ΧΑΡΑΚΤΗΡΙΣΜΟΙ ΚΑΤΑΝΟΜΩΝ
ΜΕ ΕΜΦΑΣΗ ΣΤΙΣ ΚΑΤΑΝΟΜΕΣ
PARETO ΚΑΙ YULE

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΚΑΤΕΡΙΝΑ ΔΗΜΑΚΗ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΑΘΗΝΑ 1991

Θα ήθελα να αφιερώσω μερικές γραμμές για να ευχαριστήσω ορισμένα πρόσωπα των οποίων η βοήθεια στα δύσκολα αυτά για μένα χρόνια υπήρξε πολύτιμη και πήγη δύναμης.

Καταρχήν οι ευχαριστίες μου απευθύνονται στην καθηγήτρια κ. *Ευδοκία Ξεκαλάκη* τόσο για την ουσιαστική επιστημονική συμβολή της όσο και για την αμέριστη ηθική της συμπαράσταση. Χωρίς τη βοήθεια και στήριξή της το έργο μου ίσως ήταν ακατόρθωτο.

Επίσης σημαντική θεωρώ τη συμπαράσταση που μου προσέφεραν ο καθηγητής κ. *Παναγιώτης Τζωρτζόπουλος* και ο επίκουρος καθηγητής κ. *Μιχάλης Μαγδαληνός*, μέλη της συμβουλευτικής επιτροπής.

Ακόμη θα ήθελα να ευχαριστήσω τις συναδέλφους μου Δρ. *Ελένη Καυδηλώρου* , *Δήμητρα Δημάκου* και *Μαρία Δημοπούλου* για τη βοήθεια τους στη συγκέντρωση της βιβλιογραφίας. Για τον ίδιο λόγο ευχαριστώ και τον πτυχιούχο του Πανεπιστημίου μας και υποψήφιο Διδάκτορα του Πανεπιστημίου Illinois κ. *Γιάννη Μπίλια*.

Ιδιαίτερες ευχαριστίες οφείλω στην αδελφή μου *Τζούλια* και στον πολύ καλό μου φίλο *Λευτέρη Μπεληγιάννη* που όλα αυτά τα χρόνια αφιέρωσαν ένα μεγάλο μέρος από τον πολύτιμο χρόνο τους για να με απαλλάξουν από άχαρες καθημερινές εργασίες ώστε να αφοσιωθώ περισσότερο στη διατριβή μου.

Τέλος ένα μεγάλο ευχαριστώ στον άνδρα μου *Βασίλη* και στα γλυκά μου παιδάκια *Θάνο* και *Δημήτρη* για την ηθική τους συμπαράσταση και την μεγάλη τους υπομονή .

Π Ε Ρ Ι Ε Χ Ο Μ Ε Ν Α

ΠΡΟΛΟΓΟΣ 2

ΚΕΦΑΛΑΙΟ 1 3

Γενική Εισαγωγή

1.1 Κατανομές Pareto και Yule. Σύντομη Ανασκόπηση 4

1.2 Γενικά περί χαρακτηρισμών 8

1.3 Χαρακτηρισμοί της κατανομής Pareto βασισμένοι στις διατεταγμένες στατιστικές συναρτήσεις 15

1.4 Χαρακτηρισμοί της κατανομής Pareto μέσω της γραμμικής λογαριθμο-εκθετικής οικογένειας 24

ΚΕΦΑΛΑΙΟ 2 27

Διδιάστατες Κατανομές Pareto και Yule

2.1 Εισαγωγή 28

2.2 Διδιάστατες Κατανομές Pareto τύπου I 29

2.3 Διδιάστατες Κατανομές Pareto τύπου II 32

2.3.1 Διδιάστατη κατανομή Pareto τύπου Gumbel 33

2.3.2 Διδιάστατη κατανομή Pareto τύπου Marshall-Olkin 35

2.3.3 Διδιάστατη κατανομή Pareto τύπου Mardia II 37

2.4 Διδιάστατες Κατανομές Pareto τύπου III 38

2.5 Γενικευμένες Μορφές Διδιάστατων Κατανομών Pareto 49

2.6 Μίγματα Διδιάστατων Κατανομών Pareto 51

2.7 Αλλες Διδιάστατες Κατανομές συνδεόμενες με τις Διδιάστατες κατανομές Pareto 55

2.8 Διδιάστατες Κατανομές Yule 60

ΚΕΦΑΛΑΙΟ 3

Χαρακτηριστικές ιδιότητες βασισμένες στην υπό συνθήκη κατανομή ή/και στην υπό συνθήκη αναμενόμενη τιμή 65

3.1 Εισαγωγή 66

3.2 Χαρακτηριστικές ιδιότητες βασισμένες στην υπό συνθήκη κατανομή 67

3.3 Χαρακτηριστικές ιδιότητες βασισμένες στην υπό συνθήκη κατανομή και στην υπό συνθήκη αναμενόμενη τιμή 75

3.4 Χαρακτηριστικές ιδιότητες βασισμένες στην υπό συνθήκη αναμενόμενη τιμή. Συνεχής Περίπτωση 91

3.5 Χαρακτηριστικές ιδιότητες βασισμένες στην υπό συνθήκη αναμενόμενη τιμή. Διακριτή Περίπτωση 102

ΚΕΦΑΛΑΙΟ 4

Χαρακτηριστικές ιδιότητες Διδιάστατων Κατανομών Pareto κάι Yule 121

4.1 Εισαγωγή 122

4.2 Ιδιότητες της Διδιάστατης κατανομής Pareto τύπου Marshall-Olkin 123

4.3 Ιδιότητες της Διδιάστατης κατανομής Pareto τύπου Mardia I 131

4.4 Ιδιότητες της Διδιάστατης Κατανομής Yule 136

ΚΕΦΑΛΑΙΟ 5

Οι κατανομές Pareto και Yule στο πλαίσιο της θεωρίας αξιοπιστίας 140

5.1 Εισαγωγή 141

5.2 Στοιχεία από τη θεωρία Αξιοπιστίας 144

5.3 Οι μονοδιάστατες Κατανομές Pareto και Yule στο πλαίσιο της θεωρίας αξιοπιστίας 153

5.4 Οι δισδιάστατες κατανομές Pareto και Yule στο πλαίσιο της θεωρίας αξιοπιστίας 170

ΚΕΦΑΛΑΙΟ 6

Μονοδιάστατα Συστήματα Κατανομών 182

6.1 Εισαγωγή 183

6.2 Μονοδιάστατες Σταθμισμένες Κατανομές 183

6.3 Μονοδιάστατα Συστήματα Κατανομών 199

6.3.1 Α' Περίπτωση: Διακριτή Γεννήτωρ Κατανομή 199

6.3.2 Β' Περίπτωση: Συνεχής Γεννήτωρ Κατανομή 215

ΚΕΦΑΛΑΙΟ 7

Διδιάστατα Συστήματα Κατανομών 229

7.1 Εισαγωγή 230

7.2 Διδιάστατες Σταθμισμένες Κατανομές 230

7.3 Διδιάστατα Συστήματα Κατανομών 233

7.3.1 Α' Περίπτωση: Διακριτή Διδιάστατη Γεννήτωρ Κατανομή 234

7.3.1 Α' Περίπτωση: Συνεχής Διδιάστατη Γεννήτωρ Κατανομή 248

ΚΕΦΑΛΑΙΟ 8

Πρότυπο Διαστρέβλωσης 262

8.1 Εισαγωγή 263

8.2 Χαρακτηριστικές ιδιότητες στο πλαίσιο του Underreporting: Συνεχής Περίπτωση 268

8.3 Χαρακτηριστικές ιδιότητες στο πλαίσιο του Overreporting Συνεχής Περίπτωση 285

8.4 Χαρακτηριστικές ιδιότητες στο πλαίσιο του Underreporting: Διακριτή Περίπτωση 294

8.3 Χαρακτηριστικές ιδιότητες στο πλαίσιο του Overreporting Διακριτή Περίπτωση 306

ΕΠΙΛΟΓΟΣ 312

ΒΙΒΛΙΟΓΡΑΦΙΑ 313

ABSTRACT 339

Κ Ε Φ Α Λ Α Ι Ο 1

ΓΕΝΙΚΗ ΕΙΣΑΓΩΓΗ

1.1 ΚΑΤΑΝΟΜΕΣ PARETO ΚΑΙ YULE. ΣΥΝΤΟΜΗ ΑΝΑΣΚΟΠΗΣΗ

Είναι γνωστό ότι οι κατανομές πολλών κοινωνικών και οικονομικών μεγεθών εμφανίζουν έντονα ασύμμετρες μορφές. Κατα καιρούς έχουν προταθεί διάφορες κατανομές για την περιγραφή των έντονα ασύμμετρων αυτών μορφών. Μεταξύ αυτών σημαντικότερες είναι οι κατανομές Pareto και Yule.

Η κατανομή Pareto πήρε το όνομά της από τον Ελβετό καθηγητή των Οικονομικών Vilfredo Pareto ο οποίος διαπίστωσε ότι ο λογάριθμος του ποσοστού των ατόμων που έχουν εισόδημα μεγαλύτερο ή ίσο από κάποια τιμή x είναι μιας αρνητικής κλίσης γραμμική συνάρτηση του λογαρίθμου αυτής της τιμής. Η πιο γνωστή εφαρμογή της κατανομής Pareto είναι στη μελέτη των εισοδημάτων φυσικών προσώπων. Χρησιμοποιείται επίσης για τη περιγραφή των διακυμανσέων στις τιμές των μετοχών, της δυναμικότητας των κοιτασμάτων πρώτων υλών, της συσσώρευσης λαθών σε δίκτυα επικοινωνίας, του μεγέθους βιομηχανικών εγκαταστάσεων, του πληθυσμού πόλεων και σε προβλήματα αξιοπιστίας. Τέλος η Xekalaki (1988a) στα πλαίσια του μοντέλου ανανέωσης αποθέματος και κάτω από την υπόθεση ότι τόσο ο χρόνος αναμονής μιας παραγγελίας όσο και το απόθεμα είναι τυχαίες μεταβλητές και ότι η ζήτηση για το προϊόν ακολουθεί την κατανομή Pareto, προσδιορίζει την "κρίσιμη" τιμή του αποθέματος που είναι ενδεικτική της ανάγκης ανανέωσης. Μια αναλυτική καταγραφή των εφαρμογών της κατανομής Pareto στη μελέτη των εισοδημάτων έχει γίνει μεταξύ άλλων και από τον Arnold (1982).

Στα επόμενα ο όρος κατανομή Pareto θα αναφέρεται στην κατανομή με συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \begin{cases} \alpha\,\theta^{\alpha}\,x^{-(\alpha+1)} & x \geq \theta,\ \theta>0,\ \alpha>0. \\ 0 & \text{οπουδήποτε αλλού.} \end{cases} \qquad (1.1)$$

Η κατανομή Pareto εμφανίζεται συχνά στη βιβλιογραφία με μια ακόμη παράμετρο την μ που είναι παράμετρος θέσης. Ονομάζουμε την κατανομή αυτή Pareto II και η συνάρτηση πυκνότητας πιθανότητάς της δίνεται από την

$$f(x) = \begin{cases} \frac{\alpha}{\sigma}\left[1+\left(\frac{x-\mu}{\sigma}\right)\right]^{-\alpha} & x>\mu,\ \alpha,\sigma>0 \\ 0 & \text{οπουδήποτε αλλού.} \end{cases} \qquad (1.2)$$

Ας σημειωθεί ότι ο Arnold (1982) ορίζει μια ιεραρχία *(hierarchy)* κατανομών Pareto που προοδευτικά γίνονται πιο σύνθετες. Ξεκινάει με την κλασσική κατανομή Pareto και, εισάγοντας διαδοχικά παραμέτρους θέσης *(location)*, κλίμακας *(scale)*, συντελεστή Gini *(Gini index)* και μορφής *(shape)* οδηγείται σε μια οικογένεια την οποία ονομάζει οικογένεια κατανομών Pareto IV (ουσιαστικά τις κατανομές Burr).

Η εξέλιξη και βελτίωση του μοντέλου Pareto έστρεψε την προσοχή σε επιθυμητές ιδιότητες των μέτρων που αναφέρονται σε εισοδηματικές ανισότητες. Μερικά από τα μέτρα αυτά προυπέθεταν έμμεσα ότι τα ατομικά εισοδήματα ακολουθούν την κατανομή Pareto, άλλα ήταν γενικότερα μέτρα διασποράς. Στη σχετική βιβλιογραφία η καμπύλη Lorenz και ο συντελεστής Gini είναι εκείνα που κυριαρχούν. Αρκετές γενιές Ιταλών στατιστικών ακολουθώντας τα βήματα του Gini εργάστηκαν για τη βελτίωση των μέτρων που αναφέρονται στις

εισοδηματικές ανισότητες. Στο σημείο αυτό αξίζει να παρατηρηθεί ότι υπάρχει φυσική ερμηνεία της παραμέτρου α στο πλαίσιο της κατανομής των εισοδημάτων και αυτό γιατί η παράμετρος α συνδέται μονότονα με το συντελεστή του Gini $\left(g = 1/(2\alpha-1)\right)$ και κατά συνέπεια μπορεί να χρησιμοποιηθεί ως μέτρο ανισότητας των τιμών της κατανομής. Επίσης το θ αντιπροσωπεύει το ελάχιστο ύψος εισοδήματος.

Στη δεκαετία του 60 εμφανίζονται στη βιβλιογραφία οι διδιάστατες κατανομές Pareto (Mardia, 1962, 1964). Οι κατανομές αυτές είναι χρήσιμες για την περιγραφή μεταξύ άλλων και της από κοινού κατανομής του εισοδήματος ενός ατόμου σε διαφορετικές χρονικές στιγμές ή στην από κοινού κατανομή του εισοδήματος ατόμων που συνδέονται με κάποια σχέση (αντρόγυνα). Εγινε επίσης προσπάθεια από τους Blitz and Britain (1964) και αργότερα τον Taguchi (1972a,b , 1973) προς την κατεύθυνση των πολυμεταβλητών καμπύλων συγκέντρωσης. Πάντως η ανάπτυξη της θεωρίας στο θέμα αυτό παραμένει ιδιαίτερα ρηχή. Να σημειωθεί επίσης ότι μικρή είναι η προσπάθεια που έχει γίνει για τον ορισμό μοντέλου το οποίο να οδηγεί σε πολυμεταβλητές κατανομές τύπου Pareto.

Ο Simon (1955) επεχείρησε μια ενοποιημένη ερμηνεία ορισμένων έντονα ασύμμετρων κατανομών συχνοτήτων οι οποίες περιγράφουν κοινωνικά, βιολογικά και οικονομικά φαινόμενα. Συγκεκριμένα θεώρησε την κατανομή των λέξεων σε κείμενα με βάση τη συχνότητα εμφάνισης τους, την κατανομή των πόλεων με βάση τον πληθυσμό τους, την κατανομή των βιολογικών γενών *(biological genera)* με βάση των αριθμό των ειδών *(species)*, την κατανομή των επιστημόνων με βάση τις δημοσιεύσεις τους και την κατανομή των εισοδημάτων με βάση το ύψος τους. Οι κατανομές αυτές ορίζονται στους θετικούς ακεραίους

και ονομάζονται κατανομές Yule.

Στα επόμενα ο όρος κατανομή Yule θα αναφέρεται στην κατανομή με συνάρτηση πιθανότητας:

$$p_x = \frac{\rho\, x!}{(\rho+2)_{(x)}}, \quad \rho>0,\ x = 1,2,... \qquad (1.3)$$

$$\text{όπου } \alpha_{(\beta)} = \Gamma(\alpha+\beta)/\Gamma(\alpha), \quad \alpha,\beta \in \mathbb{R}^+.$$

Η κατανομή Yule έχει εφαρμογές στο χώρο της βιολογίας, οικονομίας, βιβλιογραφικής έρευνας (Xekalaki (1981,1984b)). Επιπλέον στο πλαίσιο του μοντέλου ανανέωσης αποθέματος η κατανομή Yule προκύπτει ως η κατανομή ζήτησης (Xekalaki (1988b)). Τέλος να σημειωθεί ότι διδιάστατες επεκτάσεις της κατανομής Yule ορίζονται και μελετώνται από την Xekalaki (1986).

Ο Irwin (1975) κατέληξε ότι το διακριτό ανάλογο της γενικευμένης κατανομης Waring είναι στη γενική περίπτωση τύπου Pearson VI. Ξεκινώντας από το ότι η μονοδιάστατη γενικευμένη κατανομή Waring έχει ως ειδική περίπτωση την κατανομή Yule και δεδομένου ότι η κατανομή Pareto μπορεί να θεωρηθεί ως ειδική περίπτωση της οικογένειας Pearson VI καταλήγουμε στο συμπέρασμα ότι η κατανομή Pareto είναι το συνεχές ανάλογο της κατανομής Yule. Επειδή λοιπόν η κατανομή Yule είναι το διακριτό ανάλογο της κατανομής Pareto (Kendall (1961)) οποτεδήποτε συνεχή δεδομένα για διάφορα οικονομικά μεγέθη μετατρέπονται σε διακριτά είτε λόγω ομαδοποιήσεων είτε λόγω έλλειψης ακρίβειας των μετρήσεων η κατανομή Yule μπορεί να χρησιμοποιηθεί αντί της Pareto.

1.2 ΓΕΝΙΚΑ ΠΕΡΙ ΧΑΡΑΚΤΗΡΙΣΜΩΝ

Το κύριο αντικείμενο της μελέτης αυτής είναι η διατύπωση και απόδειξη χαρακτηριστικών ιδιοτήτων κατανομών οι οποίες χρησιμοποιούνται στην ανάλυση και μελέτη κυρίως οικονομικών φαινομένων. Οι χαρακτηρισμοί κατανομών τοποθετούνται στη διαχωριστική γραμμή που υπάρχει μεταξύ θεωρίας πιθανοτήτων και μαθηματικής στατιστικής και πέρα από το μαθηματικό τους ενδιαφέρον, είναι χρήσιμοι και στην εφαρμοσμένη στατιστική. Ο κύριος λόγος βρίσκεται στο γεγονός ότι αυτές οι ιδιότητες είναι μοναδικές για τη χαρακτηριζόμενη κατανομή. Το γεγονός αυτό μπορεί να καθοδηγήσει κάποιον στην επιλογή των υποθέσεων που πρέπει να επιβάλλει σε ένα συγκεκριμένο πρόβλημα ή να τον βοηθήσει να μετατρέψει ένα πολύπλοκο πρόβλημα σε άλλο, ισοδύναμο μεν, αλλά πιθανότατα απλούστερο. Για τις αποδείξεις των θεωρημάτων χρησιμοποιούνται τα κλασσικά εργαλεία της μαθηματικής ανάλυσης και η θεωρία των συναρτησιακών εξισώσεων. Θα ξεκινήσουμε αυτή την εισαγωγική στο θέμα ενότητα περιγράφοντας το πρόβλημα του χαρακτηρισμού μιας στατιστικής κατανομής.

Εστω η τυχαία μεταβλητή X και έστω $X_1, X_2, \ldots, X_n$ η ανεξάρτητες παρατηρήσεις πάνω σ' αυτήν. Εκείνο το οποίο συνήθως επιδιώκουμε είναι η εξαγωγή συμπερασμάτων σχετικά με την κατανομή $F_X(x)$ της X. Η συνηθισμένη προσέγγιση στο πρόβλημα είναι να ξεκινήσουμε με μια οικογένεια κατανομών, έστω F, και να διαλέξουμε από αυτή την οικογένεια μια κατανομή $F(x)$ η οποία να είναι η πιο αποδεκτή με βάση καθορισμένα κριτήρια. Δυστυχώς όμως στις περισσότερες περιπτώσεις η οικογένεια F αποτελείται από μια μόνο συνάρτηση που εξαρτάται από μια ή περισσότερες παραμέτρους και οι

παρατηρήσεις τις οποίες έχουμε χρησιμοποιούνται για την εκτίμηση των παραμέτρων αυτής. Η συνάρτηση που με αυτό τον τρόπο προσδιορίζεται θεωρούμε ότι είναι η ζητούμενη συνάρτηση κατανομής $F_X(x)$. Οι χαρακτηριστικές ιδιότητες αποτελούν τη μοναδική μέθοδο για το μη υποκειμενικό καθορισμό της οικογένειας F και οδηγούν στην ακριβή μορφή της $F_X(x)$ μέσω απλών ιδιοτήτων. Για να δείξουμε πόσο σημαντικός είναι ο ακριβής προσδιορισμός της κατανομής του πληθυσμού παραθέτουμε ένα παράδειγμα από το βιβλίο του J. Galambos (1978).

Εστω Y τυπική λογαριθμοκανονική μεταβλητή και έστω $X = (Y^t-1)/t$, $t>0$. Είναι γνωστό ότι αν το t είναι πολύ μικρό η κατανομή της X προσεγγίζει την τυπική κανονική κατανομή. Ας συγκρίνουμε λοιπόν τις τιμές H(2.6) της συνάρτησης κατανομής H(x) του μεγίστου 50 παρατηρήσεων της X κάτω από δυο διαφορετικές υποθέσεις: η X είναι κανονική και $X = (Y^{0.1}-1)/0.1$. Εύκολα βρίσκουμε ότι

H(2.6) = 0.59 όταν υποθέσουμε κανονικότητα

H(2.6) = 0.78 όταν υποθέσουμε λογαριθμοκανονικότητα.

Είναι σαφές από την ανάλυση που κάνει ο Kotz (1973) ότι ο έλεγχος καλής προσαρμογής δέχεται και τις δυο υποθέσεις αναφορικά με την κατανομή της τυχαίας μεταβλητής X. Είναι προφανές ότι η απόφαση θα έπρεπε να ήταν διαφορετική σε καθεμιά απο τις δυο περιπτώσεις. Προφανώς η κλασσική στατιστική θεωρία δεν είναι σε θέση να αντιμετωπίσει αυτή την αντίφαση και αυτό γιατί είναι γνωστό ότι ο έλεγχος καλής προσαρμογής δεν μπορεί να διακρίνει δυο κατανομές στην περίπτωση που προσεγγίζουν πολύ η μία την άλλη. Εν τούτοις αν

υπήρχε αποδεκτή χαρακτηριστική ιδιότητα θα μπορούσαμε να οδηγηθούμε σε μια και μοναδική ξεκάθαρη επιλογή.

Στη συνέχεια παραθέτουμε μερικά παραδείγματα τα οποία δίνουν μια ιδέα της πρακτικής σημασίας και αξίας των χαρακτηριστικών ιδιοτήτων.

Παράδειγμα 1 Εστω ότι ένα κατάστημα εξυπηρετεί μια ομάδα n ατόμων. Τα άτομα αυτά επισκέπτονται το κατάστημα ανεξάρτητα το ένα από το άλλο και έστω ότι οι χρόνοι άφιξης στο κατάστημα ακολουθούν την ίδια κατανομή. Επομένως, τα n άτομα μπορούν να συνδεθούν με n ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την ίδια κατανομή και έστω X_j, $1 \le j \le n$, οι τυχαίοι χρόνοι άφιξής τους στο κατάστημα. Ο καταστηματάρχης παρατηρεί τα X_j προφανώς σε αύξουσα τάξη. Εστω $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ οι διαδοχικές αφίξεις στο κατάστημα. Υποθέτουμε ακόμη ότι ο καταστηματάρχης παρατηρεί ότι τα μεσοδιαστήματα $X_{1:n}, X_{2:n}-X_{1:n}, \ldots, X_{n-1:n}-X_{n:n}$ είναι επίσης ανεξάρτητα. Εχει αποδειχθεί (Fisz (1958), Rossberg (1972), Govindarajulu (1966)) ότι στην περίπτωση που περιγράψαμε η κοινή κατανομή $F(x)$ των X_j είναι η εκθετική. Αξίζει να σημειώσουμε ότι μέσω της χαρακτηριστικής ιδιότητας φθάσαμε σε μια και μοναδική δυνατή μορφή για την $F(x)$. Αυτό σημαίνει ότι οι συμβουλές που θα δοθούν στα καταστήματα τα οποία επιδιώκουν καλυτέρευση της εξυπηρέτησης των πελατών τους και που σχετίζονται με ενδεχόμενη αύξηση του αριθμού των υπαλλήλων τους ή ενέργειες για να υπάρχει επάρκεια αγαθών κ.λ.π. θα βασίζονται σε ένα και μοναδικό, καλά ορισμένο μοντέλο.

