γάμμα με παραμέτρους λ, p, $\frac{a+p}{p}$.

Απόδειξη Με βάση τη σχέση (4.1) για την κατανομή Weibull και για $w(x) = x^a$ έχουμε:

$$f_X^w(x) = \frac{w(x)f_X(x)}{E[w(X)]} = \frac{x^a p\lambda^p x^{p-1}\, exp[-(\lambda x)^p]}{E(X^a)} = \frac{x^a p\lambda^p x^{p-1}\, exp[-(\lambda x)^p]}{\frac{a\Gamma\left(\frac{a}{p}\right)}{p\lambda^\alpha}}$$

$$= \frac{x^{a+p-1}p\lambda^{\alpha+p} exp[-(\lambda x)^p]}{\frac{a}{p}\Gamma\left(\frac{a}{p}\right)} = \frac{x^{a+p-1}p\lambda^{\alpha+p} exp[-(\lambda x)^p]}{\Gamma\left(\frac{a+p}{p}\right)}$$

$$= \frac{p\lambda^{p\frac{a+p}{p}} x^{p\frac{a+p}{p}-1} exp[-(\lambda x)^p]}{\Gamma\left(\frac{\alpha+p}{p}\right)}$$

Στην ειδική περίπτωση όπου $\alpha = 1$, $X^* \sim \Gamma\varepsilon\nu.\Gamma\acute{\alpha}\mu\mu\alpha(\lambda, p, \frac{1+p}{p})$.

Κατανομή Γάμμα

Η κατανομή Γάμμα διατηρεί αμετάβλητο τον τύπο της υπό μεροληπτική λόγω μεγέθους δειγματοληψία τάξης α.

Θεώρημα 12 Έστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Γάμμα με παραμέτρους β και λ. Η σταθμισμένη τυχαία μεταβλητή $X^{*\alpha}$ ακολουθεί την κατανομή Γάμμα με παραμέτρους $\beta+\alpha$ και λ.

Απόδειξη Με βάση τη σχέση (4.1) για την κατανομή Γάμμα και για $w(x) = x^a$ έχουμε:

$$f_X^w(x) = \frac{w(x)f_X(x)}{E[w(X)]} = \frac{x^a \dfrac{\lambda^\beta x^{\beta-1}e^{-\lambda x}}{\Gamma(\beta)}}{\dfrac{\Gamma(\beta+\alpha)}{\lambda^\alpha \Gamma(\beta)}} = \frac{\lambda^{\beta+\alpha}x^{(\beta+\alpha)-1}e^{-\lambda x}}{\Gamma(\beta+\alpha)}$$

Στην ειδική περίπτωση όπου $\alpha = 1$, $X^*\sim \Gamma \acute{\alpha} \mu\mu\alpha(\beta+1,\lambda)$.

<u>Λογαριθμοκανονική Κατανομή</u>

Η λογαριθμοκανονική κατανομή διατηρεί αμετάβλητο τον τύπο της υπό μεροληπτική λόγω μεγέθους δειγματοληψία τάξης α.

Θεώρημα 13 Έστω ότι η τυχαία μεταβλητή X ακολουθεί τη λογαριθμοκανονική κατανομή με παραμέτρους μ και σ^2. Η σταθμισμένη τυχαία μεταβλητή $X^{*\alpha}$ ακολουθεί τη λογαριθμοκανονική κατανομή με παραμέτρους $\mu+\alpha\sigma^2$ και σ^2.

Απόδειξη Με βάση τη σχέση (4.1) για τη λογαριθμοκανονική κατανομή και για $w(x) = x^a$ έχουμε:

$$\begin{aligned}
f_X^w(x) = \frac{w(x)f_X(x)}{E[w(X)]} &= \frac{x^a \dfrac{1}{\sqrt{2\pi}\sigma x} exp\left[-\dfrac{1}{2}(\dfrac{lnx-\mu}{\sigma})^2\right]}{E(X^a)} \\
&= \frac{x^a \dfrac{1}{\sqrt{2\pi}\sigma x} exp\left[-\dfrac{1}{2}(\dfrac{lnx-\mu}{\sigma})^2\right]}{exp\left(\alpha\mu + \dfrac{\alpha^2\sigma^2}{2}\right)} \\
&= x^a \frac{1}{\sqrt{2\pi}x\sigma} exp\left(-\frac{1}{2}\frac{(lnx-\mu)^2}{\sigma^2} - \alpha\mu - \frac{1}{2}\alpha^2\sigma^2\right) \\
&= \frac{1}{\sqrt{2\pi}x\sigma} exp\left(lnx^a - \frac{1}{2}\frac{(lnx-\mu)^2}{\sigma^2} - \alpha\mu - \frac{1}{2}\alpha^2\sigma^2\right) \\
&= \frac{1}{\sqrt{2\pi}x\sigma} exp\left(-\frac{1}{2}\frac{(-lnx+\mu+\alpha\sigma^2)^2}{\sigma^2}\right)
\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}x\sigma} exp\left(-\frac{1}{2}\frac{\left(lnx - (\mu + \alpha\sigma^2)\right)^2}{\sigma^2}\right)$$

Στην ειδική περίπτωση όπου $\alpha = 1$, $X^* \sim \Lambda(\mu + \sigma^2, \sigma^2)$.

<u>Γεωμετρική Κατανομή</u>

Η γεωμετρική κατανομή δεν διατηρεί αμετάβλητο τον τύπο της υπό μεροληπτική λόγω μεγέθους δειγματοληψία παραγοντικής τάξης r.

Θεώρημα 14 Έστω ότι η τυχαία μεταβλητή X ακολουθεί τη γεωμετρική κατανομή με παράμετρο p. Η σταθμισμένη τυχαία μεταβλητή $X^{*(r)}$ ακολουθεί τη μετατοπισμένη κατά r μονάδες δεξιά αρνητική διωνυμική κατανομή με παραμέτρους $1 + r$ και p.

Απόδειξη Με βάση τη σχέση (4.4) για τη γεωμετρική κατανομή έχουμε:

$$f_X^{*(r)}(x) = \frac{x^{(r)} q^x p}{E[X^{(r)}]} = \frac{x^{(r)} q^x p}{r!\, q^r p^{-r}} = \binom{x}{r} q^{x-r} p^{r+1} = \binom{x}{x-r} q^{x-r} p^{r+1}$$

Στην ειδική περίπτωση όπου $r = 1$, η X^* ακολουθεί τη μετατοπισμένη κατά μία μονάδα δεξιά αρνητική διωνυμική κατανομή με παραμέτρους 2 και p.

ΚΕΦΑΛΑΙΟ 5
ΜΕΤΡΑ ΑΞΙΟΠΙΣΤΙΑΣ ΣΤΑΘΜΙΣΜΕΝΩΝ ΚΑΤΑΝΟΜΩΝ

Ιδιαίτερο ενδιαφέρον στο πλαίσιο της θεωρίας αξιοπιστίας παρουσιάζουν οι σχέσεις που αφορούν στα μέτρα αξιοπιστίας μεταξύ αρχικής και σταθμισμένης κατανομής και οι οποίες εξαρτώνται κυρίως από τον τύπο της συνάρτησης στάθμισης. Η συνεχής περίπτωση μελετήθηκε από τους Jain et al. (1989), ενώ η διακριτή από τις Δημάκη και Γιακουμάτου (1998). Παραδείγματα σχετικά με τα αποτελέσματα των θεωρημάτων δίνονται μέσω της εκθετικής και της γεωμετρικής κατανομής αντίστοιχα για $w(x) = x$. Τέλος, ενδιαφέροντα είναι και τα αποτελέσματα που προκύπτουν αναφορικά με τα κριτήρια γήρανσης και τις διάφορες ανισοτικές σχέσεις.

5.1 Συνεχής Περίπτωση

Πριν δώσουμε το θεώρημα που αφορά στη σχέση των συναρτήσεων επιβίωσης των τυχαίων μεταβλητών X και X^w κρίνεται σκόπιμο να παραθέσουμε το παρακάτω λήμμα το οποίο θα μας βοηθήσει στην απόδειξη του θεωρήματος.

Λήμμα 1 (Δημάκη και Ξεκαλάκη (1996)) Έστω X συνεχής τυχαία μεταβλητή με συνάρτηση επιβίωσης $\bar{F}_X(t)$ για όλα τα $t \geq 0$. Έστω επίσης $w(t)$ μια διαφορίσιμη συνάρτηση τέτοια ώστε $E[w(X)] < \infty$. Τότε

$$E[w(X)|\, X > t] = w(t) + \frac{1}{\bar{F}_X(t)} \int_t^{\infty} \bar{F}_X(x) dw(x) \qquad (5.1)$$

Θεώρημα 15 Έστω X μια μη αρνητική συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(t)$ και συνάρτηση κατανομής $F_X(t)$. Έστω ότι η συνάρτηση στάθμισης $w(t)$ είναι θετική με $0 \neq E[w(X)] < \infty$ και $E(X^2) < \infty$. Η αντίστοιχη σταθμισμένη τυχαία μεταβλητή X^w έχει συνάρτηση πυκνότητας πιθανότητας $f_X^w(t)$ και συνάρτηση κατανομής $F_{X^w}(t)$. Τότε η συνάρτηση επιβίωσης $\bar{F}_{X^w}(t)$ μπορεί να εκφραστεί ως εξής:

$$\bar{F}_{X^w}(t) = \frac{1}{E[w(X)]}\bar{F}_X(t)E[w(X)|\, X > t] \tag{5.2}$$

Απόδειξη

$$\bar{F}_{X^w}(t) = 1 - F_{X^w}(t)$$

$$= 1 - \int_0^t f_{X^w}(x)dx = \int_t^\infty f_{X^w}(x)dx = \frac{1}{E[w(X)]}\int_t^\infty w(x)f_X(x)dx$$

$$= \frac{1}{E[w(X)]}\int_t^\infty w(x)F_X'(x)dx = \frac{1}{E[w(X)]}\int_t^\infty w(x)[1-\bar{F}_X(x)]'dx$$

$$= -\frac{1}{E[w(X)]}\int_t^\infty w(x)\bar{F}_X'(x)dx =$$

$$-\frac{1}{E[w(X)]}\left[-w(t)\bar{F}_X(t) - \int_t^\infty \bar{F}_X(x)dw(x)\right]$$

$$= \frac{w(t)\bar{F}_X(t)}{E[w(X)]} + \frac{1}{E[w(X)]}\int_t^\infty \bar{F}_X(x)dw(x)$$

$$= \frac{\bar{F}_X(t)}{E[w(X)]}\left[w(t) + \frac{1}{\bar{F}_X(t)}\int_t^\infty \bar{F}_X(x)dw(x)\right]$$