Παράδειγμα 2 Ασφαλιστική εταιρεία θέλει να συγκεντρώσει στοιχεία σχετικά με τα ατυχήματα που έχουν συμβεί σε οδηγούς. Η εταιρεία χαρακτηρίζει έναν οδηγό "καλό" αν αυτός έχει την παρακάτω ιδιότητα: η πιθανότητα να έχει ένα δυστύχημα παραμένει περίπου η ίδια με το πέρασμα του χρόνου. Η ιδιότητα αυτή σε μαθηματικούς όρους μεταφέρεται ως εξής: Εστω X το χρονικό διάστημα μέχρι το πρώτο δυστύχημα του συγκεκριμένου οδηγού. Τότε,

$$P(X \geq s+t | X \geq t) = P(X \geq s).$$

Η παραπάνω σχέση δεν είναι άλλη από τη μαθηματική έκφραση της έλλειψης αθροιστικής μνήμης η οποία όπως είναι γνωστό χαρακτηρίζει την εκθετική κατανομή. Η ασφαλιστική εταιρεία επομένως δεν χρειάζεται να ριψοκινδυνεύει προσπαθώντας να μαντέψει, αλλά μπορεί να προσδιορίσει το ύψος του ασφάλιστρου κάνοντας χρήση ενός καλά ορισμένου μοντέλου.

Παράδειγμα 3 Η περίπτωση την οποία θα περιγράψουμε αποτελεί ίσως τον πρώτο δημοσιευμένο χαρακτηρισμό της κατανομής Pareto. Ο Hagstroem (1925) θεώρησε το μέσο εισόδημα όλων των ατόμων των οποίων το εισόδημα υπερέβαινε μια τιμή έστω t (σε δολλάρια). Εστω τώρα ότι λόγω πληθωρισμού το εισόδημα καθενός πολλαπλασιάζεται επί έναν παράγοντα έστω k. Αν το μέσο εισόδημα των ατόμων των οποίων το εισόδημα υπερβαίνει την τιμή t παραμείνει αμετάβλητο τότε η κατανομή των εισοδημάτων είναι η Pareto.

Παράδειγμα 4 Το παράδειγμα αυτό αναφέρεται στις εργασίες της Xekalaki (1983a,1984b). Εστω X το εισόδημα μιας οικογένειας και Z τα έξοδα της για κατανάλωση. Ας υποθέσουμε ότι θέλουμε να

εκτιμήσουμε τη σχέση

$E(Z|X=x) = \alpha + \beta x, \qquad \beta \neq 0.$

Είναι γνωστό ότι τα άτομα τείνουν να δηλώνουν εισόδημα μικρότερο του πραγματικού, κατά συνέπεια είναι πιθανό να μήν είναι διαθέσιμα στοιχεία σχετικά με το πραγματικό εισόδημα X. Εστω Y το δηλούμενο εισόδημα (παρατηρούμενο μέρος της τυχαίας μεταβλητής X). Είναι λογικό να υποθέσουμε ότι μεταξύ πραγματικής τιμής X και παρατηρούμενου μέρους Y υπάρχει η σχέση

$Y = [RX]$

όπου R τυχαία μεταβλητή ανεξάρτητη από τις X και Z η οποία κατανέμεται ομοιόμορφα στο διάστημα (0,1). Είναι επίσης γνωστό, (για παράδειγμα Simon (1955)) ότι η διακριτή μορφή των εισοδημάτων ακολουθεί κατανομή Yule. Ας υποθέσουμε λοιπόν ότι από τα διαθέσιμα στοιχεία προκύπτει ότι το παρατηρούμενο εισόδημα Y ακολουθεί κατανομή Yule με παράμετρο ρ. Τότε αποδεικνύεται (Xekalaki (1984b)) ότι

$E(Z|Y=y) = \gamma + \delta y, \qquad y=0,1,2,\ldots .$

Ετσι με βάση τις διαθέσιμες παρατηρήσεις μπορούμε να εκτιμήσουμε τα γ και δ. Στη συνέχεια υπολογίζουμε τα ζητούμενα α και β μεσω των σχέσεων

$\gamma = \alpha + \beta(\rho+1)/\rho \quad$ και $\quad \delta = \beta(\rho+1)/\rho$

οι οποίες αποδεικνύεται ότι ισχύουν (Xekalaki (1984b)).

Στο πλαίσιο του παραπάνω παραδείγματος αξίζει να αναφέρουμε συμπληρωματικά ότι με βάση το συμπέρασμα της Xekalaki (1983a)

αποδεικνύεται ότι και η κατανομή του πραγματικού εισοδήματος X είναι επίσης η Yule (ρ). Δηλαδή παρά το γεγονός ότι που δεν έχουμε παρατηρήσεις για την τυχαία μεταβλητή X μπορούμε να συμπεράνουμε τη μορφή της κατανομής της με βάση τα διαθέσιμα στοιχεία για το παρατηρούμενο μέρος αυτής Y.

Παράδειγμα 5 Με βάση το αρθρο του Krishnaji (1970) τα όσα περιγράφονται στο παράδειγμα 4 μπορούν να μεταφερθούν στη συνεχή περίπτωση. Η κατανομή την οποία υποθέτουμε οτι ακολουθεί εδώ το παρατηρούμενο εισόδημα Y είναι η Pareto $(1,\alpha)$.

Σε καθένα από τα παραδείγματα που περιγράψαμε παραπάνω εμπλέκεται ένα και μόνο μοντέλο το συμπέρασμα δε στο οποίο οδηγούμεθα βασίζεται σε χαρακτηριστική ιδιότητα του μοντέλου. Είναι λίγα από τα πολλά παραδείγματα που θα μπορούσε κάποιος να αναφέρει προκειμένου να τονίσει, πέραν των άλλων, την αξία των χαρακτηρισμών στις πρακτικές εφαρμογές. Στις χαρακτηριστικές ιδιότητες τις οποίες θα διατυπώσουμε και θα αποδείξουμε στη συνέχεια της μελέτης θα προσπαθήσουμε, στο μέτρο του δυνατού, να αντιστοιχήσουμε πρακτικές καταστάσεις στις οποίες θα μπορούσαν να έχουν εφαρμογή και να διευκολύνουν την εξαγωγή συμπερασμάτων.

Οι χαρακτηριστικές ιδιότητες με τις οποίες θα ασχοληθούμε μπορούν να ταξινομηθούν στις παρακάτω κατηγορίες

1. Χαρακτηρισμοί της κατανομής Pareto που βασίζονται σε διατεταγμένες στατιστικές συναρτήσεις (order statistics)

   α) βασισμένοι σε ιδιότητες των κατανομών των διατεταγμένων στατιστικών συναρτήσεων,

β) βασισμένοι στην ανεξαρτησία συναρτήσεων των διατεταγμένων στατιστικών συναρτήσεων και

γ) βασισμένοι σε σχέσεις μεταξύ ροπών των διατεταγμένων στατιστικών συναρτήσεων.

2. Χαρακτηρισμοί της κατανομής Pareto μέσω της γραμμικής λογαριθμο-εκθετικής οικογένειας.

3. Χαρακτηρισμοί κατανομών που βασίζονται σε υπό συνθήκη πιθανότητες ή/και σε υπό συνθήκη αναμενόμενες τιμές.

4. Χαρακτηρισμοί κατανομών στο πλαίσιο της θεωρίας αξιοπιστίας.

5. Χαρακτηρισμοί κατανομών μέσω σταθμισμένων κατανομών ή άλλων συστημάτων κατανομών.

6. Χαρακτηρισμοί κατανομών στο πλαίσιο του προτύπου διαστρέβλωσης.

Από τις παραπάνω κατηγορίες η πρώτη και η δεύτερη έχουν μελετηθεί αναλυτικά από την Dimaki (1977). Στο εισαγωγικό αυτό σε θέματα χαρακτηρισμών κεφάλαιο θα γίνει πολύ σύντομη αναφορά στους χαρακτηρισμούς αυτούς, χωρίς αποδείξεις, με σχετική ενημέρωση της αντίστοιχης βιβλιογραφίας. Οι υπόλοιπες κατηγορίες αποτελούν το κύριο αντικείμενο της μελέτης και θα εξεταστούν αναλυτικά στα κεφάλαια που ακολουθούν. Αξίζει να σημειωθεί ότι θα αντιμετωπίσουμε τα επιμέρους προβλήματα τόσο στη συνεχή όσο και στη διακριτή περίπτωση γίνεται δε προσπάθεια μεταφοράς αρκετών συμπερασμάτων στο χώρο των δυο διαστάσεων.

1.3 ΧΑΡΑΚΤΗΡΙΣΜΟΙ ΤΗΣ ΚΑΤΑΝΟΜΗΣ PARETO ΒΑΣΙΣΜΕΝΟΙ ΣΤΙΣ ΔΙΑΤΕΤΑΓΜΕΝΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ.

Εστω ότι έχουμε n παρατηρήσεις από μια απόλυτα συνεχή κατανομή F και ας συμβολίσουμε με $X_{1:n}$, $X_{2:n}$,...,$X_{n:n}$ τις διατεταγμένες στατιστικές συναρτήσεις. Ο Malik (1970) απέδειξε ότι κάτω από γενικές συνθήκες η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto αν και μόνον αν οι διατεταγμένες στατιστικές συναρτήσεις $X_{m+1:n}/X_{m:n}$ και $X_{m+1:n}$, για καθορισμένο m και $1 \le m \le n-1$, είναι ανεξάρτητες. Το 1972 ο Samanta (1972) προσπάθησε να βελτιώσει το παραπάνω θεώρημα σε δυο κατευθύνσεις. Αφενός εμετρίασε την υπόθεση της απόλυτης συνέχειας της κατανομής της X απαιτώντας συνέχεια από τα δεξιά και αντικατέστησε την υπόθεση ανεξαρτησίας των $X_{m+1:n}/X_{m:n}$ και $X_{m+1:n}$ με εκείνη της ανεξαρτησίας των $X_{m+1:n}/X_{m:n}$ και $X_{k:n}$ για $m \neq k$. Εν τούτοις ο Huang (1975) διατύπωσε δυο αντι-παραδείγματα με τα οποία έδειξε ότι η συνέχεια από τα δεξιά δεν είναι αρκετή. Ουσιαστική γενίκευση της πρότασης επέτυχαν το 1973 οι Ahsanullah και Kabir (1973) με το παρακάτω θεώρημα,

Θεώρημα 1.3.1 (Ahsanullah and Kabir (1973)) Εστω X τυχαία μεταβλητή με απόλυτα συνεχή συνάρτηση κατανομής $F_X(x)$ για $\theta \le x < \infty$, $\theta > 0$. Ικανή και αναγκαία συνθήκη ώστε η X να ακολουθεί την κατανομή Pareto (θ, α) είναι η ανεξαρτησία των διατεταγμένων

στατιστικών συναρτήσεων $X_{r:n}$ και $X_{s:n}/X_{r:n}$ για κάποια r και s τέτοια ώστε $1 \le r < s \le n$.

Παρόμοια αποτελέσματα από τα οποία θα μπορούσε να προκύψει, ως πόρισμα, η περίπτωση της κατανομής Pareto, έχουν αποδειχθεί από τους Rogers (1963), Govindarajulu (1966), Ferguson (1967). Ο τελευταίος μάλιστα αντικαθιστά την υπόθεση της ανεξαρτησίας με τη σχέση $E\left(X_{i:n} | X_{i+1:n}=x\right) = \alpha x + \beta$. Στο πνεύμα αυτό ο Dallas (1976) αποδεικνύει το παρακάτω συμπέρασμα για την περίπτωση της κατανομής Pareto.

<u>Θεώρημα 1.3.2</u> (Dallas (1976)) Εστω $X_{i:n} < X_{i+1:n}$ δυο διαδοχικές διατεταγμένες στατιστικές συναρτήσεις από μια συνεχή κατανομή $F_X(x)$ με $F_X(\beta) = 0$ για $\beta > 0$. Τότε

$$E\left(X^r_{i+1:n} / X^r_{i:n} \,|X_{i:n}=x\right) = C, \quad C \text{ ανεξάρτητο του } x, \text{ και } r>0$$

αν και μόνον αν η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto, με την προυπόθεση ότι η αναμενόμενη τιμή είναι πεπερασμένη.

Ανάλογο συμπέρασμα για την κατανομή Pareto II αποδεικνύουν οι Wang και Srivastava (1980) οι οποίοι υποθέτοντας ότι η συνάρτηση κατανομής είναι συνεχής και ότι

$$E\left[\frac{1}{n-k}\sum_{i=k+1}^{n}\left(X_{i:n} - X_{k:n}\right) \;\middle|\; X_{k:n} = x\right] = \alpha x + \beta \quad \text{για κάποιο } \alpha>0$$

οδηγούνται στο συμπέρασμα ότι τα X_i ακολουθούν την κατανομή Pareto II.

Στο σημείο αυτό αξίζει να αναφερθούν μερικά συμπεράσματα τα οποία συνδέονται στενά με τα θεωρήματα που προαναφέρθηκαν αφού βασίζονται στην ανεξαρτησία συναρτήσεων ορισμένων στατιστικών συναρτήσεων. Το πρώτο διατυπώθηκε από τον Srivastava (1965) ο οποίος απέδειξε ότι κάτω από συγκεκριμένους περιορισμούς η X ακολουθεί την κατανομή Pareto (θ,α) αν και μόνον αν οι στατιστικές συναρτήσεις $X_{1:n}$ και $\left(X_{1:n}+\ldots+X_{n:n}\right)\Big/X_{1:n}$ είναι ανεξάρτητες. Στη συνέχεια αναφέρουμε χαρακτηρισμό που οφείλεται στον Samanta (1972) και ο οποίος βασίζεται στην ανεξαρτησία ανάμεσα σε ένα διάνυσμα πηλίκων διατεταγμένων στατιστικών συναρτήσεων και μιας διατεταγμένης στατιστικής συνάρτησης. Συγκεκριμένα απέδειξε το παρακάτω θεώρημα,

<u>Θεώρημα 1.3.3</u> (Samanta (1972)) Εστω $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$ οι διατεταγμένες στατιστικές συναρτήσεις ενός τυχαίου δείγματος μεγέθους n ($n \geq 2$) από πληθυσμό με συνάρτηση κατανομής $F_X(x)$ η οποία είναι συνεχής από τα δεξιά και $F_X(x) = 0$ αν $x \geq \theta$, $\theta > 0$. Το διάνυσμα των τυχαίων μεταβλητών $\left(X_{2:n}\Big/X_{1:n},\ X_{3:n}\Big/X_{1:n},\ \ldots,\ X_{n:n}\Big/X_{1:n}\right)$ και η $X_{1:n}$ είναι ανεξάρτητα αν και μόνον αν $F_X(x) = 1 - \theta^{\alpha}\, x^{-\alpha}$ για $\theta > 0$.

Τέλος ο Dallas (1976) αποδεικνύει το παρακάτω θεώρημα.

<u>Θεώρημα 1.3.4</u> (Dallas (1976)) Εστω $X_{i:n} < X_{i+1:n}$ δυο διαδοχικές

και W είναι ανεξάρτητες, τότε υπάρχει σταθερά c τέτοια ώστε

$$\int_z^\infty g(z,x)\, f_X(x)\, dx = C\left[1 - F_X(x)\right] \qquad (1.3.1)$$

για όλα τα z, όπου $f_X(x) = F_X'(x)$.

Παρατηρούμε ότι το παραπάνω θεώρημα δίνει μια ικανή μόνο συνθήκη. Μπορεί όμως να θεωρηθεί σα χαρακτηρισμός αφού το αντίστροφο σε κάθε συγκεκριμένη περίπτωση είναι προφανές. Ας θεωρήσουμε για παράδειγμα την περίπτωση της κατανομής Pareto.

Θεώρημα 1.3.6 Εστω ότι ισχύουν οι υποθέσεις του θεωρήματος 1.3.5. Θέτουμε

$$g(x,y) = \begin{cases} y/x & \text{αν } \theta \le x \le y \\ 0 & \text{οπουδήποτε αλλού.} \end{cases} \qquad (1.3.2)$$

Τότε η κατανομή της X ορίζεται μονοσήμαντα ως Pareto.

Απόδειξη

Ευθύ: Εστω ότι ισχύει η σχέση (1.3.1) δηλαδή,

$$z^{-1}\int_z^\infty x\, f_X(x)\, dx = C\left[1-F_X(z)\right]$$

ή

$$-z\, f_X(z) = C\left[1-F_X(z)\right] - C\, z\, f_X(z).$$

Επομένως,

$$z \frac{\partial \bar{F}_X(x)}{\partial z} = C \bar{F}_X(z) + C z \frac{\partial \bar{F}_X(x)}{\partial z} .$$

Δηλαδή,

$$(1-C) \ln \bar{F}_X(z) = C \ln z + \ln C_1.$$

Κατά συνέπεια,

$$F_X(z) = 1 - C_2 z^{C/(1-C)}.$$

Αλλά το γεγονός ότι $F_X(\infty) = 1$ συνεπάγεται ότι $C/(1-C) = -\alpha$ όπου $\alpha>0$. Δηλαδή,

$$F_X(z) = 1 - C_2 z^{-\alpha}.$$

Επίσης επειδή $F_X(\theta) = 0$ έπεται ότι $C_2 = \theta^{\alpha}$. Δηλαδή,

$$F(z) = 1 - \theta^{\alpha} z^{-\alpha} , \qquad z>\theta, \ \alpha>0, \ \theta>0.$$

Επειδή δε η $F_X(x)$ είναι απόλυτα συνεχής, καταλήγουμε στο συμπέρασμα ότι,

$$f_X(x) = \begin{cases} \alpha \theta^{\alpha} x^{-(\alpha+1)} & \text{αν } x \geq \theta, \ \alpha>0, \ \theta>0 \\ 0 & \text{οπουδήποτε αλλού.} \end{cases}$$

Αντίστροφο: Εστω ότι η $f_X(x)$ είναι η συνάρτηση πυκνότητας πιθανότητας της κατανομής Pareto και έστω ότι η $g(x,y)$ δίνεται από τη σχέση (1.3.2). Τότε,

$$\int_z^\infty g(z,x)\ f_X(x)\ dx = \int_z^\infty z^{-1}\ x\ \alpha\ \theta^\alpha\ x^{-\alpha-1}\ dx$$

$$= \frac{\alpha}{\alpha-1}\ \theta^\alpha\ z^{-\alpha}$$

$$= \frac{\alpha}{\alpha-1}\ \left[1 - F_X(x)\right]$$

Δηλαδή, θέτοντας $C = \frac{\alpha}{\alpha-1}$ συνεπάγεται ότι ισχύει η σχέση (3.2.1).

Σε χαρακτηρισμούς που βασίζονται σε ιδιότητες κατανομών διατεταγμένων στατιστικών συναρτήσεων αναφέρεται το θεώρημα που ακολουθεί. Συγκεκριμένα αποδεικνύεται ότι η κατανομή Pareto χαρακτηρίζεται από την κατανομή των στατιστικών συναρτήσεων $X_{s_i:n}/X_{r:n}$ και $X_{s_i-r:n-r}$ για δυο τιμές s_1 και s_2 $\left(X_{s_i-r:n-r}\right.$ είναι η διατεταγμένη στατιστική συνάρτηση τάξης (s_i-r) από ένα δείγμα μεγέθους $\left.n-r\right)$.

<u>Θεώρημα 1.3.6</u> (Δημάκη (1988a)) Εστω X τυχαία μεταβλητή με απόλυτα συνεχή και αξουσα συνάρτηση κατανομής $F_X(x)$ για όλα τα $x\geq 1$ και $F_X(x) = 0$ για όλα τα $x<1$. Τότε η τυχαία μεταβλητή X έχει την κατανομή Pareto τότε και μόνο τότε αν για δοθέν r και δυο αριθμούς s_1 και s_2 $(1\leq r<s_1<s_2\leq n)$ οι τυχαίες μεταβλητές $X_{s_i:n}/X_{r:n}$ και $X_{s_i-r:n-r}$ είναι ισόνομες για $i=1,2$.

Ενα άλλο ενδιαφέρον συμπέρασμα το οποίο θα διατυπώσουμε χαρακτηρίζει την κατανομή Pareto με τη βοήθεια μιας συνάρτησης

διατεταγμένων στατιστικών συναρτήσεων η οποία έχει την ίδια κατανομή με εκείνη του πληθυσμού. Η συνάρτηση που επιλέγεται στην προκειμένη περίπτωση είναι η $\min(X_1^n,\ldots,X_n^n)$. Συγκεκριμένα αποδεικνύεται (Dimaki, 1988b) ότι κάτω από ορισμένους περιορισμούς τα συμπεράσματα

(i) $X_{1:n}^n \overset{d}{=} X$

και

(ii) $X \sim$ Pareto (θ,α)

είναι ισοδύναμα.

Στην κατηγορία των χαρακτηρισμών που έχουμε περιγράψει στην ενότητα αυτή μπορεί να περιληφθεί ένας ακόμη χαρακτηρισμός που διατυπώθηκε από τους Beg και Kirmani (1974). Συγκεκριμένα απέδειξαν ότι οι προτάσεις

(i) $E\left[\dfrac{X_1}{X_{1:n}} \;\middle|\; X_{1:n} = x\right] = C > 0$

και

(ii) $X \sim$ Pareto

είναι ισοδύναμες.

Το τρίτο σκέλος των χαρακτηρισμών που βασίζονται σε διατεταγμένες στατιστικές συναρτήσεις είναι εκείνοι στους οποίους εμπλέκονται σχέσεις μεταξύ ροπών των διατεταγμένων στατιστικών

συναρτήσεων. Χαρακτηρισμοί αυτού του τύπου έχουν αποδειχθεί από την Dimaki (1977). Επίσης παρόμοιοι χαρακτηρισμοί για διάφορες κατανομές μεταξύ των οποίων και η Pareto έχουν αποδειχθεί και από τον Lin (1988). Στα επόμενα παραθέτουμε τέσσερις από αυτούς τους χαρακτηρισμούς αφού προηγουμένως περιγράψουμε τη γενική υπόθεση στο πλαίσιο της οποίας αποδεικνύεται ότι ισχύουν.

Εστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές η καθεμιά από τις οποίες έχει κατανομή $F(x)$ με $F(1)=0$ και $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ οι διατεταγμένες στατιστικές συναρτήσεις ενός δείγματος μεγέθους n από αυτή την κατανομή. Εστω επίσης $U_1, U_2, \ldots$ ανεξάρτητες τυχαίες μεταβλητές που η καθεμιά κατανέμεται ομοιόμορφα στο (0,1). Ορίζουμε την $H(u) = \inf \left\{x | F(x) \geq u\right\}$, $0<u<1$. Τότε για $0<u<1$ οι σχέσεις (i) $H(u) \leq x$ και (ii) $u \leq F(x)$ είναι ισοδύναμες και η κατανομή των $H(U_1)$, $H(U_2), \ldots$ είναι ίδια με την κατανομή των $X_1, X_2, \ldots$. Υποθέτουμε ότι η F είναι απόλυτα συνεχής, κατά συνέπεια η $H'(x)$ υπάρχει παντού για $0<u<1$. Τέλος εισάγουμε τον παρακάτω συμβολισμό,

$$I(u;k,m) = u^k (1-u)^m \qquad 0 < u < 1.$$

$$\mu_{k:n}(r) = E(X^r_{k:n}) \quad , \quad \mu_{k:n}(1) \equiv \mu_{k:n}$$

$$\sigma^2_{k:n} = \mathrm{Var}(X_{k:n})$$

Τότε καθεμιά από τις παρακάτω σχέσεις ορίζει μονοσήμαντα την κατανομή της X ως Pareto,

(i) $\mu_{i+1:n}(2) - \mu_{i:n}(2) = 2\,(n-i)^{-1}\,\frac{1}{\alpha}\,\mu_{i+1:n}(2)$,

$$i=1,2,\dots\ ,\ n=i+1,i+2,\dots$$

(ii) $$\mu_{i+1:n}(2) - \mu_{i:n-1}(2) = \frac{2}{n\,\alpha}\,\mu_{i+1:n}(2),$$

$$i=1,2,\dots\ ,\ n=i+1,i+2,\dots$$

(iii) $$\mu_{i+1:n}(2)\,\mu_{i+1:n} + \mu_{i:n}(n)\,\mu_{i+1:n} = \mu_{i+1:n}(2)\,\mu_{i:n},$$

$$i=1,2,\dots\ ,\ n=i+1,i+2,\dots$$

(iv) $$\sigma^2_{i+1:n} - \sigma^2_{i:n-1} = \mu_{i+1:n}(2)\left[1 - \mu_{i:n}\mu^{-1}_{i+1:n}\right] - \left[\mu^2_{i+1:n} - \mu^2_{i:n-1}\right]$$

$$i=1,2,\dots\ ,\ n=i+1,i+2,\dots$$

1.4 ΧΑΡΑΚΤΗΡΙΣΜΟΙ ΤΗΣ ΚΑΤΑΝΟΜΗΣ PARETO ΜΕΣΩ ΤΗΣ ΓΡΑΜΜΙΚΗΣ ΛΟΓΑΡΙΘΜΟ-ΕΚΘΕΤΙΚΗΣ ΟΙΚΟΓΕΝΕΙΑΣ

Οπως είναι γνωστό η κατανομή Pareto είναι μέλος της γραμμικής λογαριθμο-εκθετικής οικογένειας. Αποδεικνύεται λοιπόν, (Dimaki (1977)) ότι μπορούμε να χαρακτηρίσουμε την κατανομή Pareto μέσω της οικογένειας αυτής. Στην ενότητα αυτή θα δώσουμε τον ορισμό της γραμμικής λογαριθμο-εκθετικής οικογένειας και θα παραθέσουμε ενδεικτικά δυο συμπεράσματα που αποδεικνύονται στην εργασία που αναφέρθηκε παραπάνω.

<u>Ορισμός 1.4.1</u> Μια οικογένεια κατανομών σε έναν Ευκλείδειο χώρο

λέγεται ότι ανήκει στη γραμμική λογαριθμο-εκθετική οικογένεια αν ορίζεται από συναρτήσεις πυκνότητας πιθανότητας της μορφής,

$$f(x,\theta) = x^{\theta} \exp\left[C(x) - B(\theta)\right]$$

Είναι γνωστό ότι αν ο λογάριθμος της χαρακτηριστικής συνάρτησης μιας μεταβλητής αναπτυχθεί σε δυναμοσειρά οι συντελεστές αυτού του αναπτύγματος ονομάζονται ημιαναλλοίωτες ή αθροιστικές (παράμετροι) και συμβολίζονται με k_r $r=0,1,2,\ldots$. Η συνάρτηση μιας μεταβλητής t η οποία όταν αναπτυχθεί σε δυνάμεις του t έχει σαν συντελεστές του αναπτύγματος τις ημιαναλλοίωτες μιας κατανομής ονομάζεται γεννήτρια συνάρτηση αθροιστικών παραμέτρων. Δηλαδή,

$$\Psi_X(t) = \log \varphi_X(t) = \sum_{r=0}^{\infty} k_r \frac{(it)^r}{r!}$$

όπου $\Psi_X(.)$ είναι η γεννήτρια συνάρτηση αθροιστικών παραμέτρων της τυχαίας μεταβλητής X, $\varphi_X(.)$ είναι η χαρακτηριστική συνάρτηση της τυχαίας μεταβλητής X και k_r είναι η ημιαναλλοίωτη ή αθροιστική παράμετρος τάξης r της τυχαίας μεταβλητής X.

Κατ' αναλογία ορίζουμε τα παρακάτω.

<u>Ορισμός 1.4.2</u> Εστω $K_X(s)$ ο λογάριθμος του μετασχηματισμού Mellin μιας τυχαίας μεταβλητής X η οποία έχει συνάρτηση πυκνότητας πιθανότητας $f_X(x)$. Θεωρούμε το ανάπτυγμα σε δυναμοσειρά του $K_X(s)$ και συμβολίζουμε με v_r το συντελεστή του $s^r/r!$ στο ανάπτυγμα αυτό.

Δηλαδή,

$$K_X(s) = \log M_X(s) = \sum_{r=0}^{\infty} v_r \frac{s^r}{r!}$$

Ονομάζουμε την $K_X(s)$ γεννήτρια συνάρτηση λογαριθμο-αθροιστικών παραμέτρων της τυχαίας μεταβλητής X και την v_r λογαριθμο-αναλλοίωτη ή λογαριθμο-αθροιστική παράμετρο τάξης r της τυχαίας μεταβλητής X.

Θα προχωρήσουμε παραθέτοντας δυο χαρακτηριστικές ιδιότητες. Η πρώτη χαρακτηρίζει γενικά τη γραμμική λογαριθμο-εκθετική οικογένεια.

Θεώρημα 1.4.1 (Dimaki (1977)) Εστω $K_X(s)$ ο λογάριθμος του μετασχηματισμού Mellin της τυχαίας μεταβλητής X η οποία έχει συνάρτηση πυκνότητας πιθανότητας την $f_X(x;\theta)$. Τότε η $f_X(x,\theta)$ ανήκει στην γραμμική λογαριθμο-εκθετική οικογένεια με παράμετρο θ αν και μόνον αν οι v_r ικανοποιούν τη σχέση

$$v_{r+1} = \frac{\partial v_r}{\partial \theta}, \qquad r = 1,2,\ldots$$

Το δεύτερο θεώρημα στο οποίο θα αναφερθούμε χαρακτηρίζει την κατανομή Pareto μέσω της γραμμικής λογαριθμο-εκθετικής οικογένειας.

Θεώρημα 1.4.2 (Dimaki (1977)) Η γραμμική λογαριθμο-εκθετική οικογένεια έχει την κατανομή Pareto αν και μόνον αν $v_2 = v_1^2$.

Κ Ε Φ Α Λ Α Ι Ο 2

ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ

PARETO & YULE

2.1 ΕΙΣΑΓΩΓΗ

Οπως έχει ήδη αναφερθεί η μονοδιάστατη κατανομή Pareto εφαρμόζεται κυρίως στη μελέτη των εισοδημάτων φυσικών προσώπων, χρησιμοποιείται επίσης για την περιγραφή των διακυμάνσεων των τιμών των μετοχών, της δυναμικότητας κοιτασμάτων πρώτων υλών, της συσσώρευσης λαθών σε δίκτυα επικοινωνίας, του μεγέθους βιομηχανικών εγκαταστάσεων καθώς και σε προβλήματα αξιοπιστίας. Κατά συνέπεια παρουσιάζει ιδιαίτερο ενδιαφέρον η μελέτη διδιάστατων κατανομών των οποίων οι περιθώριες κατανομές είναι μονοδιάστατες Pareto.

Θα λέμε ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί μια διδιάστατη κατανομή Pareto αν και οι δυο περιθώριες κατανομές είναι μονοδιάστατες Pareto.

Ειναι γνωστό βέβαια ότι οι περιθώριες κατανομές δεν προσδιορίζουν μονοσήμαντα τη διδιάστατη κατανομή. Για να περιορίσουμε την τάξη των δυνατών κατανομών πρέπει να επιβάλλουμε κάποια επιπλέον συνθήκη. Το να είναι οι δεσμευμένες κατανομές του ίδιου τύπου με τις περιθώριες είναι ένας λογικός περιορισμός. Είναι όμως γνωστό πως υπάρχουν αποδεκτές και χρησιμοποιούμενες κατανομές που δεν έχουν περιθώριες και δεσμευμένες κατανομές του ίδίου τύπου, η διδιάστατη Poisson για παράδειγμα (Seshadri, and Patil, (1963)). Γενικά μια διδιάστατη κατανομή έχει ξεχωριστή σημασία αν διατηρεί κάποιες, αν όχι όλες από τις χαρακτηριστικές ιδιότητες της αντίστοιχης μονοδιάστατης κατανομής.

Στο κεφάλαιο αυτό θα ασχοληθούμε κυρίως με τη μελέτη και συστηματοποίηση των διαδικασιών γένεσης διδιάστατων μορφών

κατανομών Pareto. Οι προκύπτουσες μορφές διδιάστατων κατανομών μπορούν να διακριθούν στις εξής κατηγορίες:

α) Διδιάστατες μορφές προκύπτουσες από το γενικό ορισμό διδιάστατης κατανομής του Mongenstern.

β) Διδιάστατες μορφές που προκύπτουν με εφαρμογή της συναρτησιακής σχέσης που υπάρχει μεταξύ εκθετικής μεταβλητής και μεταβλητής Pareto στις ήδη γνωστές και αποδεκτές διδιάστατες εκθετικές κατανομές.

γ) Διδιάστατες μορφές που προκύπτουν ως μίγματα εκθετικών με κατάλληλες μορφές (μονοδιάστατες ή διδιάστατες Gamma).

2.2 ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ PARETO ΤΥΠΟΥ I

Υπάρχουν πολλές τεχνικές για να κατασκευάσει κανείς διδιάστατες κατανομές με καθορισμένες περιθώριες $F_X(x)$ και $F_Y(y)$. Ο Morgenstern (1956) εισήγαγε μια μονοπαραμετρική οικογένεια με συνάρτηση κατανομής

$$F_{X,Y}(x,y)=F_X(x)F_Y(y)[1+\theta(1-F_X(x))(1-F_Y(y))] \qquad (2.2.1)$$

όπου $-1 \le \theta \le 1$.

Στη περίπτωση που το τυχαίο διάνυσμα (X,Y) είναι συνεχές η συνάρτηση πυκνότητας πιθανότητας δίνεται από την

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)[1+\theta(2F_X(x)-1)(2F_Y(y)-1)] \qquad (2.2.2)$$

όπου $-1 \le \theta \le 1$.

Εστω ότι η $F_X(x)$ και $F_Y(y)$ είναι μονοδιάστατες Pareto δηλαδή

$$F_X(x)=1-x^{-\alpha} \qquad x>1,\ \alpha>0$$

και

$$F_Y(y)=1-y^{-\beta} \qquad y>1,\beta>0 \ .$$

Τοτε οι σχέσεις (2.2.1) και (2.2.2) γίνονται αντίστοιχα,

$$F_{X,Y}(x,y)=(1-x^{-\alpha})(1-y^{-\beta})(1+\theta x^{-\alpha}y^{-\beta}) \qquad (2.2.3)$$

όπου $x>1, y>1, \alpha>0, \beta>0, \ -1 \le \theta \le 1$

και

$$f_{X,Y}(x,y)=\alpha\beta x^{-\alpha-1}y^{-\beta-1}\left[1+\theta(1-2x^{-\alpha})(1-2y^{-\beta})\right] \qquad (2.2.4)$$

όπου $x>1, y>1, \alpha>0, \beta>0, \ -1 \le \theta \le 1$.

Παρατηρούμε ότι για $\theta=0$ προκύπτει η διπλή Pareto. Η διδιάστατη κατανομή Pareto τύπου Morgenstern έχει μελετηθεί αναλυτικά από την Dimaki (1977).

Στη συνέχεια παρατίθενται μερικές από τις κυριώτερες ιδιότητες της κατανομής αυτής. Επειδή η συνάρτηση κατανομής και η συνάρτηση πυκνότητας πιθανότητας ειναι συμμετρικές ως προς x και y, αρκεί η μελέτη της κατανομής ως προς τη μια μεταβλητή, εστω τη

X. Η περιθώρια κατανομή $F_X(x)$, όπως είναι φυσικό, είναι μονοδιάστατη Pareto. Αντίθετα, η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας δίνεται από τον τύπο

$$f_{X|(Y=y)}(x|y)=\alpha x^{-(\alpha+1)}[1+\theta(1-2x^{-\alpha})(1-2y^{-\beta})].$$

Είναι προφανές επομένως ότι οι δεσμευμένες περιθώριες κατανομές δεν είναι τύπου Pareto.

Η συνάρτηση παλινδρόμησης της X επί της Y δίνεται από τον τύπο

$$E(X|Y=y) = \frac{\alpha}{\alpha-1} + \frac{\alpha\theta(1-2y^{-\beta})}{(\alpha-1)(2\alpha-1)} .$$

Παρατηρούμε ότι όταν $\theta=0$ τότε $E(X|Y=y) = \alpha/(\alpha-1)$. Το αποτέλεσμα αυτό είναι αναμενόμενο αφου η περίπτωση $\theta=0$ συνεπάγεται την ανεξαρτησία των X και Y οπότε

$$E(X|Y=y) = E(X) = \alpha/(\alpha-1).$$

Επίσης αν θεωρήσουμε δυο τιμές του θ ίσες κατά απόλυτη τιμή βρίσκουμε ότι

$$\frac{1}{2}\left\{E_{-\theta}(X|Y=y)+E_{+\theta}(X|Y=y)\right\} = \frac{\alpha}{\alpha-1} .$$

Υπάρχει δηλαδή συμμετρία ως προς την τιμή $\alpha/(\alpha-1)$ ή ισοδύναμα ως προς E(X). Γενικώτερα , μπορεί να αποδειχθεί οτι

$$\frac{1}{2}\left\{E_{-\theta}(X^r|Y=y)+E_{+\theta}(X^r|Y=y)\right\} = \frac{\alpha}{\alpha-r} = E(X^r).$$

Η σκεδαστική συνάρτηση var(X|Y=y) υπολογίζεται εύκολα και

ισούται

$$\mathrm{var}(X|Y=y) = \frac{\alpha}{(\alpha-2)(\alpha-1)^2} + \frac{\alpha\theta(1-2y^{-\beta})}{(\alpha-1)(\alpha-2)(2\alpha-1)^2} - \frac{\alpha^2\theta^2(1-2y^{-\beta})^2}{(\alpha-1)(2\alpha-1)^2}\ .$$

Και πάλι, όπως είναι φυσικό, όταν $\theta=0$ συνεπάγεται ότι $\mathrm{var}(X|Y=y)=\alpha/(\alpha-2)(\alpha-1)^2$ δηλαδή $\mathrm{var}(X|Y=y)=\mathrm{var}(X)$.
Τέλος ο συντελεστής συσχέτισης των X και Y δίνεται από τον τύπο

$$\rho=\theta\,[\alpha\beta(\alpha-2)(\beta-2)]^{1/2}(2\alpha-1)^{-1}(2\beta-1)^{-1}.$$

2.3 ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ PARETO ΤΥΠΟΥ II

Οι δισδιάστατες κατανομές Pareto που θα περιγράψουμε στην ενότητα αυτή προκύπτουν με εφαρμογή της συναρτησιακής σχέσης που υπάρχει μεταξύ μιας μεταβλητής Pareto και μιας εκθετικής μεταβλητής στις ήδη γνωστές δισδιάστατες εκθετικές κατανομές.
Είναι γνωστό ότι υπάρχει ισοδυναμία μεταξύ των προτάσεων

(i) $X \sim$ Pareto (θ, α)

(ii) $\ln X \sim$ Εκθετική $(\ln \theta, 1/\alpha)$.

Η ισοδυναμία αυτή μεταφέρεται και ισχύει και στο χώρο των δυο διαστάσεων, όπως φαίνεται από τα παραδείγματα που αναπτύσσονται στις ενότητες που ακολουθούν. Αποδεικνύεται δηλαδή ότι οι προτάσεις

(i) $(X,Y) \sim$ Διδιάστατη Pareto

(ii) $(\ln X, \ln Y) \sim$ Διδιάστατη Εκθετική

είναι ισοδύναμες.

2.3.1 Διδιάστατη κατανομή Pareto τύπου Gumbel

Ο Gumbel (1960) εισήγαγε μια διδιάστατη εκθετική κατανομή με συνάρτηση κατανομής

$$F_{Z,W}(z,w)=1-e^{-z}-e^{-w}+e^{-z-w-\theta zw} \qquad z \geq \theta, w \geq 0 .$$

Εστω,

$$Z=\ln X^{\alpha} \qquad W=\ln Y^{\beta} \qquad \alpha > 0,\ \beta > 0.$$

Τότε η κατανομή του τυχαίου διανύσματος (X,Y) της οποίας η συνάρτηση κατανομής δίνεται από τον τύπο

$$F_{X,Y}(x,y)=1-x^{-\alpha}-y^{-\beta}+y^{-\beta}x^{-\alpha(1+\beta\theta \ln y)} \quad ; \qquad (2.3.1)$$

όπου $x>1, y>1, \alpha, \beta>0, 0 \leq \theta \leq 1$,

μπορεί να θεωρηθεί ως η διδιάστατη επέκταση της κατανομής Pareto τύπου Gumbel. Προφανώς οι περιθώριες συναρτήσεις κατανομής των μεταβλητών X και Y δίνονται αντίστοιχα από τους τύπους

$$F_X(x)=1-x^{-\alpha} \qquad x>1, \alpha>0$$

και

$$F_X(y)=1-y^{-\beta} \qquad y>1, \beta>0 \ .$$

Είναι δηλαδή οι περιθώριες κατανομές μονοδιάστατες Pareto.
Η συνάρτηση πυκνότητας πιθανότητας της κατανομής αποδεικνύεται ότι ισούται με

$$f_{X,Y}(x,y)=\alpha\beta y^{-\beta-1}x^{-\alpha(1+\beta\theta \ln y)-1}\left[(1+\alpha\theta \ln x)(1+\beta\theta \ln y)-\theta\right] \qquad (2.3.2)$$

όπου $x>1, y>1, \ \alpha,\beta>0, \ 0 \le \theta \le 1.$

Η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας ισούται με

$$f_{X|(Y=y)}(x|y)=\alpha x^{-\alpha(1+\beta\theta \ln y)-1}\left[(1+\alpha\theta \ln x)(1+\beta\theta \ln y)-\theta\right] \ .$$

$$x,y>1, \ \alpha,\beta>0, \ \ 0 \le \theta \le 1$$

ενώ

$$E(X|Y=y)= \frac{\alpha B}{\alpha B-1} + \frac{\alpha\theta}{(\alpha B-1)^2} \quad \text{για } \alpha B>1, \text{ όπου } \ B=1+\beta\theta \ln y \ .$$

Αντίστοιχες είναι οι εκφράσεις των συναρτήσεων $f_{Y|(X=x)}$ και $E(Y|X=x)$ λόγω συμμετρίας της $f_{X,Y}(x,y)$ ως προς x και y. Τέλος αξίζει να αναφερθεί ότι ο συντελεστής συσχέτισης ρ δίνεται από τον τύπο

$$\rho=-\left[(\alpha-2)(\beta-2)/\alpha\beta\right]^{1/2}\left[1-me^{m}E_1(m)\right] \quad \text{όπου } m=(\alpha-1)(\beta-1)/\alpha\beta\theta$$

$$\text{και } E_1(m)=\int_m^{\infty} t^{-1}e^{-t}dt.$$

Παρατηρώντας την τελευταία αυτή σχέση συμπεραίνουμε ότι ο συντελεστής συσχέτισης δεν μπορεί να πάρει αρνητικές τιμές. Ειδικώτερα στην περίπτωση που $\alpha=\beta=d$ και $\vartheta=1$ παίρνει τη μορφή

$$\rho=-\left[(d-2)/d\right]\left[1-m_1e^{m_1}E_1(m_1)\right] \qquad \text{όπου } m_1=(d-1)^2/d^2 .$$

Επιπλέον όταν $d \to 2$ τότε $\rho \to 0$ ενώ όταν $d \to \infty$ τότε $m_1 \to 1$ και κατά συνέπεια $\rho \to -0.403664$.

2.3.2 Διδιάστατη Pareto τύπου Marshall-Olkin.

Μια άλλη διδιάστατη εκθετική στην οποία θα εφαρμόσουμε τη συναρτησιακή σχέση που υπάρχει μεταξύ της μεταβλητής Pareto και της εκθετικής, είναι εκείνη που εισήγαγαν οι Marshall και Olkin (1967). Η συνάρτηση κατανομής της διδιάστατης εκθετικής την οποία όρισαν δίνεται από τη σχέση

$$F_{U,V}(u,v)=1-e^{-(k+r)u}-e^{-(l+r)v}+e^{-ku-lv-r\max(u,v)}.$$

Τότε η συνάρτηση κατανομής του τυχαίου διανύσματος $(X=e^{U/\alpha}, Y=e^{V/\alpha})$ δίνεται από τη σχέση

$$F_{X,Y}(x,y)=1-x^{-\alpha(k+r)}-y^{-\alpha(l+r)}+x^{-\alpha k}y^{-\alpha l}\left[\max(x^{\alpha},y^{\alpha})\right]^{-r}. \qquad (2.3.3)$$

Την κατανομή η οποία ορίζεται από την (2.3.3) θα ονομάζουμε διδιάστατη κατανομή Pareto τύπου Marshall-Olkin.

Παρατηρούμε ότι

$$F_X(x)=F_{X,Y}(x,\infty)=1-x^{-\alpha(k+r)}$$

και

$$F_Y(y)=F_{X,Y}(\infty,y)=1-y^{-\alpha(1+r)},$$

δηλαδή και οι δύο περιθώριες κατανομές είναι μονοδιάστατες Pareto.

Μια ενδιαφέρουσα παρατήρηση που μπορούμε να κάνουμε σχετικά με τη διδιάστατη αυτή Pareto είναι ότι αποτελείται από δυο μέρη: ένα απόλυτα συνεχές *(absolutely continuous)* και ένα ιδιάζον *(singular)*. Ετσι, η συνάρτηση επιβίωσης της κατανομής μπορεί να γραφεί ως εξής:

$$\bar{F}_{X,Y}(x,y)=c\bar{F}_{abs}(x,y)+(1-c)\bar{F}_{sin}(x,y) \qquad (2.3.4)$$

και επομένως,

$$\frac{\partial^2\bar{F}_{X,Y}(x,y)}{\partial x\partial y} = cf_{abs}(x,y).$$

Επίσης μπορεί να αποδειχθεί ότι

$$f_{abs}(x,y) = \begin{cases} \dfrac{\alpha^2(k+r)lR}{k+1}\, x^{-\alpha(k+r)-1}\, y^{-al-1} & \text{αν} \quad x>y \\ \dfrac{\alpha^2(l+r)kR}{k+1}\, y^{-\alpha(l+r)-1}\, x^{-\alpha k-1} & \text{αν} \quad x<y \end{cases} \qquad (2.3.5)$$

όπου $R=k+r+1$.

2.3.3 Διδιάστατη κατανομή Pareto τύπου Mardia II

Τη λογική που εφαρμόσαμε ήδη στις διδιάστατες εκθετικές των Gumbel και Marshall-Olkin εφάρμοσε και ο Mardia (1962) στη διδιάστατη εκθετική κατανομή των Wicksell-Kibble. Ετσι κατέληξε σε διδιάστατη κατανομή Pareto με συνάρτηση πυκνότητας πιθανότητας

$$f_{X,Y}(x,y) = \frac{\alpha\beta}{xy(1-\gamma)} \left[(\theta_1/x)^{\alpha}\ (\theta_2/y)^{\beta}\right] I_0(Q) \qquad (2.3.6)$$

$$\text{όπου} \quad Q= \frac{2\left(\gamma\alpha\beta\ \ln\ (x/\theta_1)\ln\ (y/\theta_2)\right)^{1/2}}{1-\gamma}\ ,$$

$x>\theta_1>0$, $y>\theta_2>0$, $\alpha>0$, $\beta>0$, $0 \le \gamma < 1$ και $I_0(.)$ ειναι η μετασχηματισμένη συναρτηση Bessel πρώτου είδους και μηδενικής τάξης. Στο άρθρο που προαναφέρθηκε ο Mardia υπολογίζει τη δεσμευμένη συνάρτηση πυκνότητας πιθανότητας, τη συνάρτηση παλινδρόμησης καθώς και τη σκεδαστική συνάρτηση της κατανομής.