$$= \frac{\bar{F}_X(t)}{E[w(X)]}E[w(X)|\, X > t]$$

Παράδειγμα 1: Η εκθετική κατανομή

Έστω X μια συνεχής τυχαία μεταβλητή η οποία ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda > 0$. Στο προηγούμενο κεφάλαιο αποδείχθηκε ότι αν $X \sim exp\,(\lambda)$, τότε $X^* \sim Γάμμα(2, \lambda)$. Κατά συνέπεια χρησιμοποιώντας τη σχέση (3.7) έχουμε ότι

$$\bar{F}_{X^*}(t) = \int_t^\infty \frac{\lambda^2}{\Gamma(2)}e^{-\lambda y}ydy = \lambda^2\int_t^\infty e^{-\lambda y}\,ydy = \lambda^2\left[\frac{t}{\lambda}e^{-\lambda t} + \frac{1}{\lambda}\int_t^\infty e^{-\lambda y}dy\right]$$

$$= \lambda^2\left[\frac{t}{\lambda}e^{-\lambda t} + \frac{1}{\lambda^2}e^{-\lambda t}\right] = \lambda t e^{-\lambda t} + e^{-\lambda t} = e^{-\lambda t}(\lambda t + 1)$$

Θα επιβεβαιώσουμε το αποτέλεσμα αυτό χρησιμοποιώντας το θεώρημα 15. Με βάση το λήμμα 1

$$E(X|\, X>t) = t + e^{\lambda t}\int_t^{\infty} e^{-\lambda x}dx = t + e^{\lambda t}\frac{1}{\lambda}e^{-\lambda t} = t + \frac{1}{\lambda}$$

Άρα από τη σχέση (5.2) έχουμε ότι

$$\bar{F}_{X^*}(t) = \frac{\bar{F}_X(t)E(X|\, X>t)}{E(X)} = \frac{e^{-\lambda t}\left(t+\frac{1}{\lambda}\right)}{\frac{1}{\lambda}} = e^{-\lambda t}(\lambda t+1)$$

Θεώρημα 16 Έστω X και X^w τυχαίες μεταβλητές όπως ορίστηκαν στο Θεώρημα 15. Έστω επίσης $h_X(t)$ η συνάρτηση κινδύνου της X. Τότε η συνάρτηση κινδύνου $h_{X^w}(t)$ της σταθμισμένης τυχαίας μεταβλητής X^w μπορεί να εκφραστεί ως εξής:

$$h_{X^w}(t) = \frac{w(t)h_X(t)}{E[w(X)|\, X>t]} \tag{5.3}$$

Απόδειξη

$$h_X(t) = \frac{f_X(t)}{\bar{F}_X(t)} = \frac{\frac{w(t)f_X(t)}{E[w(X)]}}{\frac{1}{E[w(X)]}\bar{F}_X(t)E[w(X)|\, X>t]} = \frac{w(t)f_X(t)}{\bar{F}_X(t)E[w(X)|\, X>t]}$$
$$= \frac{w(t)h_X(t)}{E[w(X)|\, X>t]}$$

<u>Παράδειγμα 2</u>: Η εκθετική κατανομή

Έστω X μια συνεχής τυχαία μεταβλητή η οποία ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda>0$. Γνωρίζοντας ότι αν $X\sim exp\,(\lambda)$, τότε $X^*\sim \Gamma\acute{\alpha}\mu\mu\alpha(2,\lambda)$ και χρησιμοποιώντας τη σχέση (3.10) έχουμε ότι

$$h_{X^*}(t) = \frac{\lambda^2 te^{-\lambda t}}{\Gamma(2)[1-I_2(\lambda t)]} = \frac{\lambda^2 te^{-\lambda t}}{1-\int_0^{\lambda t}e^{-y}ydy} = \frac{\lambda^2 te^{-\lambda t}}{1-\left[-e^{-\lambda t}\lambda t+\int_0^{\lambda t}e^{-y}dy\right]}$$
$$= \frac{\lambda^2 te^{-\lambda t}}{1-[-e^{-\lambda t}\lambda t+(-e^{-\lambda t}+1)]} = \frac{\lambda^2 te^{-\lambda t}}{1+e^{-\lambda t}\lambda t+e^{-\lambda t}-1} = \frac{\lambda^2 t}{\lambda t+1}$$

Θα επιβεβαιώσουμε το αποτέλεσμα αυτό χρησιμοποιώντας το θεώρημα 16. Προηγουμένως και με βάση το λήμμα 1 βρήκαμε ότι

$$E(X|\, X > t) = t + \frac{1}{\lambda}$$

Άρα από τη σχέση (5.3) έχουμε ότι

$$h_{X^*}(t) = \frac{t h_X(t)}{E[X|\, X > t]} = \frac{t\lambda}{t + \frac{1}{\lambda}} = \frac{\lambda^2 t}{\lambda t + 1}$$

Θεώρημα 17 Έστω X και X^w τυχαίες μεταβλητές όπως ορίστηκαν στο Θεώρημα 15. Έστω επίσης $\mu^X(t)$ η μέση υπολειπόμενη ζωή της X. Τότε η μέση υπολειπόμενη ζωή $\mu^{X^w}(t)$ της σταθμισμένης τυχαίας μεταβλητής X^w μπορεί να εκφραστεί ως εξής:

$$\mu^{X^w}(t) = \frac{\mu^X(t)}{E[w(X)|\, X > t]} \int_t^{\infty} \frac{E[w(X)|\, X > x]}{\mu^X(x)} exp\left[-\int_t^x \frac{du}{\mu^X(u)}\right] dx \qquad (5.4)$$

Απόδειξη

$$\mu^{X^w}(t) = \int_t^{\infty} \frac{\bar{F}_{X^w}(x)}{\bar{F}_{X^w}(t)} dx$$

$$= \int_t^{\infty} \frac{\bar{F}_X(x) E[w(X)|\, X > x]}{\bar{F}_X(t) E[w(X)|\, X > t]} dx$$

$$= \int_t^{\infty} \frac{\frac{\mu^X(0)}{\mu^X(x)} exp\left[-\int_0^x \frac{du}{\mu^X(u)}\right] E[w(X)|\, X > x]}{\frac{\mu^X(0)}{\mu^X(t)} exp\left[-\int_0^t \frac{du}{\mu^X(u)}\right] E[w(X)|\, X > t]} dx$$

$$= \frac{\mu^X(t)}{E[w(X)|\, X > t]} \int_t^{\infty} \left\{ \frac{E[w(X)|\, X > x]}{\mu^X(x)} exp\left[-\int_0^x \frac{du}{\mu^X(u)} + \int_0^t \frac{du}{\mu^X(u)}\right] \right\} dx$$

$$= \frac{\mu^X(t)}{E[w(X)|\, X > t]} \int_t^{\infty} \left\{ \frac{E[w(X)|\, X > x]}{\mu^X(x)} exp\left[-\int_0^x \frac{du}{\mu^X(u)} - \int_t^0 \frac{du}{\mu^X(u)}\right] \right\} dx$$

$$= \frac{\mu^X(t)}{E[w(X)|\, X > t]} \int_t^{\infty} \frac{E[w(X)|\, X > x]}{\mu^X(x)} exp\left[-\int_t^x \frac{du}{\mu^X(u)}\right] dx$$

Παράδειγμα 3: Η εκθετική κατανομή

Έστω X μια συνεχής τυχαία μεταβλητή η οποία ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda > 0$. Στο παράδειγμα 1 βρήκαμε ότι $\bar{F}_{X^*}(t) = e^{-\lambda t}(\lambda t + 1)$. Άρα αντικαθιστώντας στη σχέση (2.8) έχουμε

$$\mu^{X^*}(t) = \frac{1}{\bar{F}_{X^*}(t)} \int_t^{\infty} \bar{F}_{X^*}(x)dx$$

$$= \frac{1}{e^{-\lambda t}(\lambda t + 1)} \int_t^{\infty} e^{-\lambda x}(\lambda x + 1)dx$$

$$= \frac{1}{e^{-\lambda t}(\lambda t + 1)} \left[\frac{1}{\lambda} e^{-\lambda t}(\lambda t + 1) + \int_t^{\infty} e^{-\lambda x}dx\right]$$

$$= \frac{1}{e^{-\lambda t}(\lambda t + 1)} \left[\frac{1}{\lambda} e^{-\lambda t}(\lambda t + 1) + \frac{1}{\lambda} e^{-\lambda t}\right] = \frac{te^{-\lambda t} + \frac{2}{\lambda} e^{-\lambda t}}{e^{-\lambda t}(\lambda t + 1)} = \frac{t + \frac{2}{\lambda}}{\lambda t + 1}$$

$$= \frac{\lambda t + 2}{\lambda(\lambda t + 1)}$$

Θα επιβεβαιώσουμε το αποτέλεσμα αυτό χρησιμοποιώντας το θεώρημα 17. Προηγουμένως και με βάση το λήμμα 1 βρήκαμε ότι

$$E(X|\, X > t) = t + \frac{1}{\lambda}$$

Άρα από τη σχέση (5.4) έχουμε ότι

$$\mu^{X^*}(t) = \frac{\frac{1}{\lambda}}{t + \frac{1}{\lambda}} \int_t^{\infty} \frac{x + \frac{1}{\lambda}}{\frac{1}{\lambda}} exp\left[-\int_t^{x} \frac{du}{\frac{1}{\lambda}}\right] dx = \frac{1}{\lambda t + 1} \int_t^{\infty} (\lambda x + 1) exp\left[-\int_t^{x} \lambda du\right] dx$$

$$= \frac{1}{\lambda t + 1} \int_t^{\infty} (\lambda x + 1) exp(\lambda t - \lambda x)dx$$

$$= \frac{e^{\lambda t}}{\lambda t + 1} \int_t^{\infty} e^{-\lambda x}(\lambda x + 1)dx = \frac{e^{\lambda t}}{\lambda t + 1}\left(te^{-\lambda t} + \frac{2}{\lambda} e^{-\lambda t}\right) = \frac{t + \frac{2}{\lambda}}{\lambda t + 1}$$

$$= \frac{\lambda t + 2}{\lambda(\lambda t + 1)}$$

Θεώρημα 18 (Jain et al. (1989)) Έστω X και X^w τυχαίες μεταβλητές όπως ορίστηκαν στο Θεώρημα 15 κι έστω $h_{X^w}(t)$ η συνάρτηση κινδύνου της X^w. Τότε η συνάρτηση κινδύνου $h_X(t)$ της X δίνεται από τη σχέση

$$h_X(t) = \frac{h_{X^w}(t)/w(t)}{\int_t^\infty \frac{h_{X^w}(x)}{w(x)} exp\left[-\int_t^x h_{X^w}(u)du\right]dx} \tag{5.5}$$

Επίσης ισχύει

$$E[w(X)|\, X > t] = \frac{exp\left[-\int_0^t \frac{du}{\mu^{X^w}(u)}\right]}{\mu^{X^w}(t)\int_t^\infty \frac{1+\mu'^{X^w}(x)}{w(x)(\mu^{X^w}(x))^2} exp\left[-\int_0^x \frac{du}{\mu^{X^w}(u)}\right]dx} \tag{5.6}$$

Με τα παρακάτω θεωρήματα οι Navarro, del Aguila & Ruiz (2001) εξέφρασαν τη συνάρτηση επιβίωσης της X σαν συνάρτηση του λόγου $\frac{\bar{F}_{X^w}}{\bar{F}_X}$ και του λόγου $\frac{h_X}{h_{X^w}}$.