2.4 ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ PARETO ΤΥΠΟΥ III

Στο πλαίσιο της θεωρίας ουρών η μονοδιάστατη κατανομή Pareto εμφανίζεται ως κατανομή του χρόνου εξυπηρέτησης πελατών. Προκύπτει στην περίπτωση εκείνη κατά την οποία οι χρόνοι εξυπηρέτησης ακολουθούν εκθετική κατανομή η δε εκθετική παράμετρος είναι τυχαία μεταβλητή η οποία κατανέμεται σύμφωνα με την κατανομή Gamma. Με άλλα λόγια, μίγμα εκθετικής (θ) με Gamma (α,1) οδηγεί σε Pareto (1,α). Στα επόμενα θα εξετάσουμε κατά πόσον η ιδιότητα αυτή μεταφερόμενη στις δυο διαστάσεις οδηγεί σε διδιάστατη κατανομή η οποία μπορεί να θεωρηθεί διμεταβλητή επέκταση της κατανομής Pareto. Στο μοντέλο που θα θεωρήσουμε έχει υποτεθεί ότι σε κάθε υποπληθυσμό, οι τυχαίες μεταβλητές X και Y ειναι ανεξάτητες και ακολουθούν την ίδια εκθετική κατανομή με παράμετρο θ. Η παράμετρος θ μεταβάλλεται μεταξύ των υποπληθυσμών ακολουθώντας κατανομή Gamma. Το θεώρημα 2.4.1 συνοψίζει τόσο το μοντέλο όσο και το συμπέρασμα που αναφέρεται στην κατανομή του τυχαίου διανύσματος (X,Y).

Θεώρημα 2.4.1 Εστω δυο τυχαίες μεταβλητές X και Y των οποίων η από κοινού κατανομή είναι η διπλή εκθετική με συνάρτηση κατανομής

$$F_{X,Y}(x,y) = 1 - e^{-\theta(x+y)} \qquad x,y>0.$$

Εστω ότι το θ είναι μια Gamma ($1,\alpha$) μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$f_{\theta}(\theta) = \frac{\theta^{\alpha-1}}{\Gamma(\alpha)} e^{-\theta} \qquad \theta>0.$$

Τότε η κατανομή του τυχαίου διανύσματος (X,Y) έχει συνάρτηση επιβίωσης την

$$\bar{F}(x,y) = (x+y+1)^{-\alpha} \qquad \alpha>0. \qquad (2.3.7)$$

Απόδειξη Με βάση τον ορισμό της συνάρτησης επιβίωσης ενός μίγματος έχουμε ότι

$$\bar{F}(x,y) = \int_0^\infty f_\theta(\theta)\ \bar{F}_{X,Y}(x,y)\ d\theta$$

$$= \int_0^\infty \frac{\theta^{\alpha-1}}{\Gamma(\alpha)}\ e^{-\theta}\ e^{-\theta(x+y)}\ d\theta$$

$$= \left(1/\Gamma(\alpha)\right)\ (x+y+1)^{-\alpha} \int_0^\infty \{\theta(x+y+1)\}^{\alpha-1}\ e^{-\theta(x+y+1)}\ d\{\theta(x+y+1)\}$$

$$= (x+y+1)^{-\alpha}\ .$$

Η παραπάνω κατανομή έχει μελετηθεί από τον Mardia (1962). Θα ονομάζουμε αυτήν **διδιάστατη κατανομή Pareto τύπου Mardia I.** Για την κατανομή αυτή αξίζει να αναφέρουμε μερικά βασικά συμπεράσματα. Η συνάρτηση πυκνότητας πιθανότητας της κατανομής δίνεται από την

$$f_{X,Y}(x,y) = \alpha(\alpha+1)(x+y+1)^{-(\alpha+2)} \qquad (2.3.8)$$

όπου $x>0,\ y>0,\ \alpha>0.$

Από τη σχέση αυτή προκύπτει ότι

$$f_X(x) = \alpha x^{-(\alpha+1)} \qquad \text{και} \qquad f_Y(y) = \alpha y^{-(\alpha+1)} \qquad x>1,\ y>1,\ \alpha>0$$

δηλαδή και οι δυο περιθώριες κατανομές είναι μονοδιάστατες Pareto$(1,\alpha)$. Επίσης η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας δινεται από τη σχέση

$$f_{X|(Y=y)}(x|y) = (\alpha+1)y^{\alpha+1}(x+y+1)^{-(\alpha+2)}.$$

Είναι δηλαδή και αυτή μονοδιάστατη Pareto με μετατοπισμένη την αρχή. Είναι λοιπόν, άξιο παρατήρησης το ότι οι περιθώριες κατανομές, υπό συνθήκη ή μη, της διδιάστατης κατανομής Pareto τύπου Mardia I είναι τύπου Pareto. Εύκολα μπορούμε να προσδιορίσουμε τη συνάρτηση παλινδρόμησης η οποία είναι γραμμική. Συγκεκριμένα,

$$E(X|Y=y) = 1 + y/\alpha.$$

Αξίζει να σημειώσουμε ότι η $E(X|Y=y)$ υπάρχει για όλα τα $\alpha>0$, ενώ η $E(X)=\alpha/(\alpha-1)$ υπάρχει μόνο για $\alpha>1$. Η σκεδαστική συνάρτηση της X δίνεται από τη σχέση

$$var(X|Y=y) = y^2(\alpha+1)/\alpha(\alpha-1).$$

Τέλος,

$$Corr(X,Y) = 1/\alpha \qquad \text{για} \quad \alpha>2.$$

Από τη τελευταία αυτή σχέση προκύπτει ότι ο συντελεστής συσχέτισης

δεν παίρνει αρνητικές τιμές και επιπλέον είναι μικρότερος απο το 1/2. Επίσης όταν $\alpha \to \infty$ τοτε $Corr(X,Y) \to 0$.

Στο μοντέλο το οποίο περιγράψαμε στο θεώρημα 2.4.1 οι τυχαίες μεταβλητές X και Y ήταν ανεξάρτητες και η καθεμιά ακολουθούσε εκθετική (θ) κατανομή. Στη συνέχεια θα θεωρήσουμε μια γενίκευση του μοντέλου αυτού.

Εστω ότι οι X και Y είναι ανεξάρτητες εκθετικές μεταβλητές με παραμέτρους θ_1 και θ_2 αντιστοιχα. Θεωρούμε τις εξής εναλλακτικές περιπτώσεις:

(i) Οι παράμετροι θ_1, θ_2 είναι ανεξάρτητες Gamma $(\alpha, 1)$ μεταβλητές.

(ii) Οι θ_1, θ_2 είναι συσχετισμένες και ακολουθούν κάποια γνωστή διδιάστατη Gamma.

Τις δυο παραπάνω περιπτώσεις πραγμάτεύονται τα θεωρήματα 2.4.2 και 2.4.3 αντίστοιχα.

Θεώρημα 2.4.2 Εστω δυο τυχαίες μεταβλητές X και Y των οποίων η από κοινού κατανομή είναι η διπλή εκθετική με συνάρτηση κατανομής

$$F_{X,Y}(x,y) = 1 - e^{-\theta_1 x - \theta_2 y}, \qquad x,y>0.$$

Εστω ότι το τυχαίο διάνυσμα (θ_1, θ_2) ακολουθεί διπλή Gamma με συνάρτηση πυκνότητας πιθανότητας

$$f_{\theta_1,\theta_2}(\theta_1,\theta_2) = \frac{\theta_1^{a_1-1}}{\Gamma(a_1)} e^{-\theta_1} \frac{\theta_2^{a_2-1}}{\Gamma(a_2)} e^{-\theta_2}, \qquad \theta_1,\theta_2>0.$$

Τότε η κατανομή του τυχαίου διανύσματος (X,Y) έχει συνάρτηση επιβίωσης

$$\bar{F}(x,y) = (1+x)^{-a_1} (1+y)^{-a_2} \qquad \text{όπου } a_1,a_2>0.$$

Απόδειξη Με βάση τον ορισμό της συνάρτησης επιβίωσης μίγματος και λαμβάνοντας υπόψη τις υποθέσεις του θεωρήματος προκύπτει ότι

$$\bar{F}(x,y) = \int_0^\infty \int_0^\infty f_{\theta_1,\theta_2}(\theta_1,\theta_2)\, \bar{F}_{X,Y}(x,y)\, d\theta_1\, d\theta_2$$

$$= \int_0^\infty \int_0^\infty \frac{\theta_1^{a_1-1}}{\Gamma(a_1)} e^{-\theta_1} \frac{\theta_2^{a_2-1}}{\Gamma(a_2)} e^{-\theta_2} e^{-\theta_1 x-\theta_2 y}\, d\theta_1\, d\theta_2$$

$$= \int_0^\infty \frac{\theta_1^{a_1-1}}{\Gamma(a_1)} e^{-\theta_1(1+x)}\, d\theta_1 \int_0^\infty \frac{\theta_2^{a_2-1}}{\Gamma(a_2)} e^{-\theta_2(1+y)}\, d\theta_2$$

$$= (1+x)^{-a_1} (1+y)^{-a_2}.$$

Δηλαδή, στη περίπτωση κατά την οποία το τυχαίο διάνυσμα (θ_1,θ_2) ακολουθεί διπλή Gamma, η κατανομή του τυχαίου διανύσματος (X',Y') το οποίο συνδέεται με το (X,Y) με τις σχέσεις X'=X+1,

$Y'=Y+1$ είναι η διπλή Pareto.

Το θεώρημα που ακολουθεί εξετάζει την περίπτωση κατά την οποία η κατανομή του τυχαίου διανύσματος (θ_1,θ_2) είναι διδιάστατη Gamma και συγκεκριμένα η διδιάστατη Gamma του McKay (1934).

Θεώρημα 2.4.3 Εστω δυο τυχαίες μεταβλητές X και Y των οποίων η από κοινού κατανομή είναι η διπλή εκθετική με συνάρτηση κατανομής

$$F_{X,Y}(x,y) = 1-e^{-\theta_1 x-\theta_2 y}, \qquad x,y>0.$$

Εστω ότι το τυχαίο διάνυσμα (θ_1,θ_2) ακολουθεί τη διδιάστατη Gamma του McKay με συνάρτηση πυκνότητας πιθανότητας

$$f_{\theta_1,\theta_2}(\theta_1,\theta_2) = \frac{\alpha^{\beta+\gamma}}{\Gamma(\beta)\Gamma(\gamma)}\ \theta_1^{\beta-1}\ (\theta_2-\theta_1)^{\gamma-1}\ e^{-\alpha\theta_2},$$

$$\alpha,\beta,\gamma>0 \quad \text{και} \quad \theta_2>\theta_1>0.$$

Τότε η κατανομή του τυχαίου διανύσματος (X,Y) έχει συνάρτηση επιβίωσης

$$\bar{F}(x,y) = \alpha^{\beta+\gamma}\ (\alpha+y)^{-\gamma}\ (\alpha+y+x)^{-\beta}.$$

Απόδειξη Με βάση τον ορισμό της συνάρτησης επιβίωσης μίγματος και τις υποθέσεις του θεωρήματος έχουμε ότι

$$\bar{F}(x,y) = \int_0^{\infty}\int_0^{\theta_2} \frac{\alpha^{\beta+\gamma}}{\Gamma(\beta)\Gamma(\gamma)}\ \theta_1^{\beta-1}(\theta_2-\theta_1)^{\gamma-1}\ e^{-\alpha\theta_2}\ e^{-\theta_1 x-\theta_2 y}\ d\theta_1\ d\theta_2$$

$$= \int_0^{\infty} \frac{\alpha^{\beta+\gamma}}{\Gamma(\beta)\Gamma(\gamma)} e^{\theta_2(-\alpha-y)} \int_0^{\theta_2} \theta_1^{\beta-1}(\theta_2-\theta_1)^{\gamma-1} e^{-\theta_1 x} d\theta_1 d\theta_2$$

$$= \frac{\alpha^{\beta+\gamma}}{\Gamma(\beta)\Gamma(\gamma)} \int_0^{\infty} e^{\theta_2(-\alpha-y)} \theta_2^{\beta+\gamma-1} \int_0^1 t^{\beta-1}(1-t)^{\gamma-1} e^{t(-\theta_2 x)} dt d\theta_2$$

όπου $t = \theta_1/\theta_2$.

Αλλά,

$$\int_0^1 t^{\beta-1} (1-t)^{\gamma-1} e^{t(-\theta_2 x)} dt = B(\beta,\gamma) \; {}_1F_1(\beta, \gamma+\beta, -\theta_2 x)$$

όπου $B(\beta,\gamma)=\frac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)}$ είναι η συνάρτηση Beta

και ${}_1F_1(a;b;z)=\sum_{r=0}^{\infty} \frac{a_{(r)}}{b_{(r)}} \frac{z^r}{r!}$ είναι η confluent υπεργεωμετρική

συνάρτηση. Κατά συνέπεια,

$$\overline{F}(x,y) = \frac{\alpha^{\beta+\gamma}}{\Gamma(\beta+\gamma)} \sum_{r=0}^{\infty} \frac{\beta_{(r)}}{(\beta+\gamma)_{(r)}} \int_0^{\infty} \theta_2^{\beta+\gamma-1} \frac{(-\theta_2 x)^r}{r!} e^{\theta_2(-\alpha-y)} d\theta_2$$

$$= \frac{\alpha^{\beta+\gamma}}{\Gamma(\beta+\gamma)} \sum_{r=0}^{\infty} \frac{\beta_{(r)}}{(\beta+\gamma)_{(r)}} \frac{(-x)^r}{r!} \int_0^{\infty} \theta_2^{\beta+\gamma+r-1} e^{-\theta_2(\alpha+y)} d\theta_2$$

$$= \frac{\alpha^{\beta+\gamma}}{\Gamma(\beta+\gamma)} \sum_{r=0}^{\infty} \frac{\beta_{(r)}}{(\beta+\gamma)_{(r)}} \frac{(-x)^r}{r!} \frac{\Gamma(\beta+\gamma+r)}{(\alpha+y)^{\beta+\gamma+r}}$$

$$= \left(\frac{\alpha}{\alpha+y}\right)^{\beta+\gamma} \sum_{r=0}^{\infty} \frac{\beta_{(r)}}{(\beta+\gamma)_{(r)}} \left(\frac{-x}{\alpha+y}\right)^r \frac{1}{r!} \frac{\Gamma(\beta+\gamma+r)}{\Gamma(\beta+\gamma)}$$

$$= \left(\frac{\alpha}{\alpha+y}\right)^{\beta+\gamma} \sum_{r=0}^{\infty} \beta_{(r)} \frac{z^r}{r!}$$

$$= \left(\frac{\alpha}{\alpha+y}\right)^{\beta+\gamma} {}_2F_1(.,\beta;.;z)$$

όπου $z = -x/(\alpha+y)$ και ${}_2F_1(a,b;c;z)$ είναι η υπεργεωμετρική σειρά του Gauss η οποία δίνεται από τον τύπο

$$\sum_{r=0}^{\infty} \frac{a_{(r)}\, b_{(r)}\, z^r}{c_{(r)}\, r!} = {}_2F_1(a,b;c;z)$$ και η οποία στην ειδική περίπτωση

που a=c=d παίρνει τη μορφή ${}_2F_1(d,b;d;z) = (1-z)^{-b}$. Αρα τελικά

$$\overline{F}(x,y) = \frac{\alpha^{\beta+\gamma}}{(\alpha+y)^{\beta+\gamma}} \left(1+\frac{x}{\alpha+y}\right)^{-\beta}$$

$$= \alpha^{\beta+\gamma} (\alpha+y)^{-\gamma} (\alpha+y+x)^{-\beta}.$$

Παρατηρούμε ότι η συνάρτηση επιβίωσης του τυχαίου διανύσματος (X',Y') το οποίο συνδέεται με το (X,Y) με τις σχέσεις X'=X και Y'=Y+α−1 είναι γινόμενο της συνάρτησης κατανομής μιας μονοδιάστατης κατανομής Pareto και αυτής μιας διδιάστατης κατανομής Pareto τύπου Mardia I.

Επανερχόμενοι στο θεωρήμα 2.4.1 παρατηρούμε ότι η διδιάστατη κατανομή Pareto τύπου Mardia I προκύπτει σαν Gamma μίγμα διπλής εκθετικής κατανομής. Η διδιάστατη δηλαδή αυτή κατανομή διατηρεί την ιδιότητα της μονοδιάστατης Pareto ως προς τον τρόπο γένεσης. Η παραπάνω διαδικασία μεσω της οποίας προκύπτει η διδιάστατη Pareto τύπου Mardia I υποδηλώνει ενδεχόμενη χρησιμότητα της κατανομής στην ανάλυση πινάκων συναφείας στους οποίους οι δύο διαστάσεις μπορούν να θεωρηθούν σαν εναλλακτικά μέτρα του ίδιου μεγέθους. Η μελέτη της κατανομής των τραυμάτων δύο οδηγών σε ένα τροχαίο δυστύχημα ή της σοβαρότητας της σωματικής βλάβης ενός ασθενούς όπως αυτή αξιολογείται από δύο γιατρούς αποτελούν δύο ενδεικτικές πιθανές εφαρμογές της κατανομής αυτής.

Στα μίγματα που εξετάσαμε μέχρι τώρα, οι τυχαίες μεταβλητές ήταν ανεξάρτητες και επομένως η από κοινού κατανομή τους ήταν διπλή εκθετική μιάς ή δύο παραμέτρων. Το θεώρημα 2.4.4 αντιμετωπίζει την περίπτωση κατά την οποία η από κοινού κατανομή τους είναι μία από τις γνωστές διδιάστατες εκθετικές κατανομές και συγκεκριμένα εκείνη που ορίσθηκε από τους Marshall και Olkin (1967).

Θεώρημα 2.4.4 Εστω (X,Y) τυχαίο διάνυσμα του οποίου η κατανομή είναι η διδιάστατη εκθετική των Marshall-Olkin με συνάρτηση κατανομής την

$$F_{X,Y}(x,y) = 1-e^{-(k+r)x}\ -e^{-(1+r)y}\ -e^{-kx-ly-r\max(x,y)}$$

$$x>0,\ y>0.$$

Εστω επίσης ότι $k=\theta$, $l=\theta$, $r=R\theta$ όπου το θ ακολουθεί Gamma $(\alpha,1)$ κατανομή. Τότε το τυχαίο διάνυσμα (X,Y) ακολουθεί διδιάστατη Pareto με συνάρτηση επιβίωσης

$$\overline{F}(x,y) = \left[1+x+y+R\max(x,y)\right]^{-\alpha}, \qquad R = k+l+r.$$

Απόδειξη Με βάση τον ορισμό της συνάρτησης κατανομής μίγματος και τις υποθέσεις του θεωρήματος έχουμε ότι

$$F(x,y) = \int_0^\infty \left\{1-e^{-\theta(1+R)x}\ -e^{-\theta(1+r)y}\ -e^{-\theta[x+y+R\max(x,y)]}\right\}\left\{\frac{\theta^{\alpha-1}}{\Gamma(\alpha)}\ e^{-\theta}\right\}\ d\theta$$

$$= 1-\int_0^\infty \frac{\theta^{\alpha-1}}{\Gamma(\alpha)}\ e^{-\theta[1+(1+R)x]}\ d\theta\ -\int_0^\infty \frac{\theta^{\alpha-1}}{\Gamma(\alpha)}\ e^{-\theta[1+(1+R)y]}\ d\theta$$

$$-\int_0^\infty \frac{\theta^{\alpha-1}}{\Gamma(\alpha)}\ e^{-\theta[1+x+y+R\max(x,y)]}\ d\theta$$

$$= 1-\left[1+(1+R)x\right]^{-\alpha}-\left[1+(1+R)y\right]^{-\alpha}-\left[1+x+y+R\max(x,y)\right]^{-\alpha}$$

Επειδή δε, $\bar{F}(x,y) = F(x,y) - F(\infty,y) - F(x,\infty) + 1$, έπεται ότι,

$$\bar{F}(x,y) = \left[1+x+y+Rmax(x,y)\right]^{-\alpha}$$

Το θεώρημα 2.4.4 αποδεικνύει ότι μίγμα διδιάστατης εκθετικής κατανομής τύπου Marshall-Olkin με Gamma οδηγεί σε μορφή διδιάστατης κατανομής Pareto. Δεν συμβαίνει το ίδιο αν χρησιμοποιηθεί η διδιάστατη εκθετική κατανομή του Gumbel (1960). Συγκεκριμένα, αν θεωρήσουμε τη διδιάστατη εκθετική με συνάρτηση επιβίωσης

$$\bar{F}_{X,Y}(x,y) = e^{-x-y-\theta xy}, \qquad x \geq \theta,\ y > 0,$$
$$0 \leq \theta \leq 1$$

και θεωρήσουμε ότι η παράμετρος αυτής θ μεταβάλλεται ακολουθώντας κατανομή Beta τύπου I με συνάρτηση πυκνότητας πιθανότητας

$$f_{\theta}(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\theta^{\alpha-1}\,(1-\theta)^{\beta-1} \qquad \alpha>0,\ \beta>0,$$

τότε η συνάρτηση επιβίωσης του μίγματος είναι

$$\bar{F}(x,y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,e^{-(x+y)}\int_0^1 \theta^{\alpha-1}\,(1-\theta)^{\beta-1}\,e^{\theta(-xy)}\,d\theta$$

$$= e^{-(x+y)}\,{}_1F_1(\alpha,\beta+\alpha;-xy)$$

$$= e^{-(x+y+xy)} \; {}_1F_1(\beta, \beta+\alpha; xy)$$

όπου ${}_1F_1(a;b;z)$ είναι η confluent υπεργεωμετρική σειρά.

2.5 ΓΕΝΙΚΕΥΜΕΝΕΣ ΜΟΡΦΕΣ ΔΙΔΙΑΣΤΑΤΩΝ ΚΑΤΑΝΟΜΩΝ PARETO

Στο τέλος της ενότητας 2.3.2 αναφέραμε ότι η διδιάστατη Pareto τύπου Marshall-Olkin δεν είναι απόλυτα συνεχής. Αντίθετα η συνάρτηση κατανομής (2.3.3) μπορεί να διασπασθεί σε ένα απόλυτα συνεχές και σε ένα ιδιάζον μέρος όπως φαίνεται από τη σχέση (2.3.4). Η Dimaki,C. (1977) αποδεικνύει ότι η διδιάστατη Pareto τύπου Marshall-Olkin είναι απόλυτα συνεχής στην ειδική μόνο περίπτωση που οι τυχαίες μεταβλητές X και Y είναι ανέξαρτητες. Μπορούμε όμως, και ίσως παρουσιάζει κάποιο ενδιαφέρον, να ορίσουμε διδιάστατη Pareto με συνάρτηση επιβίωσης εκείνην της διδιάστατης Pareto τύπου Masrhall-Olkin απαλλαγμένη όμως από το ιδιάζον μέρος. Ονομάζουμε μια τέτοια κατανομή **γενικευμένη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin** και η συνάρτηση κατανομής της δίνεται από τη σχέση

$$\bar{F}(x,y) = \frac{R}{k+1}\left\{x^{-k}\, y^{-l} \left[\max(x,y)\right]^{-r}\right\} - \frac{r}{k+1}\left[\max(x,y)\right]^{-R} \qquad (2.4.1)$$

$$x,y \geq 1 \quad \text{και} \quad R=k+l+r \; .$$

Οι περιθώριες κατανομές της (2.4.1) δίνονται από τις παρακάτω δυο

σχέσεις

$$\bar{F}_X(x) = \frac{R}{k+1} x^{-k-r} - \frac{r}{k+1} x^{-R} \qquad x \geq 1$$

$$\bar{F}_Y(y) = \frac{R}{k+1} y^{-l-r} - \frac{r}{k+1} y^{-R} \qquad y \geq 1.$$

Οπως είναι φυσικό, οι περιθώριες συναρτήσεις κατανομής δεν είναι τύπου Pareto, αλλά σταθμικοί μέσοι συναρτήσεων κατανομών Pareto.