Θεώρημα 19 Αν $a(x) = \frac{\bar{F}_{X^w(x)}}{\bar{F}_{X(x)}}$ και $k = E[w(X)]$, τότε

$$\bar{F}_X(t) = exp\left\{-\int_{-\infty}^{t} \frac{a'(x)dx}{a(x) - w(x)/k}\right\} \tag{5.7}$$

Θεώρημα 20 Αν η $w(x)$ είναι διαφορίσιμη και $\beta(x) = \frac{h_X(x)}{h_{X^w}(x)}$, τότε

$$\bar{F}_X(t) = exp\left\{-\int_{-\infty}^{t} \frac{\beta'(x)w(x) + \beta(x)w'(x)}{[\beta(x) - 1]w(x)}dx\right\} \tag{5.8}$$

5.2 Διακριτή Περίπτωση

Πριν δώσουμε το θεώρημα που αφορά στη σχέση των συναρτήσεων επιβίωσης των τυχαίων μεταβλητών X και X^w κρίνεται σκόπιμο να παραθέσουμε το παρακάτω λήμμα το οποίο θα μας βοηθήσει στην απόδειξη του θεωρήματος.

Λήμμα 2 (Δημάκη και Ξεκαλάκη (1996)) Έστω X διακριτή τυχαία μεταβλητή με συνάρτηση επιβίωσης $\bar{F}_X(t)$ για όλα τα $t = 0, 1, 2, \ldots$. Έστω επίσης $w(t)$ μια πραγματική συνάρτηση τέτοια ώστε $E[w(X)] < \infty$. Τότε

$$E[w(X)|\, X > t] = w(t+1) + \frac{1}{\bar{F}_X(t)} \sum_{x=t+1}^{\infty} [w(x+1) - w(x)]\bar{F}_X(x) \tag{5.9}$$

Θεώρημα 21 (Δημάκη και Γιακουμάτου (1998)) Έστω X μια μη αρνητική διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας p_t. Έστω επίσης ότι η συνάρτηση στάθμισης $w(t)$ είναι μη αρνητική, γνήσια μονότονη κι έστω ότι η $E[w(X)]$ υπάρχει. Η αντίστοιχη σταθμισμένη τυχαία μεταβλητή X^w έχει συνάρτηση πιθανότητας p_t^w. Τότε η συνάρτηση επιβίωσης $\bar{F}_{X^w}(t)$ μπορεί να εκφραστεί ως εξής:

$$\bar{F}_{X^w}(t) = \frac{1}{E[w(X)]} \bar{F}_X(t) E[w(X)|\, X > t] \tag{5.10}$$

Απόδειξη

$$\begin{aligned}
\bar{F}_{X^w}(t) &= P(X^w > t) \\
&= \sum_{x=t+1}^{\infty} P(X^w = t) = \sum_{x=t+1}^{\infty} \frac{w(x)P(X=x)}{E[w(X)]} \\
&= \frac{1}{E[w(X)]} \sum_{x=t+1}^{\infty} w(x)P(X=x) \\
&= \frac{1}{E[w(X)]} \sum_{x=t+1}^{\infty} w(x)[P(X \geq x) - P(X \geq x+1)] \\
&= \frac{1}{E[w(X)]} \left[\sum_{x=t+1}^{\infty} w(x)P(X \geq x) - \sum_{x=t+1}^{\infty} w(x)P(X \geq x+1) \right] \\
&= \frac{1}{E[w(X)]} \Bigg[w(t+1)P(X \geq t+1) \\
&\quad + \sum_{x=t+1}^{\infty} w(x)P(X \geq x) - \sum_{x=t+1}^{\infty} w(x)P(X \geq x+1) \Bigg] \\
&= \frac{1}{E[w(X)]} \Bigg[w(t+1)P(X > t) \\
&\quad + \sum_{x=t+1}^{\infty} [w(x+1) - w(x)]P(X > x) \Bigg] \\
&= \frac{1}{E[w(X)]} \left[w(t+1)\bar{F}_X(t) + \sum_{x=t+1}^{\infty} [w(x+1) - w(x)]\bar{F}_X(x) \right]
\end{aligned}$$

$$= \frac{\bar{F}_X(t)}{E[w(X)]}\left[w(t+1) + \frac{1}{\bar{F}_X(t)}\sum_{x=t+1}^{\infty}[w(x+1)-w(x)]\bar{F}_X(x)\right]$$
$$= \frac{1}{E[w(X)]}\bar{F}_X(t)E[w(X)|\, X>t]$$

<u>Παράδειγμα 4</u>: Η γεωμετρική κατανομή

Έστω X μια διακριτή τυχαία μεταβλητή η οποία ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p>0$. Στο προηγούμενο κεφάλαιο αποδείχθηκε ότι αν $X\sim G\,(p)$, τότε η X^* έχει συνάρτηση πιθανότητας $f_X^*(t) = tp^2q^{t-1}$. Κατά συνέπεια η συνάρτηση επιβίωσης δίνεται από τη σχέση

$$\bar{F}_{X^*}(t) = q^t(pt+1)$$

Θα επιβεβαιώσουμε το αποτέλεσμα αυτό χρησιμοποιώντας το θεώρημα 21. Με βάση το λήμμα 2

$$E(X|\, X>t) = t+1+q^{-(t+1)}\sum_{x=t+1}^{\infty}q^{x+1} = t+1+\frac{q}{1-q} = \frac{1+pt}{p}$$

Άρα από τη σχέση (5.10) έχουμε ότι

$$\bar{F}_{X^*}(t) = \frac{1}{E[w(X)]}\bar{F}_X(t)E[w(X)|\, X>t] = \frac{p}{q}q^{t+1}\frac{1+pt}{p} = q^t(1+pt)$$

Θεώρημα 22 (Δημάκη & Γιακουμάτου (1998)) Έστω X και X^w τυχαίες μεταβλητές όπως ορίστηκαν στο Θεώρημα 21. Έστω επίσης $h_X(t)$ η συνάρτηση κινδύνου της X. Τότε η συνάρτηση κινδύνου $h_{X^w}(x)$ της σταθμισμένης τυχαίας μεταβλητής X^w μπορεί να εκφραστεί ως εξής:

$$h_{X^w}(t) = \frac{h_X(t)}{\frac{E[w(X)|\, X>t]}{w(t)} + \left[1-\frac{E[w(X)|\, X>t]}{w(t)}\right]h_X(t)} \qquad (5.11)$$

Απόδειξη

$$h_{X^w}(t) = \frac{P(X^w=t)}{P(X^w\geq t)} = \frac{P(X^w=t)}{P(X^w>t)+P(X^w=t)}$$
$$= \frac{\frac{w(t)P(X=t)}{E[w(X)]}}{P(X>t)\frac{E[w(X)|\, X>t]}{E[w(X)]} + \frac{w(t)P(X=t)}{E[w(X)]}}$$

$$= \frac{P(X=t)}{\frac{E[w(X)|\, X>t]}{w(t)}P(X>t)+P(X=t)}$$

$$= \frac{P(X=t)}{\frac{E[w(X)|\, X>t]}{w(t)}[P(X\geq t)-P(X=t)]+P(X=t)}$$

$$= \frac{P(X=t)}{\frac{E[w(X)|\, X>t]}{w(t)}P(X\geq t)+\left[1-\frac{E[w(X)|\, X>t]}{w(t)}\right]P(X=t)}$$

$$= \frac{h_X(t)P(X\geq t)}{\frac{E[w(X)|\, X>t]}{w(t)}P(X\geq t)+\left[1-\frac{E[w(X)|\, X>t]}{w(t)}\right]P(X\geq t)h_X(t)}$$

$$= \frac{h_X(t)}{\frac{E[w(X)|\, X>t]}{w(t)}+\left[1-\frac{E[w(X)|\, X>t]}{w(t)}\right]h_X(t)}$$

<u>Παράδειγμα 5</u>: Η γεωμετρική κατανομή

Έστω X μια διακριτή τυχαία μεταβλητή η οποία ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p>0$. Στο προηγούμενο κεφάλαιο αποδείχθηκε ότι αν $X \sim G\,(p)$, τότε $f_X^*(t)=tp^2q^{t-1}$. Κατά συνέπεια αντικαθιστώντας στη σχέση (2.4) έχουμε ότι

$$h_{X^*}(t)=\frac{P(X^*=t)}{P(X^*>t)+P(X^*=t)}=\frac{tp^2q^{t-1}}{q^t(pt+1)+tp^2q^{t-1}}=\frac{tp^2}{tp+q}$$

Θα επιβεβαιώσουμε το αποτέλεσμα αυτό χρησιμοποιώντας το θεώρημα 22. Προηγουμένως και με βάση το λήμμα 2 βρήκαμε ότι

$$E(X|\, X>t)=\frac{1+pt}{p}$$

Άρα από τη σχέση (5.11) έχουμε ότι

$$h_{X^*}(t)=\frac{h_X(t)}{\frac{E(X|X>t)}{w(t)}+\left[1-\frac{E(X|X>t)}{w(t)}\right]h_X(t)}=\frac{p}{\frac{1+pt}{pt}+\left(1-\frac{1+pt}{pt}\right)p}$$

$$=\frac{p}{\frac{1+pt+p^2t-p-p^2t}{pt}}=\frac{tp^2}{tp+(1-q)}=\frac{tp^2}{tp+q}$$

Θεώρημα 23 (Δημάκη & Γιακουμάτου (1998)) Έστω X και X^w τυχαίες μεταβλητές όπως ορίστηκαν στο Θεώρημα 21. Έστω επίσης $\mu^X(t)$ η μέση υπολειπόμενη ζωή της X. Τότε η μέση υπολειπόμενη ζωή $\mu^{X^w}(t)$ της σταθμισμένης τυχαίας μεταβλητής X^w μπορεί να εκφραστεί ως εξής:

$$\mu^{X^w}(t) = \frac{1}{\bar{F}_X(t)E[w(X)|\, X > t]}\sum_{x=t}^{\infty}\bar{F}_X(x)E[w(X)|\, X > x] \tag{5.12}$$