Τελειώνοντας, αξίζει να αναφέρουμε τον ορισμό μιάς ακόμη γενικευμένης διδιάστατης κατανομής τύπου Pareto. Η συνάρτηση πυκνότητας πιθανότητας της κατανομής αυτής που ορίσθηκε από τον Arnold (1987) είναι

$$f_{X,Y}(x,y) = \text{constant.}\ (1+\lambda_1 x+\lambda_2 y+\lambda_{12} xy)^{-(\alpha+2)}$$

$$\text{με} \quad \lambda_1,\lambda_2>0 \quad \text{και} \quad \lambda_{12} \geq 0 \qquad \text{αν} \qquad \alpha>1$$

$$\text{ή} \quad \lambda_{12} > 0 \qquad \text{αν} \quad 0 < \alpha \leq 1.$$

Το ιδιαίτερο χαρακτηριστικό της κατανομής αυτής είναι ότι οι υπό συνθήκη περιθώριες κατανομές της είναι Pareto $(1,\alpha)$. Παρατηρούμε επίσης ότι αν $\lambda_{12} = \lambda_1\lambda_2$ τότε προκύπτει η διπλή Pareto, ενώ αν $\lambda_{12} = 0$ προκύπτει η διδιάστατη Pareto τύπου Mardia I.

2.6 ΜΙΓΜΑΤΑ ΔΙΔΙΑΣΤΑΤΩΝ ΚΑΤΑΝΟΜΩΝ PARETO

Στην ενότητα αυτή θα ασχοληθούμε με τις κατανομές οι οποίες προκύπτουν όταν οι παράμετροι των διδιάστατων κατανομών Pareto είναι τυχαίες μεταβλητές. Οπως θα παρατηρήσουμε οι κατανομές στις οποίες οδηγούμαστε δεν είναι ούτε γνωστές ούτε εύχρηστες. Για την πληρότητα όμως της παρουσίασης παραθέτουμε ενδεικτικά κάποια συμπεράσματα. Επιπλέον για να τονίσουμε ότι το πρόβλημα δεν εντοπίζεται στις δυο διάστάσεις αλλά είναι εξίσου έντονο και στη μονοδιάστατη περίπτωση ξεκινάμε υπολογίζοντας τη συνάρτηση επιβίωσης μίγματος Pareto $(1,\alpha)$ με Gamma $(\beta,1)$.

$$\bar{F}(x) = \int_0^\infty \frac{\alpha^{\beta-1}}{\Gamma(\beta)} e^{-\alpha} x^{-\alpha} \, d\alpha$$

$$= \int_0^\infty \frac{\alpha^{\beta-1}}{\Gamma(\beta)} e^{-(1+\ln x)} \, d\alpha$$

$$= (1+\ln x)^{-\beta} \int_0^\infty \frac{[\alpha(1+\ln x)]^{\beta-1}}{\Gamma(\beta)} e^{-\alpha(1+\ln x)} \, d[\alpha(1+\ln x)]$$

$$= (1+\ln x)^{-\beta}$$

δηλαδή $(1+\ln X) \sim$ Pareto $(1,\beta)$.

Για να μεταφέρουμε το συμπέρασμα στις δυο διαστάσεις υπολογίζουμε τη συνάρτηση επιβίωσης του παρακάτω μίγματος

Διδιάστατη Pareto τύπου Mardia I $\overset{\alpha}{\wedge}$ Gamma(β,1)

$$\bar{F}(x,y) = \int_0^1 \frac{\alpha^{\beta-1}}{\Gamma(\beta)} e^{-\alpha} (x+y+1)^{-\alpha} d\alpha$$

$$= \int_0^\infty \frac{\alpha^{\beta-1}}{\Gamma(\beta)} e^{-\alpha} [1+\ln (x+y+1)] d\alpha$$

$$= \left[1+\ln (x+y+1)\right]^{-\beta}$$

δηλαδή, $1+\ln (X+Y+1) \sim$ Pareto $(1,\beta)$.

Στη συνέχεια θα εξετάσουμε περιπτώσεις στις οποίες η παράμετρος τόσο της μονοδιάστατης όσο και της διδιάστατης κατανομής Pareto ακολουθεί την κατανομή Beta. Ξεκινώντας λοιπόν από τη μονοδιάστατη περίπτωση, υπολογίζουμε τη συνάρτηση επιβίωσης μίγματος Pareto $(1,\alpha)$ με Beta II (β,γ).

$$\bar{F}(x) = \int_0^\infty \frac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)} \alpha^{\beta-1} (1+\alpha)^{-(\beta+\gamma)} x^{-\alpha} d\alpha$$

$$= \frac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)} \int_0^\infty \alpha^{\beta-1} \ (1+\alpha)^{-(\beta+\gamma)} \ e^{-\alpha \ln x} \ d\alpha$$

$$= \frac{\Gamma(\beta+\gamma)}{\Gamma(\gamma)} \ \Psi(\beta, 1-\gamma; \ln x)$$

όπου $\Psi(a;c;z) = \frac{1}{\Gamma(a)} \int_0^\infty t^{a-1} \ (1+t)^{c-a-1} \ e^{-zt} \ dt$.

Επειδή η υπεργεωμετρική σειρά $\Psi(a;c;z)$ συνδέεται με την confluent υπεργεωμετρική σειρά ${}_1F_1(a;c;z)$ με τη σχέση

$$\Psi(a,c;z) = \frac{\Gamma(1-c)}{\Gamma(1+a-c)} \ {}_1F_1(a;c;z)$$

καταλήγουμε στο συμπέρασμα ότι

$$\overline{F}_X(x) = {}_1F_1(\beta; 1-\gamma; \ln x).$$

Τέλος θα υπολογίσουμε τη συνάρτηση επιβίωσης του παρακάτω μίγματος

Διπλή Pareto $\overset{(\theta_1,\theta_2)}{\wedge}$ **Bivariate Beta II**

$$\overline{F}(x,y) = \int_0^\infty \int_0^\infty \frac{\Gamma(k+m+\rho)}{\Gamma(k)\Gamma(m)\Gamma(\rho)} \ \theta_1^{k-1} \ \theta_2^{m-1} \ (1+\theta_1+\theta_2)^{-(k+m+\rho)} x^{-\theta_1} \ y^{-\theta_2} d\theta_1 d\theta_2$$

$$= \frac{\Gamma(k+m+\rho)}{\Gamma(k)\Gamma(m)\Gamma(\rho)} \int_0^\infty \int_0^\infty \theta_1^{k-1}\, \theta_2^{m-1}\, (1+\theta_1+\theta_2)^{-(k+m+\rho)} x^{-\theta_1}\, y^{-\theta_2} d\theta_1 d\theta_2$$

$$= \frac{\Gamma(k+m+\rho)}{\Gamma(k)\Gamma(m)\Gamma(\rho)} \frac{\Gamma(1-\rho-k-m)\Gamma(k)\Gamma(m)}{\Gamma(1-\rho)} \Phi_2(k,m;1-\rho;\ln x, \ln y)$$

όπου

$$\Phi_2(a,b;c;z,w) = \frac{\Gamma(c)}{\Gamma(c-a-b)\Gamma(a)\Gamma(b)} \int_0^\infty \int_0^\infty e^{-xz} e^{-yw}\, x^{a-1} y^{b-1} (1+x+y)^{c-a-b-1}\, dxdy$$

είναι η διμεταβλητή confluent υπεργεωμετρική σειρά.

Επειδή,

$$\frac{\Gamma(1-\rho-k-m)}{\Gamma(1-\rho)} = \frac{(-\rho-k-m)!}{(-\rho)!}$$

$$= \frac{(\rho-1)!}{(\rho+k+m-1)!} (-1)^{k+m}$$

έπεται ότι

$$\overline{F}(x,y) = (-1)^{k+m} \Phi_2(k,m;1-\rho;\ln x, \ln y).$$

2.7 ΑΛΛΕΣ ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ ΣΥΝΔΕΟΜΕΝΕΣ ΜΕ ΤΙΣ ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ PARETO

Υπάρχουν πολλές σημαντικές κατανομές στενά συνδεδεμένες με τη μονοδιάστατη Pareto $(1,\alpha)$. Είναι γνωστές οι συναρτησιακές σχέσεις που υπάρχουν μεταξύ της μεταβλητής Pareto και μεταβλητών που προέρχονται από κατανομές όπως η εκθετική, η Weibull, η power function, η κατανομή Burr ή η λογιστική κατανομή. Είναι επίσης γνωστό ότι η ύπαρξη των συναρτησιακών αυτών σχέσεων βοηθάει στο να μεταφέρουμε υπάρχοντα συμπεράσματα της κατανομής Pareto, κυρίως χαρακτηριστικές ιδιότητες, σε συμπεράσματα που να ισχύουν για τις συνδεόμενες με αυτήν κατανομές.

Στην ενότητα αυτή θα επιχειρήσουμε τη μεταφορά των ισοδυναμιών μεταξύ Pareto και άλλων κατανομών στο χώρο των δυο διαστάσεων. Θα συγκεντρώσουμε την προσοχή μας στη διδιάστατη Pareto τύπου Mardia I με συνάρτηση πυκνότητας πιθανότητας

$$f_{X,Y}(x,y) = \alpha(\alpha+1)(x+y-1)^{-(\alpha+1)} \qquad (2.7.1)$$

όπου $x>1, y>1, \alpha>0$

κυρίως διότι η διδιάστατη αυτή κατανομή διατηρεί πολλές από τις ιδιότητες της μονοδιάστατης Pareto.

Οι ισοδυναμίες μεταξύ των κατανομών που προαναφέρθηκαν θεμελιώνονται με τα παρακάτω πέντε θεωρήματα. Το θεώρημα 2.7.1 καθορίζει τη συναρτησιακή σχέση που υπάρχει μεταξύ τυχαίου διανύσματος προερχομένου από διδιάστατη κατανομή Pareto τύπου Mardia I και διδιάστατης εκθετικής κατανομής. Αξίζει να

σημειωθεί ότι η διδιάστατη εκθετική στην οποία οδηγούμεθα δεν αναφέρεται στη βιβλιογραφία και κατά συνέπεια δέν έχει μελετηθεί.

Θεώρημα 2.7.1 Εστω (X,Y) τυχαίο διάνυσμα που ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I. Τότε το τυχαίο διάνυσμα (Z=lnX, W=lnY) ακολουθεί διδιάστατη εκθετική κατανομή με συνάρτηση πυκνότητας πιθανότητας την

$$f_{Z,W}(z,w) = \alpha(\alpha+1)\, e^{z+w}\, (e^{z}+e^{w}-1)^{-(\alpha+2)} \qquad (2.7.2)$$

όπου $z,w \geq 0$ $\alpha > 0$.

Απόδειξη Εστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με συνάρτηση πυκνότητας πιθανότητας την (2.7.1). Με εφαρμογή του θεωρήματος αλλαγής μεταβλητής δεδομένου ότι η Ιακωβιανή του μετασχηματισμού $Z = \ln X$, $W = \ln Y$ δίνεται από τον τύπο $J = e^{Z+W}$ προκύπτει εύκολα η (2.7.2).

Η κατανομή αυτή είναι διδιάστατη εκθετική αφού

$$f_Z(z) = \alpha\, e^{-\alpha z} \quad \text{και} \quad f_W(w) = \alpha\, e^{-\alpha w}.$$

Είναι δηλαδή και οι δύο περιθώριες κατανομές μονοδιάστατες εκθετικές (α).

Θεώρημα 2.7.2 Εστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί διδιάστατη κατανομή Pareto τύπου Mardia I με συνάρτηση πυκνότητας πιθανότητας την (2.7.1). Τότε το τυχαίο διάνυσμα (Z,W) όπου

$X=\exp(Z-\theta)^c$ και $Y=\exp(W-\theta)^c$, όπου $Z>\theta$, $W>\theta$, $\theta \geq 0$, $c>-1$, ακολουθεί διδιάστατη κατανομή Weibull με συνάρτηση πυκνότητας πιθανότητας την

$$f_{Z,W}(z,w)=$$

$$=\alpha(\alpha+1)c^2\left[(z-\theta)(w-\theta)\right]^{c-1}\exp\left[(z-\theta)^c+(w-\theta)^c\right]\left[e^{(z-\theta)^c}+e^{(w-\theta)^c}-1\right]^{-(\alpha+2)} \qquad (2.7.3)$$

όπου $z,w>\theta$, $\theta \geq 0$, $c>-1$.

<u>Απόδειξη</u> Εφαρμόζοντας τον μετασχηματισμό $Z = \theta+(\ln X)^{1/c}$, $W = \theta+(\ln Y)^{1/c}$ στην (2.7.1) και λαμβάνοντας υπόψη ότι η Ιακωβιανή είναι $J = c^2(z-\theta)^{c-1}(w-\theta)^{c-1}\exp(z-\theta)^c \exp(w-\theta)^c$ καταλήγουμε στο ζητούμενο. Η διδιάστατη κατανομή που έχει συνάρτηση πυκνότητας πιθανότητας την (2.7.3) είναι διδιάστατη Weibull αφού και οι δυο περιθώριες κατανομές είναι μονοδιάστατες Weibull.

Συνεχίζουμε ορίζοντας, με αντίστοιχο τρόπο, τη διδιάστατη κατανομή Burr.

<u>Θεώρημα 2.7.3</u> Εστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με συνάρτηση πυκνότητας πιθανότητας την (2.7.1). Τότε το τυχαίο διάνυσμα (Z,W) όπου $X = 1+Z^{\beta}$ και $Y = 1+W^{\beta}$, $\beta>1$ ακολουθεί τη διδιάστατη κατανομή Burr με συνάρτηση πυκνότητας πιθανότητας

$$f_{Z,W}(z,w) = \alpha(\alpha+1)\beta^2\, z^{\beta-1}\, w^{\beta-1}\, (1+z^{\beta}+w^{\beta})^{-(\alpha+2)} \qquad (2.7.4)$$

$$z,w\geq 0,\ \alpha>0,\ \beta>1.$$

<u>Απόδειξη</u> Η συνάρτηση κατανομής του τυχαίου διάνυσμα (X,Y) δίνεται από τον τύπο

$$F_{X,Y}(x,y) = (x+y-1)^{\alpha} - x^{-\alpha} - y^{-\alpha} + 1 \qquad x\geq 1,\ y\geq 1,\ \alpha>0\ .$$

Τότε, η συνάρτηση κατανομής του τυχαίου διανύσματος (Z,W) δίνεται από τη σχέση

$$F_{Z,W}(z,w) = \left(1+z^{\beta}+w^{\beta}\right)^{-\alpha} - \left(1+z^{\beta}\right)^{-\alpha} - \left(1+w^{\beta}\right)^{-\alpha} + 1.$$

η οποία συνεπάγεται ότι η $f_{Z,W}(z,w)$ δίνεται από τη σχέση (2.7.4). Η διδιάστατη κατανομή που ορίζεται με αυτό τον τρόπο είναι μια διδιάστατη κατανομή Burr αφού

$$F_Z(z) = F_{Z,W}(z,\infty) = 1 - \left(1+z^{\beta}\right)^{-\alpha}, \qquad \alpha>0,\ \beta>1,\ z\geq 0$$

και

$$F_W(w) = F_{Z,W}(\infty,w) = 1 - \left(1+w^{\beta}\right)^{-\alpha}, \qquad \alpha>0,\ \beta>1,\ w\geq 0.$$

Είναι δηλαδή και οι δυο περιθώριες κατανομές μονοδιάστατες Burr.

Το θεώρημα που ακολουθεί καθορίζει τη σχέση που υπάρχει μεταξύ τυχαίων διανυσμάτων προερχομένων από τη διδιάστατη κατανομή Pareto τύπου Mardia I και από εκείνη την οποία ονομάζουμε διδιάστατη power function κατανομή.

Θεώρημα 2.7.4 Εστω (X,Y) τυχαίο διάνυσμα που ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με συνάρτηση πυκνότητας πιθανότητας την (2.7.1). Τότε το τυχαίο διάνυσμα $(Z=X^{-1},W=Y^{-1})$ ακολουθεί τη διδιάστατη power function κατανομή με συνάρτηση πυκνότητας πιθανότητας την

$$f_{Z,W}(z,w) = \alpha(\alpha+1)\, z^{\alpha}\, w^{\alpha}\, (z+w-zw)^{-(\alpha+2)}, \qquad 0<z\le 1,\ 0<w\le 1,\ \alpha>0. \tag{2.7.5}$$

Απόδειξη Ισχύει ότι

$$F_{X,Y}(x,y) = (x+y-1)^{-\alpha} - x^{-\alpha} - y^{-\alpha} + 1 \qquad x\ge 1,\ y\ge 1,\ \alpha>0.$$

Τότε

$$F_{Z,W}(z,w) = \left(z^{-1} + w^{-1} - 1\right)^{-\alpha} - z^{\alpha} - w^{\alpha} + 1$$

και άρα η $f_{Z,W}(z,w)$ δίνεται από τη σχέση (2.7.5). Η κατανομή που ορίζεται με αυτό τον τρόπο είναι μια διδιάστατη power function κατανομή αφού

$$F_Z(z) = 1 - z^{-\alpha} \qquad \text{και} \qquad F_W(w) = 1 - w^{-\alpha}$$

είναι δηλαδή και οι δυο περιθώριες μονοδιάστατες power function κατανομές.

θα τελειώσουμε την ενότητα ορίζοντας, με αντίστοιχο τρόπο, τη διδιάστατη λογιστική κατανομή.

Θεώρημα 2.7.5 Εστω το τυχαίο διάνυσμα (X,Y) που ακολουθεί διδιάστατη κατανομή Pareto τύπου Mardia I με συνάρτηση πυκνότητας πιθανότητας τη (2.7.1) και παράμετρο α=1. Τότε το τυχαίο διάνυσμα (Z,W) όπου X = 1 + exp(-Z), Y = 1 + exp(-W) ακολουθεί διδιάστατη λογιστική κατανομή με συνάρτηση πυκνότητας πιθανότητας

$$f_{Z,W}(z,w) = 2\, e^{-(z+w)} \left(e^{-z} + e^{-w} + 1\right)^{-3}$$

Απόδειξη Ομοια με την προηγούμενη.

2.8 ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ YULE

Οπως ήδη αναφέρθηκε στην εισαγωγή, με βάση τα όσα αποδεικνύονται από τον Irwin (1975) προκύπτει ότι η μονοδιάστατη Pareto είναι το συνεχές ανάλογο της κατανομής Yule. Η Xekalaki (1984a) μεταφέρει και αποδεικνύει την ισχύ της ιδιότητας αυτής στο χώρο των δυο διαστάσεων. Συγκεκριμένα αποδεικνύει ότι το συνεχές ανάλογο της διδιάστατης γενικευμένης κατανομής Waring της οποίας η συνάρτηση πιθανότητας δίνεται από την

$$p_{x,y} \equiv P(X=x, Y=y) = \frac{\rho_{(k+m)}}{(a+\rho)_{(k+m)}} \frac{a_{(x+y)}\, k_{(x)}\, m_{(y)}}{(a+k+m+\rho)_{(x+y)}} \frac{1}{x!} \frac{1}{y!} \qquad (2.8.1)$$

όπου $a>0, k>0, m>0$ και $\rho>0$ με $x=0,1,2,\ldots$ $y=0,1,2,\ldots$

και $\alpha_{(\beta)} = \Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)$, $\alpha>0, \beta \in R$.

είναι η διδιάστατη Beta II με συνάρτηση πυκνότητας πιθανότητας την

$$f_{X,Y}(x,y) = \frac{\Gamma(\alpha+\beta+\rho)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\rho)} x^{\alpha-1} y^{\beta-1} (x+y+1)^{-(\alpha+\beta+\rho)} \qquad (2.8.2)$$

όπου $\alpha, \beta>0$ και $x, y>0$.

Επιπλέον η Xekalaki (1986) ονομάζει την κατανομή του τυχαίου διανύσματος (X,Y) του οποίου η συνάρτηση πιθανότητας δίνεται από την (2.8.1) για $a=k=m=1$ διδιάστατη Yule με παράμετρο ρ. Η συνάρτηση πιθανότητας της δι-διάστατης κατανομής Yule έτσι όπως αυτή ορίζεται στο άρθρο που προαναφέραμε δίνεται επομένως από τη σχέση

$$p_{x,y} = P(X=x, Y=y) = \frac{\rho_{(2)}}{(\rho+1)_{(x+y+2)}} (x+y)! \qquad x,y=0,1,2,\ldots \quad \rho>0. \qquad (2.8.3)$$

Ομως, η ειδική περίπτωση της (2.8.2) για $\alpha=\beta=1$ είναι η σχέση (2.3.8) δηλαδή η συνάρτηση πυκνότητας πιθανότητας της διδιάστατης κατανομής Pareto τύπου Mardia I. Κατά συνέπεια η διδιάστατη Yule έτσι όπως ορίζεται από την (2.8.3) είναι το διακριτό ανάλογο της

διδιάστατης Pareto τύπου Mardia I. Οπως θα παρατηρηθεί στα επόμενα η δυαδικότητα αυτή που υπάρχει μεταξύ των δυο αυτών κατανομών εδραιώνεται μέσω της αναλογίας που υπάρχει μεταξύ των χαρακτηριστικών ιδιοτήτων των δυο κατανομών.

Η διδιάστατη Yule είναι η φυσική επέκταση στο χώρο των δυο διαστάσεων της μονοδιάστατης κατανομής Yule αφού και οι δυο περιθώριες κατανομές της είναι μονοδιάστατες Yule (ρ). Πριν προχωρήσουμε στη διατύπωση ορισμένων βασικών ιδιοτήτων της διδιάστατης κατανομής Yule θα παρεμβάλλουμε τον ορισμό μιας ακόμη διδιάστατης κατανομής η οποία προκύπτει σαν ειδική περίπτωση της (2.8.1). Ειδικώτερα η Xekalaki (1986) ονομάζει την κατανομή του τυχαίου διανύσματος (X,Y) του οποίου η συνάρτηση πιθανότητας δίνεται από την (2.8.1) για k=m=1 διδιάστατη Waring με παραμέτρους (a;ρ). Η συνάρτηση πιθανότητας της διδιάστατης κατανομής Waring δίνεται από τη σχέση

$$p_{x,y} \equiv P(X=x,Y=y) = \frac{\rho_{(2)}}{(a+\rho)_{(2)}} \frac{a_{(x+y)}}{(a+\rho+2)_{(x+y)}}, \quad x,y=0,1,2,\ldots \qquad (2.8.4)$$

$$\rho>0$$

Προφανώς ισχύει ότι:

Διδιάστατη κατανομή Waring (1;ρ) ~ Διδιάστατη κατανομή Yule (ρ).

Μερικές βασικές ιδιότητες της διδιάστατης κατανομής Yule συνοψίζονται στο παρακάτω θεώρημα.

Θεώρημα 2.8.1 Εστω το τυχαίο διάνυσμα (X,Y) με ακέραιες και μη αρνητικές συνιστώσες το οποίο ακολουθεί τη διδιάστατη κατανομή Yule (ρ). Τότε

(i) X ~ μονοδιάστατη Yule (ρ) , Y ~ μονοδιάστατη Yule (ρ).

(ii) X+Y ~ μονοδιάστατη γενικευμένη κατανομή Waring (1,2;ρ).

(iii) X|(Y=y) ~ μονοδιάστατη κατανομή Waring (y+1,ρ+1)

Y|(X=x) ~ μονοδιάστατη κατανομή Waring (x+1,ρ+1).

Για την απόδειξη του θεωρήματος παραπέμπουμε στο άρθρο της Xekalaki (1984a) όπου αποδεικνύεται γενικότερο θεώρημα για τη διδιάστατη γενικευμένη κατανομή Waring. Αξίζει να παρατηρήσουμε πως από την (iii) προκύπτει ότι οι συναρτήσεις παλινδρόμησης της X επί της Y και της Y επί της X είναι γραμμικές.

Οι παραγοντικές ροπές της διδιάστατης κατανομής Yule δίνονται από τη σχέση

$$\mu_{(i,j)} = \frac{(i+j)!\ i!\ j!}{(\rho-1)(\rho-2)\ldots(\rho-i-j)} , \qquad i,j=0,1,2,\ldots$$

Παρατηρούμε ότι για ρ≤i+j οι παραγοντικές ροπές και κατά συνέπεια και οι ροπές απειρίζονται. Μέσω των γνωστών μετασχηματισμών προκύπτει ότι

$$\mu'_{1,0} \equiv \mu_X = \frac{1}{\rho-1} , \qquad \mu'_{0,1} \equiv \mu_Y = \frac{1}{\rho-1}$$

$$\mu_{1,1} \equiv \sigma_{XY} = \frac{\rho}{(\rho-1)^2(\rho-2)}$$

$$\mu_{2,0} \equiv \sigma_X^2 = \frac{\rho^2}{(\rho-1)^2(\rho-2)} , \qquad \mu_{0,2} \equiv \sigma_Y^2 = \frac{\rho^2}{(\rho-1)^2(\rho-2)} .$$

Επειδή θα πρέπει να ισχύει $\rho>2$ προκύπτει ότι η συνδυακύμανση σ_{XY} είναι πάντα θετική και κατά συνέπεια οι τυχαίες μεταβλητές X,Y είναι πάντα θετικά συσχετισμένες.

Τέλος να σημειώσουμε ότι οι διαδοχικές πιθανότητες της διδιάστατης κατανομής Yule (ρ) συνδεόνται με τη σχέση

$$\frac{p_{i+1,j}}{p_{i,j}} = \frac{p_{i,j+1}}{p_{i,j}} = \frac{1+i+j}{3+\rho+i+j}$$

Επιπλέον, η Xekalaki (1986) παρατηρεί ότι οι πιθανότητες $p_{x,y}$ της κατανομής αυτής είναι συμμετρικές ως προς x και y ($p_{x,y} = p_{y,x}$, $x,y=0,1,2,\ldots$) και σταθερές επί της ευθείας $x+y=c$, όπου c σταθερά $\left(p_{x,c-x} = p_{y,c-y} , \; x\neq y, \; x,y = 0,1,2,\ldots\right)$.

ΚΕΦΑΛΑΙΟ 3

ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΗΝ ΥΠΟ ΣΥΝΘΗΚΗ ΚΑΤΑΝΟΜΗ Η/ΚΑΙ ΣΤΗΝ ΥΠΟ ΣΥΝΘΗΚΗ ΑΝΑΜΕΝΟΜΕΝΗ ΤΙΜΗ

3.1 ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό περιέχονται χαρακτηρισμοί των μονοδιάστατων κατανομών Pareto τύπου I και II (συνεχής περίπτωση) και τηςμονοδιάστατης κατανομής Yule (διακριτή περίπτωση) όταν ο παρατηρητής έχει πρόσβαση σε πληροφορίες σχετικές με την υπό συνθήκη κατανομή της X δεδομένου του γεγονότος $\{Y=y\}$ ή/και με τη συνάρτηση παλινδρόμησης $E(X|Y=y)$.

Στην πρώτη ενότητα καθορίζονται οι συνθήκες οι οποίες προσδιορίζουν μονοσήμαντα την περιθώρια συνάρτηση (πυκνότητας) πιθανότητας της Y και κατά συνέπεια την από κοινού συνάρτηση (πυκνότητας) πιθανότητας των X και Y στην περίπτωση που δίνονται οι μορφές των συναρτήσεων (πυκνότητας) πιθανότητας των μεταβλητών X και $X|(Y=y)$.

Στη δεύτερη ενότητα αποδεικνύονται τέσσερα γενικά θεωρήματα από τα οποία προκύπτει ότι δεδομένης της μορφής της συνάρτησης (πυκνότητας) πιθανότητας της μεταβλητής $X|(Y=y)$ η συνάρτηση παλινδρόμησης $E(X|Y=y)$ προσδιορίζει μονοσήμαντα τη συνάρτηση (πυκνότητας) πιθανότητας της X. Αποδεικνύεται ότι συμπεράσματα ήδη γνωστά στη βιβλιογραφία αποτελούν ειδικές περιπτώσεις (πορίσματα) των τεσσάρων αυτών θεωρημάτων.

Στις δυο τελευταίες ενότητες αποδεικνύονται χαρακτηριστικές ιδιότητες για την κατανομή της X στην περίπτωση που υπάρχουν πληροφορίες για την $E\left[h(X)|X>y\right]$. Το θέμα αντιμετωπίζεται τόσο στη συνεχή περίπτωση (τρίτη ενότητα) όσο και στη διακριτή περίπτωση (τετάρτη ενότητα). Γίνεται αναδρομή στην υπάρχουσα βιβλιογραφία και γενικεύονται μερικά πολύ σημαντικά θεωρήματα.

3.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΗΝ ΥΠΟ ΣΥΝΘΗΚΗ ΚΑΤΑΝΟΜΗ

Αρχικά θα ασχοληθούμε με το πρόβλημα του καθορισμού της διδιάστατης κατανομής των X και Y στην περίπτωση που είναι γνωστή η περιθώρια κατανομή της μιας συνιστώσας, έστω της X και η υπό συνθήκη κατανομή της ίδιας συνιστώσας δοθείσης της άλλης. Στο θεώρημα που ακολουθεί αποδεικνύεται ότι όταν η περιθώρια της X είναι Pareto $(1,\alpha)$ τότε, για δεδομένη μορφή της $f_{X|(Y=y)}(x)$ η περιθώρια κατανομή της Y ορίζεται μονοσήμαντα και είναι και αυτή Pareto.

Θεώρημα 4.2.1 Εστω (X,Y) τυχαίο διάνυσμα. Υποθέτουμε ότι η περιθώρια κατανομή της X είναι Pareto $(1,\alpha)$ και η υπό συνθήκη κατανομή της X δεδομένου του γεγονότος (Y=y) είναι τέτοια ώστε

$$f_{X|(Y=y)}(x) = \alpha\, x^{-\alpha(1+\beta\theta \ln y)-1}\left[(1+\alpha\theta \ln x)(1+\beta\theta \ln y)-\theta\right] \qquad (3.2.1)$$

$$x\geq 1,\ y\geq 1,\quad \alpha,\beta\geq 0,\quad 0\leq\theta\leq 1\ .$$

Τότε η περιθώρια κατανομή της Y ορίζεται μονοσήμαντα και είναι Pareto $(1,\beta)$.

Απόδειξη Είναι γνωστό ότι

$$\int_{R_y} f_{X|(Y=y)}(x)\, f_Y(y)\, dy = f_X(x).$$

Στην προκειμένη περίπτωση η παραπάνω σχέση γίνεται

$$\int_1^{\infty} \alpha\, x^{-\alpha(1+\beta\theta \ln y)-1} \left[(1+\alpha\theta \ln x)(1+\beta\theta \ln y)-\theta\right] f_Y(y)\, dy = \alpha\, x^{-\alpha-1}$$

ή ισοδύναμα,

$$\int_1^{\infty} x^{-\alpha-\alpha\beta\theta \ln y-1+\alpha+1} \left[1+\beta\theta \ln y+\alpha\theta \ln x+\alpha\beta\theta^2 \ln x \ln y-\theta\right] f_Y(y)\, dy = 1.$$

Θέτοντας $-\alpha\beta\theta \ln x = \kappa$ προκύπτει ότι

$$\int_1^{\infty} x^{-\alpha\beta\theta \ln y} \left[1 + \beta\theta \ln y - \frac{\kappa}{\beta} - \theta\kappa \ln y - \theta\right] f_Y(y)\, dy = 1.$$

Δηλαδή,

$$\int_1^{\infty} x^{-\alpha\beta\theta \ln y} \left[(1 - \frac{\kappa}{\beta} - \theta) + (1 - \frac{\kappa}{\beta})\, \beta\theta \ln y\right] f_Y(y)\, dy = 1.$$

Η σχέση αυτή γράφεται ισοδύναμα ως εξής

$$\int_1^{\infty} y^{-\alpha\beta\theta \ln x} \left[(1 - \frac{\kappa}{\beta} - \theta) + (1 - \frac{\kappa}{\beta})\, \beta\theta \ln y\right] f_Y(y)\, dy = 1$$

ή

$$\int_1^{\infty} y^{\kappa}\, (1 - \frac{\kappa}{\beta} - \theta)\, dy + \int_1^{\infty} y^{\kappa}\, (1 - \frac{\kappa}{\beta})\, \beta\theta \ln y\, f_Y(y) dy - 1 = 0.$$

Επομένως,

$$(1 - \frac{\kappa}{\beta} - \theta)\int_1^{\infty} y^{\kappa}\, f_Y(y)\, dy+(1 - \frac{\kappa}{\beta})\beta\theta\int_1^{\infty} y^{\kappa} \ln y\, f_Y(y)dy = 1 \qquad (3.2.2)$$

θέτω $\int_1^{\infty} y^{\kappa} f_Y(y)\, dy = h(\kappa)$

οπότε $\int_1^{\infty} y^{\kappa} \ln y\, f_Y(y)\, dy = h'(\kappa)$.

Κατά συνέπεια η σχέση (4.2.2) γίνεται

$$(1 - \frac{\kappa}{\beta})\, \beta\theta\, h'(\kappa) + (1 - \frac{\kappa}{\beta} - \theta)\, h(\kappa) - 1 = 0 \qquad \text{με} \quad h(0) = 1.$$

Λύση της διαφορικής αυτής εξίσωσης είναι,

$$\overline{h(\kappa)} = \frac{\beta}{\beta - \kappa} .$$

Δηλαδή,

$$\int_1^{\infty} y^{\kappa} f_Y(\kappa)\, dy = \frac{\beta}{\beta - \kappa} .$$

Είναι όμως γνωστό ότι ο μετασχηματισμός Mellin προσδιορίζει μονοσήμαντα την κατανομή, κατά συνέπεια,

$$f_Y(y) = \beta\, y^{-(\beta+1)}, \quad y \geq 1. \qquad (3.2.3)$$

Δηλαδή $Y \sim \text{Pareto}\ (1,\beta)$.

Παρατήρηση. Είναι γνωστό ότι γενικά ισχύει

$$f_{X,Y}(x,y) = f_{X|(Y=y)}(x,y)\, f_Y(y) .$$

Αν αντικαταστήσουμε στη σχέση αυτή τις (3.2.1) και (3.2.3) καταλήγουμε ότι

$$f_{X,Y}(x,y) = \alpha\beta\, y^{-\beta-1}\, x^{-\alpha(1+\beta\theta \ln y)-1}\left[(1+\alpha\theta \ln x)(1+\beta\theta \ln y)-\theta\right] .$$

Αλλά αυτή δεν είναι άλλη από τη συνάρτηση πυκνότητας πιθανότητας της διδιάστατης κατανομής Pareto τύπου Gumbel. Εμμεσα λοιπόν το θεώρημα οδηγεί σε χαρακτηρισμό της διδιάστατης αυτής κατανομής.

<u>Πόρισμα 3.2.1</u> Εστω (X,Y) τυχαίο διάνυσμα. Υποθέτουμε ότι η υπό συνθήκη κατανομή της X δεδομένου του γεγονότος (Y=y) είναι τέτοια ώστε

$$f_{X|(Y=y)}(x) = \alpha\, x^{-\alpha(1+\beta\theta \ln y)-1}\left[(1+\alpha\theta \ln x)(1+\beta\theta \ln y)-\theta\right] \qquad (3.2.4)$$

$$x\geq 1,\ y\geq 1,\quad \alpha,\beta\geq 0,\quad 0\leq\theta\leq 1 .$$

Τότε οι προτάσεις:

(α) X ~ Pareto $(1,\alpha)$

(β) Y ~ Pareto $(1,\beta)$

είναι ισοδύναμες.

<u>Απόδειξη</u> Εστω ότι ισχύει η (3.2.4) και έστω επίσης ότι X ~ Pareto $(1,\alpha)$ τότε με βάση το θεώρημα 3.2.1 συνεπάγεται ότι Y ~ Pareto $(1,\beta)$.

Εστω τώρα ότι ισχύει η (3.2.4) και έστω επίσης ότι Y ~ Pareto $(1,\beta)$ τότε προφανώς X ~ Pareto $(1,\alpha)$.

Τα δυο θεωρήματα που ακολουθούν καθορίζουν το πλαίσιο στο οποίο η κατανομή της Y προσδιορίζεται μονοσήμαντα από την κατανομή της X στην περίπτωση που η κατανομή της X δεδομένου του γεγονότος Y=y είναι η Pareto II, μετατοπισμένη δηλαδή μορφή της κατανομής Pareto.

Θεώρημα 3.2.2 Αν η τυχαία μεταβλητή $X|(Y=y)$ ακολουθεί την Pareto II (θ=y, c=y-1, α=p+1) τότε η κατανομή της X προσδιορίζει μονοσήμαντα την κατανομή της Y, στην περίπτωση που η τελευταία είναι ανεξάρτητη της παραμέτρου p.

Απόδειξη Εστω ότι $X|(Y=y) \sim$ Pareto II (θ=y, c=y-1, α=p+1)

Δηλαδή,

$$f_{X|(Y=y)}(x) = (p+1)\, y^{p+1}\, (x+y-1)^{-(p+2)}.$$

Τότε,

$$f_X(x) = (p+1) \int_1^{\infty} y^{p+1}\, (x+y-1)^{-(p+2)}\, f_Y(y)\, dy\ .$$

Εστω $f_Y(y)$ μια λύση ανεξάρτητη της παραμέτρου p. Αν η $f_Y(y)$ δεν είναι η μόνη λύση τότε για κάθε άλλη λύση $f_Y'(y)$ ανεξάρτητη της p ισχύει .

$$\int_1^{\infty} y^{p+1}\, (x+y-1)^{-(p+2)}\, f_Y^*(y)\, dy = 0 \qquad (3.2.5)$$

όπου $f^* = f - f'$.

Είναι όμως γνωστό ότι η οικογένεια Pearson VI είναι πλήρης, κατά συνέπεια η σχέση (3.2.5) ισοδυναμεί με την $f^* = 0$ ή $f = f'$.

Αρα, η κατανομή της Y ορίζεται μονοσήμαντα.

Το θεώρημα που ακολουθεί καθορίζει τη συνθήκη ώστε η κατανομή της Y να ορίζεται μονοσήμαντα ως Pareto (1,p-1) στην περίπτωση που η δεσμευμένη κατανομή της X δεδομένου του γεγονότος Y=y είναι επίσης Pareto μετατοπισμένη όμως y μονάδες προς τα δεξιά.

Θεώρημα 3.2.3 Εστω ότι η τυχαία μεταβλητή X|(Y=y) ακολουθεί την κατανομή Pareto (y+1,p), $x \geq y > 0$. Τότε η τυχαία μεταβλητή X κατανέμεται ως beta II (2,p-1) αν και μόνον αν η Y ακολουθεί την Pareto (1,p-1).

Απόδειξη

Ευθύ: Εστω ότι $X|(Y=y) \sim$ Pareto (y+1,p), $x \geq y > 0$. Εστω επίσης ότι $Y \sim$ Pareto (p-1), $y > 0$.

Είναι γνωστό ότι

$$f_X(x) = \int_{R_Y} f_{X|(Y=y)}(x)\, f_Y(y)\, dy.$$

Η σχέση αυτή στη συγκεκριμένη περίπτωση γίνεται

$$f_X(x) = \int_0^x p(1+y)^p\, (1+x)^{-(p+1)}(p-1)\, (1+y)^{-p}\, dy$$

$$= p(p-1)(1+x)^{-(p+1)}x\ .$$

Δηλαδή, η τυχαία μεταβλητή X ακολουθεί την κατανομή beta II (2,p-1).

Αντίστροφο: Εστω ότι $X|(Y=y) \sim$ Pareto $(y+1,p)$, $x \geq y > 0$. Εστω επίσης ότι $X \sim$ beta II $(2,p-1)$. Τότε,

$$p(p-1)(1+x)^{-(p+1)}x = \int_0^x p(1+y)^p(1+x)^{-(p+1)} f_Y(y)\, dy.$$

Προφανώς, μια λύση της παραπάνω εξίσωσης είναι η συνάρτηση πυκνότητας πιθανότητας της Pareto $(1,p-1)$. Η λύση αυτή είναι και η μοναδική, διότι διαφορετικά για κάθε άλλη λύση $f_Y'(y)$ θα ισχύει

$$\int_0^x p(1+y)^p(1+x)^{-(p+1)} f_Y^*(y)\, dy = 0 \qquad (3.2.6)$$

όπου $f^* = f - f'$.

Από τη σχέση (3.2.6) προκύπτει με παραγώγιση των μελών της ότι $f^* = 0$, δηλαδή ότι $f = f'$. Αρα η κατανομή της τυχαίας μεταβλητής Y ορίζεται μονοσήμαντα ως Pareto $(1,p-1)$.

Αξίζει να σημειωθεί ότι ανάλογα αποτελέσματα για την περίπτωση διακριτών κατανομών έχουν αποδειχθεί από την Xekalaki (1981). Συγκεκριμένα έχει αποδειχθεί ότι αν η $X|(Y=y)$ ακολουθεί τη μονοδιάστατη γενικευμένη κατανομή Waring $(y+1,n;\rho+m)$ και η X τη μονοδιάστατη γενικευμένη κατανομή Waring $(1,m+n;\rho)$ τότε η κατανομή της Y προσδιορίζεται μονοσήμαντα ως μονοδιάστατη γενικευμένη κατανομή Waring $(1,m;\rho)$. Αν μάλιστα θεωρήσουμε την ειδική περίπτωση $m=1=n$ τότε η κατανομή της Y ορίζεται μονοσήμαντα ως Yule (ρ).

Διδιάστατα πιθανοθεωρητικά μοντέλα χρησιμοποιούνται συχνά για να περιγράψουν τυχαία φαινόμενα, κυρίως γιατί δίνουν μια βαθύτερη θεώρηση του ενυπάρχοντος τυχαίου μηχανισμού. Σε πολλές περιπτώσεις όπου ένα διδιάστατο μοντέλο χρησιμοποιείται, η δεσμευμένη κατανομή της μιας τυχαίας μεταβλητής δοθέντος του αθροίσματός τους μπορεί να δώσει πολλές πληροφορίες. Ετσι θα αναφερθούμε για λίγο στο πρόβλημα που παρουσιάζεται κατά τον προσδιορισμό της περιθώριας κατανομής της Y όταν στο πλαίσιο μιας διδιάστατης κατανομής των X και Y οι πληροφορίες που έχουμε αφορούν την κατανομή της X|(X+Y=Z). Το θέμα αντιμετωπίστηκε για πρώτη φορά απο τους Patil and Seshadri (1964) στην περίπτωση που οι μεταβλητές X και Y είναι ανεξάρτητες. Απέδειξαν, μεταξύ άλλων ότι κάτω από ορισμένες προυποθέσεις οι τυχαίες μεταβλητές X και Y ακολουθούν την αρνητική εκθετική κατανομή. Βελτίωση των αποτελεσμάτων επιχειρείται απο τους Menon (1966) και Galambos (1975). Γενίκευση του αποτελέσματος επιτυγχάνεται απο τους Mathai (1966) και Berk (1977) για την περίπτωση n ανεξάρτητων τυχαίων μεταβλητών όταν οι πληροφορίες που έχουμε αναφέρονται στην κατανομή των $X_i|(X_1+...+X_n=Z)$, i=1,2,...,n. Ο Panaretos (1982c) άρει την υπόθεση της ανεξαρτησίας και αντιμετωπίζει το πρόβλημα στη περίπτωση που οι τυχαίες μεταβλητές X και Y δεν είναι ανεξάρτητες, οι κατανομές που ακολουθούν είναι διακριτές και η μορφή της κατανομής της X|(X+Y=Z) είναι γενική. Τελος, αξίζει να αναφέρουμε ότι η Dimaki (1988b) προσδιορίζει τις περιθώριες κατανομές των τυχαίων μεταβλητών X και Y οι οποίες θεωρούνται ανεξάρτητες όταν δίνεται η κατανομή της X|(T=t), όπου T = XY.

3.3 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΗΝ ΥΠΟ ΣΥΝΘΗΚΗ ΚΑΤΑΝΟΜΗ ΚΑΙ ΣΤΗΝ ΥΠΟ ΣΥΝΘΗΚΗ ΑΝΑΜΕΝΟΜΕΝΗ ΤΙΜΗ.

Η υπό συνθήκη κατανομή παίζει ένα σημαντικό ρόλο στη στατιστική συμπερασματολογία κυρίως γιατί μπορεί να μετασχηματίζει ένα πρόβλημα ορισμένο σε ένα δειγματικό χώρο σε ένα άλλο πρόβλημα ορισμένο σε ένα περιορισμό του δειγματικού χώρου. Είναι όμως επίσης γνωστό ότι απλή γνώση της υπό συνθήκη κατανομής δυο τυχαίων μεταβλητών δεν είναι επαρκής για τον μονοσήμαντο καθορισμό των περιθωρίων κατανομών. Συχνά μαζί με τη μορφή της δεσμευμένης κατανομής γίνεται μια πρόσθετη υπόθεση σχετική με τη μορφή της συνάρτησης παλινδρόμησης.

Τα θεωρήματα που ακολουθούν αποδεικνύουν ότι κάτω από γενικές υποθέσεις, η συνάρτηση παλινδρόμησης $E(X|Y=y)$ χαρακτηρίζει την κατανομή μιας τυχαίας μεταβλητής X όταν είναι δοσμένη η μορφή της δεσμευμένης κατανομής της X επί μιας άλλης τυχαίας μεταβλητής Y. Θα ξεκινήσουμε την αντιμετώπιση του προβλήματος με την περίπτωση που οι τυχαίες μεταβλητές είναι διακριτές.

<u>Θεώρημα 3.3.1</u> Εστω X τυχαία μεταβλητή που παίρνει ακέραιες και θετικές τιμές και της οποίας η κατανομή είναι εκφυλισμένη. Εστω επίσης Y τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές τέτοια ώστε $Y \le X$. Υποθέτουμε ότι

$$P(X=x|Y=y) = \frac{\frac{1}{x} P(X=x)}{\sum_{y=x+1}^{\infty} \frac{1}{x} P(X=x)}, \quad x = 1,2,... \qquad (3.3.1)$$

Τότε η συνάρτηση

$$\lambda(y) = E(X|Y=y), \quad y = 0,1,2,...$$

ορίζει μονοσήμαντα την κατανομή της τυχαίας μεταβλητής X, οποτεδήποτε η $\lambda(y)$ είναι μια μη σταθερή συνάρτηση.