Απόδειξη

$$\mu^{X^w}(t) = \frac{1}{\bar{F}_{X^w}(t)}\sum_{x=t}^{\infty}\bar{F}_{X^w}(x)$$

$$= \frac{1}{\frac{1}{E[w(X)]}\bar{F}_X(t)E[w(X)|\, X > t]}\sum_{x=t}^{\infty}\frac{1}{E[w(X)]}\bar{F}_X(x)E[w(X)|\, X > x]$$

$$= \frac{1}{\bar{F}_X(t)E[w(X)|\, X > t]}\sum_{x=t}^{\infty}\bar{F}_X(x)E[w(X)|\, X > x]$$

<u>Παράδειγμα 6</u>: Η γεωμετρική κατανομή

Έστω X μια διακριτή τυχαία μεταβλητή η οποία ακολουθεί τη γεωμετρική κατανομή με παράμετρο $p > 0$. Στο προηγούμενο κεφάλαιο αποδείχθηκε ότι αν $X \sim G\,(p)$, τότε $f_X^*(t) = tp^2q^{t-1}$ και άρα με άμεσους υπολογισμούς $\bar{F}_{X^*}(t) = q^t(pt+1)$. Κατά συνέπεια αντικαθιστώντας στη σχέση (2.9) έχουμε

$$\mu^{X^*}(t) = \frac{1}{\bar{F}_{X^*}(t)}\sum_{x=t}^{\infty}\bar{F}_{X^*}(x) = \frac{1}{q^t(pt+1)}\sum_{x=t}^{\infty}q^x(px+1) = \frac{pt+q+1}{p(pt+1)}$$

Θα επιβεβαιώσουμε το αποτέλεσμα αυτό χρησιμοποιώντας το θεώρημα 23. Προηγουμένως και με βάση το λήμμα 2 βρήκαμε ότι

$$E(X|\, X > t) = \frac{1+pt}{p}$$

Άρα από τη σχέση (5.12) έχουμε ότι

$$\mu^{X^*}(t) = \frac{1}{\bar{F}_X(t)E(X|\, X > t)}\sum_{x=t}^{\infty}\bar{F}_X(x)E(X|\, X > t)$$

$$= \frac{1}{q^{t+1}\left(\frac{pt+1}{p}\right)}\sum_{x=t}^{\infty}q^{x+1}\left(\frac{px+1}{p}\right)$$

$$= \frac{1}{q^t(pt+1)} \sum_{x=t}^{\infty} (q^x px + q^x) = \frac{1}{pt+1} \left(\frac{pt+q+1}{p} \right) = \frac{pt+q+1}{p(pt+1)}$$

5.3 Ανισοτικές Σχέσεις – Κριτήρια Γήρανσης

Σε αυτή την τελευταία υποενότητα του κεφαλαίου παραθέτουμε κάποιες ανισοτικές σχέσεις μεταξύ των μέτρων αξιοπιστίας αρχικής και σταθμισμένης κατανομής αλλά και μεταξύ των τυχαίων μεταβλητών. Πριν όμως προχωρήσουμε στα θεωρήματα είναι απαραίτητο να δώσουμε τους παρακάτω ορισμούς.

Ορισμός 14 Έστω X και Y μη αρνητικές συνεχείς τυχαίες μεταβλητές με συναρτήσεις πυκνότητας πιθανότητας f και g αντίστοιχα. Λέμε ότι η X είναι μεγαλύτερη από την Y στο λόγο πιθανοφάνειας και γράφουμε $X \geq_{LR} Y$ αν η $\frac{f_X(x)}{g_Y(x)}$ είναι αύξουσα συνάρτηση του x.

Ορισμός 15 Έστω X και Y μη αρνητικές συνεχείς τυχαίες μεταβλητές με συναρτήσεις πυκνότητας πιθανότητας f και g αντίστοιχα. Λέμε ότι η X είναι μεγαλύτερη από την Y στο βαθμό κινδύνου και γράφουμε $X \geq_{HR} Y$ αν $h_X(x) \leq h_Y(x)$ για κάθε x.

Θεώρημα 24 Αν η $w(x)$ είναι αύξουσα συνάρτηση του x τότε $X^w \geq_{LR} X$, $\bar{F}_{X^w}(x) \geq \bar{F}_X(x)$ και $h_{X^w}(x) \leq h_X(x)$ για κάθε $x \geq 0$. Όλες οι προηγούμενες ανισότητες αλλάζουν φορά αν η $w(x)$ είναι φθίνουσα συνάρτηση του x.

Στο θεώρημα 24 η μονοτονία της $w(x)$ είναι ικανή αλλά όχι αναγκαία συνθήκη. Μάλιστα, ισχύουν τα παρακάτω:

1. $h_{X^w}(x) \leq h_X(x) \Leftrightarrow \frac{\bar{F}_{X^w}(x)}{\bar{F}_X(x)}$ είναι αύξουσα $\Leftrightarrow \frac{E[w(X)|\, X>x]}{E[w(X)]}$ είναι αύξουσα.
2. $\bar{F}_{X^w}(x) \geq \bar{F}_X(x) \Leftrightarrow \int_0^x h_{X^w}(u)du \leq \int_0^x h_X(u)du$.
3. $\mu^{X^w}(x) \geq \mu^X(x) \Leftrightarrow \frac{\int_x^\infty \bar{F}_{X^w}(u)du}{\int_x^\infty \bar{F}_X(u)du}$ είναι αύξουσα.

Θεώρημα 25

1. $X \geq_{LR} Y$ αν και μόνο αν $X^w \geq_{LR} Y^w$.
2. Αν $X \geq_{HR} Y$, τότε $X^w \geq_{HR} Y^w$ για κάθε αύξουσα συνάρτηση $w(x)$.
3. Αν $X \geq_{HR} Y$ και $\frac{h_X(x)}{h_Y(x)}$ αύξουσα, τότε $X^w \geq_{LR} Y^w$.