Απόδειξη Εστω ότι p_x, q_y είναι οι συναρτήσεις πιθανότητας των X και Y αντίστοιχα. Τότε με βάση τον ορισμό της $E(X|Y=y)$ προκυπτει ότι,

$$\lambda(y) = E(X|Y=y)$$

$$= \sum_{x=y+1}^{\infty} x\, P(X=x|Y=y)$$

$$= \sum_{x=y+1}^{\infty} \frac{P(X=x)}{\sum_{x=y+1}^{\infty} (1/x)\, P(X=x)} .$$

Δηλαδή,

$$\lambda(y) \sum_{x=y+1}^{\infty} \frac{p_x}{x} = \sum_{x=y+1}^{\infty} p_x \qquad (3.3.2)$$

Εξειδικεύοντας τη σχέση (3.3.2) για y=r και y=r+1 και αφαιρώντας κατά μέλη τις εξισώσεις που προκύπτουν έχουμε ότι:

$$\left\{\lambda(r) - \lambda(r+1)\right\} \sum_{x=r+2}^{\infty} \frac{p_x}{x} = p_{r+1} \left(1 - \frac{\lambda(r)}{r+1}\right)$$

Εφαρμόζοντας την ίδια τεχνική φθάνουμε στη παρακάτω εξίσωση,

$$\frac{p_{r+1}}{r+1} = \frac{p_r}{r} \frac{r-\lambda(r-1)}{\lambda(r-1)-\lambda(r)} - \frac{p_{r+1}}{r+1} \frac{r+1-\lambda(r)}{\lambda(r)-\lambda(r+1)}$$

ή

$$\frac{p_{r+1}}{r+1} \left\{1 + \frac{r+1-\lambda(r)}{\lambda(r)-\lambda(r+1)}\right\} = \frac{p_r}{r} \frac{r-\lambda(r-1)}{\lambda(r-1)-\lambda(r)} . \qquad (3.3.3)$$

Θέτω

$$\frac{p_r}{r} \frac{1}{\lambda(r-1)-\lambda(r)} = P_r \qquad (3.3.4)$$

κατά συνέπεια η σχέση (3.3.3) γράφεται

$$P_{r+1} \left\{r+1-\lambda(r+1)\right\} - P_r \left\{r-\lambda(r-1)\right\} = 0.$$

Προφανώς $r+1-\lambda(r+1) \neq 0$, αλλιώς $P_r=0$, $r=1,2,\ldots$ το οποίο είναι άτοπο. Κατά συνέπεια,

$$P_{r+1} - P_r \frac{r-\lambda(r-1)}{r+1-\lambda(r+1)} = 0 .$$

Η μοναδική λύση της εξίσωσης αυτής διαφορών δίνεται από τη σχέση

$$P_r = P_1 \prod_{i=1}^{r-1} \frac{i-\lambda(i-1)}{i+1-\lambda(i-1)}, \quad r = 2,3,... \qquad (3.3.5)$$

Αντικαθιστώντας στη σχέση (3.3.5) την (3.3.4) προκύπτει,

$$p_r = p_1 \frac{\lambda(r-1)-\lambda(r)}{\lambda(0)-\lambda(1)} \prod_{i=1}^{r-1} \frac{i-\lambda(i-1)}{i+1-\lambda(i-1)}, \quad r = 2,3,... \qquad (3.3.6)$$

όπου $\lambda(i) \neq i$, δηλαδή μη σταθερά.

Κατά συνέπεια η p_x ορίζεται μονοσήμαντα από τη συνάρτηση $\lambda(y) = E(X|Y=y)$.

Στη συνέχεια θα θεωρήσουμε την ειδική περίπτωση που η συνάρτηση παλινδρόμησης είναι γραμμική. Θα υποθέσουμε δηλαδή ότι $E(X|Y=y) = \alpha + \beta y$. Με το πόρισμα 3.3.1 αποδεικνύεται ότι στην ειδική περίπτωση που $\alpha=\beta=c$ η κατανομή της X είναι η Yule (ρ) με $\rho = 1/(c-1)$.

<u>Πόρισμα 3.3.1</u> Εστω οι τυχαίες μεταβλητές X και Y που ορίζονται όπως στο θεώρημα 3.3.1. Εστω επίσης ότι η $P(X=x|Y=y)$ δίνεται από τη σχέση (3.3.1). Τότε

$E(X|Y=y) = c + cy, \quad y = 0,1,2,..., \quad c > 1$

τότε και μόνο τότε αν η κατανομή της X είναι η Yule $\left(\rho = 1/(c-1)\right)$.

<u>Απόδειξη</u> Σύμφωνα με το θεώρημα 4.3.1 η συνάρτηση $\lambda(y)=E(X|Y=y)$ είναι γραμμική, δηλαδή

$\lambda(y) = \alpha + \beta y.$

Τότε η σχέση (3.3.6) γίνεται

$$P(X=y) = P(X=1)\ y \prod_{r=1}^{y-1} \frac{[(\beta-\alpha)+(1-\beta)r]}{[(1-\alpha-\beta)+(1-\beta)r]}$$

$$= P(X=1)\ y \prod_{r=1}^{y-1} \frac{\frac{\beta-\alpha}{1-\beta} + r}{\frac{1-\alpha-\beta}{1-\beta} + r} .$$

Θέτουμε όπου $(\beta-\alpha)/(1-\beta) = c_1$ και $(1-\alpha-\beta)/(1-\beta) = c_2$.

Τότε,

$$P(X=y) = P(X=1)\ y \prod_{r=1}^{y-1} \frac{c_1+r}{c_2+r}$$

$$= P(X=1)\ y \frac{(c_1+1)_{(y-1)}}{(c_2+1)_{(y-1)}}$$

$$= P(X=1) \frac{(c_1+1)_{(y-1)}\ 1_{(y)}}{(c_2+1)_{(y-1)}} \frac{y}{1_{(y)}}$$

$$= P(X=1) \frac{(c_1+1)_{(y-1)}\ 2_{(y-1)}}{(c_2+1)_{(y-1)}} \frac{1}{(y-1)!} .$$

Αλλά,

$$\sum_{y=1}^{\infty} P(X=y) = 1.$$

Κατά συνέπεια,

$$1 = P(X=1) \sum_{y=1}^{\infty} \frac{(c_1+1)_{(y-1)}\ 2_{(y-1)}}{(c_2+1)_{(y-1)}} \frac{1}{(y-1)!}$$

$$= P(X=1) \sum_{y=1}^{\infty} \frac{(c_1+1)_{(y)}\ 2_{(y)}}{(c_2+1)_{(y)}} \frac{1}{y!}$$

$$= P(X=1) \frac{\Gamma(c_2+1)\ \Gamma(c_2-c_1-2)}{\Gamma(c_2-c_1)\ \Gamma(c_2-1)} ,$$

δηλαδή,

$$1 = P(X=1) \frac{c_2\ (c_2-1)}{(c_2-c_1-1)\ (c_2-c_1-2)} .$$

Επομένως,

$$P(X=1) = \frac{(c_2-c_1-1)\ (c_2-c_1-2)}{c_2\ (c_2-1)} .$$

Αλλά, $(\beta-\alpha)/(1-\beta) = c_1$ και $(1-\alpha-\beta)/(1-\beta) = c_2$

Επομένως,

$$P(X=1) = \frac{\beta}{\alpha} \frac{1}{\beta+\alpha-1} .$$

Στην περίπτωση που $\alpha = \beta = c$ προκύπτει ότι,

$$P(X=1) = \frac{1}{2c-1} .$$

Η τυχαία μεταβλητή X ακολουθεί λοιπόν την κατανομή Yule με $\rho = 1/(c-1)$.

Με το θεώρημα 3.3.1 χαρακτηρίσαμε την κατανομή της X από τη συνάρτηση παλινδρόμησης στην περίπτωση που $Y \leq X$. Το θεώρημα που ακολουθεί αντιμετωπίζει το ίδιο θέμα στη περίπτωση που η μεταξύ των μεταβλητών σχέση είναι η $Y \geq X$.

Θεώρημα 3.3.2 Εστω X μια θετική τυχαία μεταβλητή με μια μη εκφυλισμένη κατανομή ορισμένη στο σύνολο των θετικών αριθμών. Εστω Y μια άλλη τυχαία μεταβλητή που παίρνει ακέραιες και θετικές τιμές και τέτοια ώστε $Y \geq X$. Υποθέτουμε ότι,

$$P(X=x|Y=y) = \frac{x\,P(X=x)}{\sum_{x=1}^{y} x\,P(X=x)} , \quad x = 1,2,... \qquad (3.3.4)$$

Τότε η συνάρτηση

$$\lambda(y) = E(X|Y=y) , \quad y = 1,2,...$$

ορίζει μονοσήμαντα την κατανομή της τυχαίας μεταβλητής X, οποτεδήποτε η $\lambda(y)$ είναι μία μη σταθερή συνάρτηση.

Απόδειξη Εστω ότι p_x, q_y είναι οι συναρτήσεις πιθανότητας των X και Y αντίστοιχα. Τότε με βάση τον ορισμό της $E(X|Y=y)$ προκύπτει ότι,

$$\lambda(y) = E(X|Y=y)$$

$$= \sum_{x=1}^{y} x\, P(X=x|Y=y)$$

$$= \sum_{x=1}^{y} \frac{x^2\, P(X=x)}{\sum_{x=1}^{y} x\, P(X=x)} ,$$

Δηλαδή,

$$\lambda(y) \sum_{x=1}^{y} x\, p_x = \sum_{x=1}^{y} x^2\, p_x \,. \qquad (3.3.4)$$

Εξειδικεύοντας τη σχέση (3.3.4) για y=r και y=r+1 και αφαιρώντας κατά μέλη τις εξισώσεις που προκύπτουν έχουμε ότι

$$\sum_{x=1}^{r} x\, p_x = \frac{[(r+1)-\lambda(r+1)]\ (r+1)}{[\lambda(r+1)-\lambda(r)]}\, p_{r+1} .$$

Εφαρμόζοντας την ίδια τεχνική φθάνουμε στην παρακάτω εξίσωση,

$$\frac{[(r+2)-\lambda(r+2)]\ (r+2)}{[\lambda(r+2)-\lambda(r+1)]}\ p_{r+2} - \frac{[(r+1)-\lambda(r)]\ (r+1)}{[\lambda(r+1)-\lambda(r)]}\ p_{r+1} = 0$$

η οποία τελικά γράφεται,

$$p_{r+1} - \frac{r\ [r-\lambda(r-1)][\lambda(r+1-\lambda(r)]}{(r+1)\ [\lambda(r)-\lambda(r-1)][(r+1)-\lambda(r+1)]}\ p_r = 0.$$

Η μοναδική λύση της εξίσωσης αυτής διαφορών δίνεται από τη σχέση

$$P(X=y)=P(X=1)\prod_{r=1}^{y-1}\frac{r}{r+1}\ \frac{[r-\lambda(r-1)][\lambda(r+1)-\lambda(r)]}{[\lambda(r)-\lambda(r-1)][(r+1)-\lambda(r+1)]},\ r=2,3,\ldots \quad (3.3.5)$$

Κατά συνέπεια η p_x ορίζεται μονοσήμαντα από τη συνάρτηση $\lambda(y) = E(X|Y=y),\ y = 1,2,\ldots$

Τα δύο τελευταία θεωρήματα της ενότητας αυτής ασχολούνται με την περίπτωση κατά την οποία οι τυχαίες μεταβλητές είναι συνεχείς.

<u>Θεώρημα 3.3.3</u> Εστω X θετική συνεχής τυχαία μεταβλητή της οποίας η κατανομή είναι μη εκφυλισμένη. Εστω επίσης Y μη αρνητική συνεχής τυχαία μεταβλητή τέτοια ώστε $Y \le X$. Υποθέτουμε ότι

$$f_{X|(Y=y)}(x) = \frac{\frac{1}{x}\ f_X(x)}{\int_y^\infty \frac{1}{x}\ f_X(x)\ dx},\ x>0 \quad (3.3.5)$$

Τότε η συνάρτηση

$\lambda(y) = E(X|Y=y)$, $y>0$

ορίζει μονοσήμαντα την κατανομή της X.

<u>Απόδειξη</u> Εστω $f_X(x)$ η συνάρτηση πυκνότητας πιθανότητας της X. Τότε με βάση τον ορισμό της $E(X|Y=y)$ προκύπτει ότι,

$$\lambda(y) = E(X|Y=y)$$

$$= \int_y^\infty x\, f_{X|(Y=y)}(x)\, dx$$

$$= \int_y^\infty x \frac{\frac{1}{x} f_X(x)}{\int_y^\infty \frac{1}{x} f_X(x)\, dx}\, dx \ .$$

Δηλαδή,

$$\lambda(y) \int_y^\infty \frac{1}{x} f_X(x)\, dx = \int_y^\infty f_X(x)\, dx \ . \qquad (3.3.6)$$

Παραγωγίζοντας τη σχέση (3.3.6) ως προς y προκύπτει ότι,

$$\lambda'(y) \int_y^\infty \frac{1}{x} f_X(x)\, dx - \lambda(y) \frac{1}{y} f_X(y) = -f_X(y) \ . \qquad (3.3.7)$$

Θέτουμε

$$K(y) = \int_y^\infty x^{-1} f_X(x)\, dx \ .$$

Τότε

$$K'(y) = - y^{-1}\, f_X(y).$$

Αντικαθιστώντας τα $K(y)$ και $K'(y)$ στην (3.3.7) προκύπτει ότι,

$$\lambda'(y)\, K(y) + \lambda(y)\, K'(y) = y\, K'(y).$$

Προφανώς $\lambda(y) \neq y$, διότι διαφορετικά $f_X(y) = 0$, $y \geq 0$ το οποίο αντίκειται στις υποθέσεις για την τυχαία μεταβλητή X. Επομένως,

$$\frac{K'(y)}{K(y)} = \frac{\lambda'(y)}{y-\lambda(y)} \ .$$

Η λύση της διαφορικής αυτής εξίσωσης είναι,

$$K(y) = C \exp \int \frac{\lambda'(y)}{y-\lambda(y)}\, dy$$

η οποία ισοδύναμα γράφεται,

$$\int_y^\infty x^{-1}\, f_X(x)\, dx = C \exp \int \frac{\lambda'(y)}{y-\lambda(y)}\, dy \ .$$

Παραγωγίζοντας ως προς y οδηγούμαστε στη σχέση

$$y^{-1} f_X(y) = - C \exp \int \frac{\lambda'(y)}{y-\lambda(y)} dy \frac{\partial}{\partial y} \left(\int \frac{\lambda'(y)}{y-\lambda(y)} dy \right),$$

δηλαδή στη σχέση

$$f_X(y) = C\, y \frac{\lambda'(y)}{\lambda(y)-y} \exp \int \frac{\lambda'(y)}{y-\lambda(y)} dy \quad . \qquad (3.3.8)$$

Από τη σχέση (3.3.8) είναι προφανές ότι η $f_X(y)$ ορίζεται μονοσήμαντα από τη συνάρτηση παλινδρόμησης $\lambda(y) = E(X|Y=y)$.

Θα εξετάσουμε τώρα την ειδική περίπτωση κατά την οποία η συνάρτηση παλινδρόμησης είναι γραμμική. Θα υποθέσουμε δηλαδή ότι $E(X|Y=y) = \alpha + \beta\, y$. Με το πόρισμα που ακολουθεί αποδεικνύεται ότι στην ειδική περίπτωση που $\alpha = 0$ η τυχαία μεταβλητή X κατανέμεται ως Pareto $\left(1, 1/(\beta-1)\right)$.

<u>Πόρισμα 3.3.2</u> Εστω οι τυχαίες μεταβλητές X και Y που ορίζονται όπως στο θεώρημα 3.3.3. Εστω επίσης ότι η $f_{X|(Y=y)}$ δίνεται από τη σχέση (3.3.5). Τότε

$E(X|Y=y) = \beta y$, $y > 0$, $\beta > 1$

τότε και μόνο τότε αν η κατανομή της X είναι Pareto $\left(1, 1/(\beta-1)\right)$.

<u>Απόδειξη</u> Οι μεταβλητές X και Y ορίζονται όπως στο θεώρημα 3.3.3 με την συνάρτηση $\lambda(y) = E(X|Y=y)$ να είναι γραμμικής μορφής, δηλαδή

$\lambda(y) = \alpha + \beta y$.

Τότε η σχέση (3.3.8) γράφεται,

$$f_X(y) = C\ y\ \frac{\beta}{\alpha+(\beta-1)y}\ \exp \int \frac{\beta}{(1-\beta)y-\alpha}\ dy\ . \qquad (3.3.9)$$

Θέτουμε όπου $\beta/(1-\beta) = \mu$ και $\alpha/(1-\beta) = \kappa$ και η σχέση (3.3.9) γίνεται,

$$f_X(y) = C\ y\ \frac{\mu}{y-\kappa}\ (y-\kappa)^{\mu}$$

ή

$$f_X(x) = C\ y\ \mu\ (y-\kappa)^{\mu-1}\ . \qquad (3.3.10)$$

Στην ειδική περίπτωση που $\alpha=0$ και κατά συνέπεια και $\kappa=0$ η σχέση (3.3.10) γίνεται,

$$f_X(y) = C\ \mu\ y^{\mu}\ . \qquad (3.3.11)$$

Αλλά,

$$\int_1^{\infty} f_X(x)\ dy = 1.$$

Κατά συνέπεια $C = (1+\mu)/\mu$. Η σχέση (3.3.11) επομένως γράφεται,

$f_X(y) = (\mu+1)\ y^{\mu}$, όπου $\mu = \beta/(1-\beta) > 0$.

Αποδείχθηκε λοιπόν, ότι στην περίπτωση που ισχύουν οι υποθέσεις του πορίσματος 3.3.2 η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $\left(1, 1/(1-\beta)\right)$.

Το τελευταίο θεώρημα της ενότητας αναφέρεται και αυτό σε συνεχείς τυχαίες μεταβλητές και χαρακτηρίζει την κατανομή της τυχαίας μεταβλητής X από τη συνάρτηση παλινδρόμησης $E(X|Y=y)$ στην περίπτωση που η μεταξύ των μεταβλητών σχέση είναι η $Y \geq X$.

Θεώρημα 3.3.5 Εστω X θετική συνεχής τυχαία μεταβλητή της οποίας η κατανομή είναι μη εκφυλισμένη. Εστω επίσης Y μη αρνητική συνεχής τυχαία μεταβλητή τέτοια ώστε $Y \geq X$. Υποθέτουμε ότι

$$f_{X|(Y=y)}(x) = \frac{x\, f_X(x)}{\int_1^y x\, f_X(x)\, dx}, \quad x > 0$$

Τότε η συνάρτηση

$$\lambda(y) = E(X|Y=y), \quad y \geq 0$$

ορίζει μονοσήμαντα την κατανομή της X.

Απόδειξη Εστω $f_X(x)$ η συνάρτηση πυκνότητας πιθανότητας της X. Τότε με βάση τον ορισμό της $E(X|Y=y)$ προκύπτει ότι,

$$\lambda(y) = E(X|Y=y)$$

$$= \int_1^y x \, f_{X|(Y=y)}(x) \, dy$$

$$= \int_1^y x \, \frac{x \, f_X(x)}{\int_1^y x \, f_X(x) \, dx} \, dx \, .$$

Δηλαδή,

$$\lambda(y) \int_1^y x \, f_X(x) \, dx = \int_1^y x^2 \, f_X(x) \, dx \, . \qquad (3.3.13)$$

Παραγωγίζοντας τη σχέση (3.3.13) ως προς y προκύπτει ότι,

$$\lambda'(y) \int_1^y x \, f_X(x) \, dx + \lambda(y) \, y \, f_X(y) = y^2 \, f_X(y). \qquad (3.3.14)$$

Θέτουμε

$$K(y) = \int_1^y x \, f_X(x) \, dx \, .$$

Επομένως,

$$K'(y) = y \, f_X(x).$$

Αντικαθιστώντας τα K(y) και K'(y) στην (3.3.14) προκύπτει ότι,

$$\lambda'(y) \, K(y) + \lambda(y) \, K'(y) = y \, K'(y).$$

Επειδή η κατανομή της τυχαίας μεταβλητής X είναι μη εκφυλισμένη,

$$= \int_1^y x \, f_{X|(Y=y)}(x) \, dy$$

$$= \int_1^y x \, \frac{x \, f_X(x)}{\int_1^y x \, f_X(x) \, dx} \, dx \, .$$

Δηλαδή,

$$\lambda(y) \int_1^y x \, f_X(x) \, dx = \int_1^y x^2 \, f_X(x) \, dx \, . \qquad (3.3.13)$$

Παραγωγίζοντας τη σχέση (3.3.13) ως προς y προκύπτει ότι,

$$\lambda'(y) \int_1^y x \, f_X(x) \, dx + \lambda(y) \, y \, f_X(y) = y^2 \, f_X(y). \qquad (3.3.14)$$

Θέτουμε

$$K(y) = \int_1^y x \, f_X(x) \, dx \, .$$

Επομένως,

$$K'(y) = y \, f_X(x).$$

Αντικαθιστώντας τα K(y) και K'(y) στην (3.3.14) προκύπτει ότι,

$$\lambda'(y) \, K(y) + \lambda(y) \, K'(y) = y \, K'(y).$$

Επειδή η κατανομή της τυχαίας μεταβλητής X είναι μη εκφυλισμένη,

ισχύει ότι $K(y) > 0$ και συνεπώς $\lambda(y) \neq y$. Αρα η τελευταία σχέση γράφεται

$$\frac{K'(y)}{K(y)} = \frac{\lambda'(y)}{y-\lambda(y)} .$$

Η λύση της διαφορικής αυτής εξίσωσης είναι,

$$K(y) = C \exp \int \frac{\lambda'(y)}{y-\lambda(y)} \, dy ,$$

η οποία ισοδύναμα γράφεται,

$$\int_1^y x \, f_X(x) \, dx = C \exp \int \frac{\lambda'(y)}{y-\lambda(y)} \, dy .$$

Παραγωγίζοντας ως προς y οδηγούμαστε στη σχέση

$$y \, f_X(y) = C \exp \int \frac{\lambda'(y)}{y-\lambda(y)} \, dy \, \frac{\partial}{\partial y} \left[\int \frac{\lambda'(y)}{y-\lambda(y)} \, dy \right] ,$$

ή ισοδύναμα ότη σχέση

$$f_X(y) = C \, y^{-1} \, \frac{\lambda'(y)}{y-\lambda(y)} \exp \int \frac{\lambda'(y)}{y-\lambda(y)} \, dy \quad . \qquad (3.3.15)$$

Από την (3.3.15) είναι προφανές ότι η $f_X(x)$ ορίζεται μονοσήμαντα από τη συνάρτηση παλινδρόμησης $\lambda(y) = E(X|Y=y)$

3.4 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΗΝ ΥΠΟ ΣΥΝΘΗΚΗ ΑΝΑΜΕΝΟΜΕΝΗ ΤΙΜΗ. ΣΥΝΕΧΗΣ ΠΕΡΙΠΤΩΣΗ.

Οι δεσμευμένες αναμενόμενες τιμές αποτελούν ένα δημοφιλές εργαλείο για την απόδειξη διαφόρων χαρακτηριστικών ιδιοτήτων ορισμένων κατανομών. Πρώτος ο Shanbhag (1970) έδωσε έναν ενδιαφέροντα χαρακτηρισμό της εκθετικής κατανομής σε σχέση με ιδιότητα της συνάρτησης $E(X|X>y)$. Το θεώρημα έτσι όπως διατυπώθηκε από τον ίδιο έχει ως εξής,

Θεώρημα 3.4.1 (Shanbhag (1970)) Εστω X μη αρνητική τυχαία μεταβλητή με συνεχή κατανομή της οποίας η αναμενόμενη τιμή υπάρχει. Η ικανή και αναγκαία συνθήκη για να ακολουθεί η X εκθετική κατανομή με $E(X) = m$ είναι

$$E(X|X > y) = y + m \ , \ y > 0 \qquad (3.4.1)$$

Ο Hamdan (1972) ακολουθώντας την ίδια μεθοδολογία επεχείρησε γενίκευση του παραπάνω συμπεράσματος και έτσι οδηγήθηκε σε χαρακτηρισμό της κατανομής Weibull και κατά συνέπεια και της εκθετικής κατανομής. Το θεώρημα έτσι όπως διατυπώθηκε από τον Hamdan έχει ως εξής,

Θεώρημα 3.4.2 (Hamdan (1972)) Εστω h μια γνησίως αύξουσα διαφορίσιμη συνάρτηση από το διάστημα $[a,b)$ στο $[0,\infty)$ και έστω $C>0$ σταθερά που ικανοποιεί τη σχέση $a < C < b$. Η τυχαία μεταβλητή X έχει αθροιστική κατανομή που δίνεται από την

$$P(X \le x) = \begin{cases} 0 & x < a \\ 1 - \exp\,[-h(x)/h(C)] & a < x < b \\ 1 & x \ge b \end{cases} \qquad (3.4.2)$$

αν και μόνον αν

$$E\left[h(X)\,|\,X>y\right] = h(y) + h(C),\ y \in [a,b)\ . \qquad (3.4.3)$$

Αξίζει να αναφερθεί ότι στην ειδική περίπτωση που $h(t) = t^c$, $t \in [0,\infty)$ και $c>0$ προκύπτει χαρακτηρισμός της κατανομής Weibull ενώ όταν $c = 1$ η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα ως εκθετική. Τέλος η περίπτωση $h(t) = -\ln(1-t)$, $t \in [0,1)$ οδηγεί σε χαρακτηρισμό της ομοιόμορφης κατανομής.