Θεώρημα 26 Έστω $w(x)$ αύξουσα, κοίλη (φθίνουσα, κυρτή) και διαφορίσιμη συνάρτηση. Ισχύουν τα παρακάτω:

1. Αν η F είναι IFR (DFR), τότε η F^w είναι IFR (DFR).
2. Αν η F είναι DMRL (IMRL), τότε η $E[w(X) - w(x)| \, X > x]$ είναι φθίνουσα (αύξουσα).
3. Αν η F είναι DMRL (IMRL), τότε $\frac{h_{X^w}(x)}{h_X(x)}$ είναι αύξουσα (φθίνουσα).
4. Αν η F είναι IFR και $\bar{F}_X(x)w(x)$ αύξουσα, τότε η F^w είναι IFR.

ΚΕΦΑΛΑΙΟ 6
ΕΠΙΛΟΓΟΣ

Σκοπός της εργασίας αυτής ήταν η μελέτη των σταθμισμένων κατανομών υπό το πρίσμα της θεωρίας αξιοπιστίας. Στο πλαίσιο αυτό, η μελέτη επικεντρώθηκε στις σημαντικότερες μονοδιάστατες κατανομές χρόνου ζωής. Θα πρέπει όμως να τονίσουμε ότι ενδιαφέροντα συμπεράσματα μπορούν να προκύψουν για πλήθος άλλων κατανομών (μονοδιάστατων ή δισδιάστατων) που δεν εμπίπτουν στο πεδίο της ανάλυσης επιβίωσης. Επιπρόσθετα, ιδιαίτερη προσοχή θα πρέπει να δοθεί στο μέλλον στις σχέσεις για τα μέτρα αξιοπιστίας μεταξύ αρχικής και σταθμισμένης κατανομής στη δισδιάστατη περίπτωση κατανομών.

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] **Δημάκη, Κ. (2008).** *Ανάλυση Επιβίωσης*. Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών, Αθήνα.

ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] **Barlow, R. E. and Proschan, F. (1975).** *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston Inc., New York.

[2] **Bartoszewicz, J. and Skolimowska, M. (2006).** Preservations of Classes of Life Distributions and Stochastic Orders under Weighting, *Statistics & Probability Letters*, 76, 587-596.

[3] **Bryson, M. C. and Siddiqui, M. M. (1969).** Some Criteria of Ageing, *Journal of the American Statistical Association*, 64, 1472-1483.

[4] **Dimaki, C. and Giakoumatou, M. (1998).** Reliability Measures for Discrete Weighted Distributions, *Technical Report No 44*, Dept. of Statistics, AUEB.

[5] **Giakoumatou, M. (1998).** *A Study of Weighted Distributions: Properties, Characterizations and Reliability Measures*. Thesis (MSc). AUEB.

[6] **Gupta, R. C. and Keating, J. P. (1986).** Relations for Reliability Measures under Length-Biased Sampling, *Scandinavian Journal of Statistics*, 13, 49-56.

[7] **Jain, K., Singh, H. and Bagai, I. (1989).** Relations for Reliability Measures of Weighted Distributions, *Communications in Statistics-Theory and Methods*, 18(12), 4393-4412.

[8] **Kirmani, S. N. U. A. and Ahsanullah, M. (1987).** A Note on Weighted Distributions, *Communications in Statistics-Theory and Methods*, 16(1), 275-280.

[9] **Kochar, S. C. and Gupta, R. P. (1987).** Some Results on Weighted Distributions for Positive-Valued Random Variables, *Probability in the Engineering and Informational Sciences*, 1, 417-423.

[10] **Lai, C. D. and Xie, M. (2006).** *Stochastic Ageing and Dependence for Reliability*. Springer, New York.

[11] **Lee, E. T. (1992*).*** *Statistical Methods for Survival Data*. Wiley & Sons Inc., New York.

[12] **Mahfoud, M. and Patil, G. P. (1982).** On Weighted Ditributions, *In Statistics and Probability: Essays in Honor of C. R. Rao,* G. Kallianpur, P. R. Krishnaiah, J. K. Ghosh, (eds.), Amsterdam: North Holland, 479-492.

[13**] Nanda, A. K. and Jain, K. (1999).** Some weighted distribution Results on Univariate and Bivariate Cases, *Journal of Statistical Planning and Inference*, 77, 169-180.

[14] **Navarro, J., Del Aguila, Y. and Ruiz, J. M. (2001).** Characterizations Through Reliability Measures from Weighted Distributions, *Statistical Papers*, 42, 395-402.

[15] **Oluyede, B. O. and George, E. O. (2002).** On Stochastic Inequalities and Comparisons of Reliability Measures for Weighted Distributions, *Mathematical Problems in Engineering,* Vol. 8, pp. 1-13.

[16] **Patil, G. P. and Ord, J. K. (1976).** On Size-Biased Sampling and Related Form-Invariant Weighted Distributions, *Sankya: The Indian Journal of Statistics-Series B*, 38, 48-61.

[17] **Patil, G. P. and Rao, C. R. (1978).** Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families, *Biometrics*, 34, 179-189.

[18] **Patil, G. P. (1991).** Encountered Data, Statistical Ecology, Enviromental Statistics and Weighted Distribution Methods, *Environmetrics*, 2(4), 377-423.