Στο ίδιο πλαίσιο το 1977 η Dimaki απέδειξε το παρακάτω θεώρημα που οδηγεί σε χαρακτηρισμό της κατανομής Pareto.

<u>Θεώρημα 3.4.3 (Dimaki (1977))</u> Εστω X τυχαία μεταβλητή που παίρνει τιμές μεγαλύτερες από τη μονάδα. Η ικανή και αναγκαία συνθήκη για να ακολουθεί η X κατανομή Pareto με συνάρτηση κατανομής την

$$F(x) = 1 - \theta^{\alpha}\, x^{-\alpha}, \qquad x > \theta,\ \theta > 0,\ \alpha > 1 \qquad (3.4.4)$$

είναι,

$$E(X|X>y) = (y/\theta)\ E(X) \quad \text{για όλα τα } y. \qquad (3.4.5)$$

Η μέθοδος που χρησιμοποιήθηκε στα τρία προαναφερθέντα

θεωρήματα γενικεύθηκε από τον Kotlarski (1978) ο οποίος με το παρακάτω θεώρημα δίνει την δυνατότητα χαρακτηρισμού ενός μεγάλου αριθμού συνεχών κατανομών.

Θεώρημα 3.4.4 (Kotlarski (1975)) Εστω X_0 και X δυο πραγματικές τυχαίες μεταβλητές με αντίστοιχες συναρτήσεις κατανομών F_0 και F. Ας υποθέσουμε ότι η F_0 είναι συνεχής και ότι υπάρχει ένα διάστημα $[a,b]$ όπου $-\infty \le a < b \le +\infty$ τέτοιο ώστε $F(a+0) = F_0(a) = 0$, $F(b-0) = F_0(b) = 1$ και ότι για κάθε $a < x < b$ ισχύει ότι $0 < F_0(x) < 1$ και $0 < F(x) < 1$. Εστω $h(.)$ μια πραγματική συνάρτηση η οποία απεικονίζει το διάστημα $[a,b]$ στο $[\alpha,\beta]$ όπου $-\infty \le \alpha < \beta \le +\infty$, συνεχής και διαφορίσιμη και έστω $h'(x) > 0$ για όλα τα $a < x < b$. Ας υποθέσουμε επίσης ότι τόσο η $E[h(X_0)]$ όσο και η $E[h(X)]$ υπάρχουν και είναι πεπερασμένες. Τότε οι X_0 και X ακολουθούν την ίδια κατανομή αν και μόνον αν ικανοποιείται μια από τις παρακάτω συνθήκες

$$E\left[h(X)\,|\,X \le y\right] = h(y) - \left[F_0(y)\right]^{-1} \int_a^y F_0(x)\, dh(x), \quad a<y\le b \qquad (3.4.6)$$

$$E\left[h(X)\,|\,X > y\right] = h(y) + \left[\overline{F}_0(y)\right]^{-1} \int_y^b \overline{F}_0(x)\, dh(x), \quad a\le y<b \qquad (3.4.7)$$

όπου $\overline{F}_0(x) = 1 - F_0(x), \qquad x \in \mathbb{R}$

Τα πορίσματα που ακολουθούν τονίζουν τη μεγάλη σημασία του θεωρήματος 3.4.4.

Πόρισμα 3.4.1 Εστω ότι ικανοποιούνται οι υποθέσεις του θεωρήματος 3.4.4 και έστω ότι,

$$\bar{F}_0(x) = x^{-\alpha} \qquad \text{δηλαδή,} \quad X_0 \sim \text{Pareto}\ (1,\alpha) \tag{3.4.8}$$

και

$$h(x) = x \tag{3.4.9}$$

Ικανή και αναγκαία συνθήκη για να ακολουθεί η X κατανομή Pareto $(1,\alpha)$ είναι η

$$E\left[X|X>y\right] = y\left(\alpha/(\alpha-1)\right).$$

Απόδειξη Αντικαθιστώντας τις σχέσεις (3.4.8) και (3.4.9) στην (3.4.7) προκύπτει ότι

$$E\left[X|X>y\right] = y + y^{\alpha}\int_y^{\infty} x^{-\alpha}\,dx$$

$$= y\,\frac{\alpha}{\alpha-1}$$

Παρατηρούμε ότι το θεώρημα 3.4.3 μπορεί να θεωρηθεί ως ειδική περίπτωση του θεωρήματος 3.4.4.

Πόρισμα 3.4.2 Εστω ότι ικανοποιούνται οι υποθέσεις του θεωρήματος 3.4.4 και έστω ότι,

$$\bar{F}_0(x) = e^{-\alpha x} \qquad \text{δηλαδή,} \quad X_0 \sim \text{Εκθετική}\ (1/\alpha) \tag{3.4.10}$$

και

$$h(x) = x \qquad (3.4.11)$$

Ικανή και αναγκαία συνθήκη για να ακολουθεί η X κατανομή εκθετική $(1/\alpha)$ είναι η

$$E\left[X|X>y\right] = y + (1/\alpha)$$

<u>Απόδειξη</u> Αντικαθιστώντας τις σχέσεις (3.4.10) και (3.4.11) στην (3.4.7) προκύπτει ότι,

$$E\left[X|X>y\right] = y + e^{\alpha y}\int_{y}^{\infty} e^{-\alpha x}\,dx$$

$$= y + \frac{1}{\alpha}$$

Παρατηρούμε ότι το θεώρημα 3.4.1 μπορεί να θεωρηθεί ως ειδική περίπτωση του θεωρήματος 3.4.4.

Το νόημα του θεωρήματος του Kotlarski είναι ότι αφού καθορισθεί η συνάρτηση h(.) και επιλεγεί η επιθυμητή συνάρτηση κατανομής F_0 προσδιορίζεται συνθήκη από τον υπολογισμό της $E\left[h(X)|X>y\right]$, για να ακολουθεί η τυχαία μεταβλητή X την κατανομή F_0. Για το σημαντικότατο αυτό θεώρημα προτείνουμε μια πιο εύχρηστη παραλλαγμένη μορφή που περιγράφεται από το παρακάτω θεώρημα.

<u>Θεώρημα 3.4.5</u> Εστω X συνεχής τυχαία μεταβλητή με συνάρτηση κατανομής $F_X(x)$ για όλα τα $x\geq\alpha$. Εστω επίσης h(.) μια γνησίως αύξουσα, διαφορίσιμη συνάρτηση. Τότε η συνάρτηση

$$E\left[h(X)\,|\,X>y\right] - h(y) = H(y), \qquad y \geq \alpha \tag{3.4.12}$$

ορίζει μονοσήμαντα την κατανομή της X.

Απόδειξη Με βάση τον ορισμό της $E\left[h(X)\,|\,X>y\right]$ προκύπτει ότι

$$E\left[h(X)\,|\,X>y\right]-h(y) = \frac{\int_y^{\infty} h(x)\ dF_X(x)}{1-F_X(x)} - h(y) .$$

Κατά συνέπεια η σχέση (3.4.12) γίνεται,

$$H(y) + h(y) = \frac{\int_y^{\infty} h(x)\ dF_X(x)}{\overline{F}_X(y)} ,$$

γράφεται δε ισοδύναμα ως εξής,

$$\overline{F}_X(y)\left[H(y)+h(y)\right] = \int_y^{\infty} h(x)\ dF_X(x)$$

$$= E\left[h(X)\right] + \int_{\alpha}^{y} h(x)\ d\overline{F}_X(x)$$

$$= E\left[h(X)\right] + h(y)\overline{F}_X(x) - h(\alpha) - \int_{\alpha}^{y} h'(x)\ \overline{F}_X(x)\ dx .$$

Συνεπώς,

$$H(y)\ \overline{F}_X(y) = E\left[h(X)\right] - h(\alpha) - \int_{\alpha}^{y} h'(x)\ \overline{F}_X(x)\ dx .$$

Παραγωγίζοντας ως προς y προκύπτει ότι,

$$H'(y)\,\bar{F}_X(y) + H(y)\,\bar{F}'_X(y) = -\,h'(y)\,\bar{F}_X(y)\ ,$$

ή ισοδύναμα,

$$\frac{\bar{F}'_X(y)}{\bar{F}_X(y)} = -\,\frac{h'(y) + H'(y)}{H(y)}\ .$$

Η λύση της διαφορικής αυτής εξίσωσης οδηγεί στη σχέση

$$\int \frac{\bar{F}'_X(y)}{\bar{F}_X(y)}\,dy = -\int \frac{h'(y) + H'(y)}{H(y)}\,dy + C^*$$

και τελικά,

$$\bar{F}_X(x) = C_1 \exp\left\{-\int \frac{h'(y)+H'(y)}{H(y)}\,dy\right\}\ . \qquad (3.4.13)$$

Από τη σχέση (3.4.13) προκύπτει ότι η $\bar{F}_X(x)$ ορίζεται μονοσήμαντα από τις συναρτήσεις h(y) και H(y).

Παρατηρούμε ότι το θεώρημα 3.4.5 οδηγεί σε χαρακτηρισμό της κατανομής της τυχαίας μεταβλητής X μέσω της σχέσης (3.4.12) αφού προκαθορισθεί η μορφή της συνάρτησης h(.). Ακολουθούν ορισμένα πορίσματα που τεκμηριώνουν τη πρακτική σημασία της παραλλαγμένης αυτής μορφής του θεωρήματος του Kotlarski.

Αρχικά να παρατηρήσουμε ότι στην ειδική περίπτωση που $h(x) = x$ και $H(y) = m = E(X)$ το θεώρημα 3.4.5 ταυτίζεται με το συμπέρασμα του Shanbhag (1970).

Σαν πρώτη ειδική περίπτωση του θεωρήματος 3.4.5 θα θεωρήσουμε εκείνη που προκύπτει όταν $h(x) = x^{\alpha}$ και $H(y) = c$. Ετσι έχουμε το παρακάτω πόρισμα.

<u>Πόρισμα 3.4.3</u> Εστω X τυχαία μεταβλητή με συνεχή συνάρτηση κατανομής έστω $F_X(x)$ που ορίζεται στο διάστημα $[0,\infty)$. Αν

$$E\left[X^{\alpha} \mid X > y\right] = y^{\alpha} + c, \quad \text{για} \quad \alpha>0,\ c>0 \quad \text{και} \quad y \in [0,\infty) ,$$

τότε η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα και είναι η Weibull.

Στην ειδική περίπτωση που $\alpha=1$ η κατανομή της X ορίζεται μονοσήμαντα ως εκθετική.

<u>Απόδειξη</u> Αντικαθιστώντας τις σχέσεις $h(x) = x^{\alpha}$ και $H(y) = c$ στην (3.4.13) προκύπτει ότι,

$$F_X(x) = 1 - C_1 \exp\left[-\frac{1}{c} x^{\alpha}\right].$$

Επειδή δε $F_X(0) = 0 \Rightarrow C_1=1$.

Αρα,

$$F_X(x) = 1 - \exp\left[-\frac{1}{c} x^{\alpha}\right] .$$

Συνεπώς η τυχαία μεταβλητή X ακολουθεί την κατανομή Weibull με

παραμέτρους $\alpha>0$ και $c^{-1}>0$ για $x \in [0,\infty)$.

Ειναι προφανές ότι αν $\alpha=1$ η τυχαία μεταβλητή X ακολουθεί την εκθετική κατανομή.

Δεύτερη ειδική περίπτωση του θεωρήματος 3.4.5 την οποία θα μελετήσουμε είναι εκείνη που προκύπτει όταν $h(x) = -\ln (1-x^{\alpha})$ και $H(y)=1$. Μπορούμε λοιπόν να διατυπώσουμε το παρακάτω πόρισμα.

<u>Πόρισμα 3.4.4</u> Εστω X τυχαία μεταβλητή με συνεχή συνάρτηση κατανομής έστω $F_X(x)$ που ορίζεται στο διάστημα [0,1). Αν

$$E\left[-\ln (1-X^{\alpha}) \mid X > y\right] = -\ln (1-y^{\alpha}) + 1 \ , \quad \text{για } \alpha>0 \text{ και } y \in [0,1),$$

τότε η κατανομή της X ορίζεται μονοσήμαντα και είναι η power function.

Στην ειδική περίπτωση που $\alpha=1$ δηλαδή όταν $h(x) = -\ln(1-\alpha)$ η κατανομή της X ορίζεται μονοσήμαντα ως ομοιόμορφη.

<u>Απόδειξη</u> Αντικαθιστώντας όπου $h(x) = -\ln (1-x^{\alpha})$ και $H(y) = 1$ στη σχεση (3.4.13) προκύπτει ότι,

$$F_X(x) = 1 - C_1 \exp\left\{\ln (1-x^{\alpha})\right\} .$$

Επειδή δε $F_X(0) = 0 \Rightarrow C_1 = 1$.

Κατά συνέπεια

$$F_X(x) = x^{\alpha}$$

Αρα η τυχαία μεταβλητή X ακολουθεί την κατανομή power function με $\alpha>0$ και $x \in [0,1)$.

Είναι προφανές ότι όταν $\alpha=1$ η κατανομή της X είναι η ομοιόμορφη. στο [0,1).

Το τρίτο πόρισμα του θεωρήματος 3.4.5 χαρακτηρίζει την κατανομή Pareto.

<u>Πόρισμα 3.4.5</u> Εστω X τυχαία μεταβλητή με συνεχή συνάρτηση κατανομής έστω $F_X(x)$ που ορίζεται στο διάστημα $[1,\infty)$. Αν

$$E\left[\ln X \mid X > y\right] = \ln y + c, \quad \text{για} \quad c > 0 \quad \text{και} \quad y \in [1,\infty),$$

τότε η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα ως Pareto $(1,c^{-1})$.

<u>Απόδειξη</u> Αντικαθιστώντας όπου $h(x) = \ln X$ και $H(y) = c$ στη σχέση (3.4.13) προκύπτει ότι

$$F_X(x) = 1 - C_1 \exp\left[-\frac{1}{c} \ln x\right].$$

Δηλαδή

$$F_X(x) = 1 - C_1 x^{-(1/c)}.$$

Επειδή δε $F_X(1) = 0 \Rightarrow C_1 = 1$.

Αρα,

$$F_X(x) = 1 - x^{-(1/c)},$$

δηλαδή η τυχαία μεταβλητή X ορίζεται μονοσήμαντα και κατανέμεται ως Pareto $(1,c^{-1})$.

Η τελευταία ειδική περίπτωση του θεωρήματος 3.4.5 με την οποία θα ασχοληθούμε είναι εκείνη η οποία προκύπτει αν θεωρήσουμε την $h(x) = \ln(1+x^{\alpha})$ και την $H(y) = c$. Το συμπέρασμα συνοψίζεται από το πόρισμα 3.4.6.

<u>Πόρισμα 3.4.6</u> Εστω X τυχαία μεταβλητή με συνεχή συνάρτηση κατανομής έστω $F_X(x)$ που ορίζεται στο διάστημα $[0,\infty)$. Αν

$$E\left[\ln(1+X^{\alpha}) \mid X > y\right] = \ln(1+y^{\alpha}) + c \quad \text{για} \quad \alpha>0,\ c>0 \quad \text{και} \quad y \in [0,\infty)$$

τότε η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα και είναι η κατανομή Burr.
Στην ειδική περίπτωση που $\alpha=1$ δηλαδή όταν $h(x) = \ln(1+x)$ η κατανομή της X ορίζεται μονοσήμαντα ως Beta II.

<u>Απόδειξη</u> Θέτοντας όπου $h(x) = \ln(1+x^{\alpha})$ και $H(y) = c$ στη σχέση (3.4.13) προκύπτει ότι,

$$F_X(x) = 1 - C_1 \exp\left[-\frac{1}{c}\ln(1+x^{\alpha})\right].$$

Επειδή δε, $F_X(0) = 0 \Rightarrow C_1 = 1$.

Κατά συνέπεια

$$F_X(x) = 1 - \left[1+x^{\alpha}\right]^{-1/c}.$$

Αρα η τυχαία μεταβλητή X κατανέμεται ως Burr με παραμέτρους $\alpha > 0$ και $c^{-1} > 0$.

Είναι προφανές ότι όταν $\alpha = 1$ η τυχαία μεταβλητή X ακολουθεί την κατανομή Beta II.

3.5 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΒΑΣΙΣΜΕΝΕΣ ΣΤΗΝ ΥΠΟ ΣΥΝΘΗΚΗ ΑΝΑΜΕΝΟΜΕΝΗ ΤΙΜΗ. ΔΙΑΚΡΙΤΗ ΠΕΡΙΠΤΩΣΗ.

Στην ενότητα αυτή θα ασχοληθούμε με το θέμα του χαρακτηρισμού διακριτών κατανομών στην περίπτωση που έχουμε πληροφορίες για τη συνάρτηση $E\left[h(X)\,|\,X>y\right]$. Ο Shanbhag (1970) διατύπωσε επίσης και έναν ενδιαφέροντα χαρακτηρισμό για τη γεωμετρική κατανομή. Το θεώρημα έτσι όπως διατυπώθηκε από τον ίδιο έχει ως εξής,

<u>Θεώρημα 3.5.1 (Shanbhag (1970))</u> Εστω X τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές της οποίας η αναμενόμενη τιμή υπάρχει. Η X κατανέμεται σύμφωνα με τη γεωμετρική κατανομή αν και μόνον αν

$$E\left[X\,|\,X>y\right] = [y] + 1 + E[X] \qquad \text{για όλα τα } y. \qquad (3.5.1)$$

Στην εργασία των Dimaki and Xekalaki (1991) αποδεικνύεται ανάλογη ιδιότητα για την κατανομή Yule. Το συμπέρασμά της συνοψίζεται από το παρακάτω θεώρημα.

<u>Θεώρημα 3.5.2 (Dimaki and Xekalaki (1991)</u> Εστω X τυχαία μεταβλητή με τιμές ακέραιες και μεγαλύτερες από την μονάδα της

οποίας η αναμενόμενη τιμή υπάρχει. Η Χ κατανέμεται σύμφωνα με τη μετατοπισμένη κατανομή Yule (ρ) αν και μόνον αν

$$E\left[X|X>y\right] = y\,E(X) \qquad \text{για όλα τα } y. \qquad (3.5.2)$$

Αξίζει να σημειωθεί η δυαδικότητα που υπάρχει μεταξύ των θεωρημάτων τα οποία αφορούν τις κατανομές εκθετική και γεωμετρική δηλαδή των 3.4.1 και 3.5.1 αφενός, καί εκείνων που αφορούν τις κατανομές Pareto και Yule δηλαδή των 3.4.3 και 3.5.2 αφετέρου.

Το συμπέρασμα του θεωρήματος 3.5.2 γενικεύεται για την περίπτωση που η κατανομή της τυχαίας μεταβλητής Χ είναι η μονοδιάστατη γενικευμένη Waring. Πριν προχωρήσουμε στην απόδειξη του θεωρήματος θα διατυπώσουμε το παρακάτω Λήμμα.

<u>Λήμμα 3.5.1 (Xekalaki, 1977)</u> Εστω Χ τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές. Τότε,

$$P(X>r) = \frac{1}{\rho}\,(\alpha+r)\,P(X=r), \qquad r = 0,1,2,\ldots \ \alpha>0,\ \rho>0 \qquad (3.5.3)$$

αν και μόνον αν η Χ ακολουθεί τη γενικευμένη μονοδιάστατη κατανομή Waring $(\alpha,1,\rho)$, $\alpha>0$, $\rho>0$.

<u>Θεώρημα 3.5.3</u> Εστω Χ τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές της οποίας η αναμενόμενη τιμή υπάρχει. Η Χ κατανέμεται σύμφωνα με τη μονοδιάστατη γενικευμένη κατανομή Waring που έχει συνάρτηση πιθανότητας την

$$p_x = P(X=x) = \frac{\rho_{(1)}}{(\alpha+\rho)_{(1)}} \frac{\alpha_{(x)}\ 1_{(x)}}{(\alpha+1+\rho)_{(x)}} \frac{1}{x!}, \quad x=0,1,2, \qquad (3.5.4)$$

$$\alpha,\rho>0.$$

αν και μόνον αν

$$E\left[X|X>y\right] = \mu + (y+1)\mu_1$$

όπου

$$\mu = E(X) = \frac{\alpha}{\rho-1}$$

και

$\mu_1 = \frac{\rho}{\rho-1}$, δηλαδή η αναμενόμενη τιμή της αντίστοιχης μετατοπισμένης Yule (ρ).

<u>Απόδειξη</u>

Ευθύ: Εστω ότι η τυχαία μεταβλητή X ακολουθεί τη μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha,1,\rho)$. Τότε με βάση τον ορισμό της $E\left[X|X>y\right]$ και το συμπέρασμα του λήμματος 3.5.1 έχουμε ότι,

$$E\left[X|X>y\right] = \sum_{x=y+1}^{\infty} x\ P(X|X>x)$$

$$= \sum_{x=y+1}^{\infty} x \frac{P(X=x)}{P(X>y)}$$

$$= \frac{\rho}{\alpha+y} \frac{1}{P(X=y)} \sum_{x=y+1}^{\infty} x \; P(X=x)$$

$$= \frac{\rho}{\alpha+y} \frac{1}{P(X=y)} \sum_{x=y+1}^{\infty} x \frac{\rho_{(1)}}{(\alpha+\rho)_{(1)}} \frac{\alpha_{(x)} \; 1_{(x)}}{(\alpha+\rho+1)_{(x)}} \frac{1}{x!}$$

$$= \frac{\rho}{\alpha+y} \frac{1}{P(X=y)} \sum_{x=0}^{\infty} (x+y+1) \frac{\rho}{\alpha+\rho} \frac{\alpha_{(x+y+1)} \; 1_{(x+y+1)}}{(\alpha+\rho+1)_{(x+y+1)}} \frac{1}{(x+y+1)!}$$

$$= \frac{\rho}{\alpha+y} \frac{1}{P(X=y)} \sum_{x=0}^{\infty} (x+y+1) \frac{\rho}{\alpha+\rho} \frac{\alpha_{(y+1)} \; (\alpha+y+1)_{(x)}}{(\alpha+\rho+1)_{(y+1)} \; (\alpha+\rho+y+2)_{(x)}}$$

$$= \frac{\rho}{\alpha+\rho} \frac{1}{P(X=y)} \frac{\rho}{\alpha+\rho} \frac{\alpha_{(y)} \; (\alpha+y)_{(1)}}{(\alpha+\rho+1)_{(y)} (\alpha+\rho+y+1)_{(1)}} \sum_{x=0}^{\infty} (x+y+1) \frac{(\alpha+y+1)_{(x)} \, 1_{(x)}}{(\alpha+\rho+y+2)_{(x)}} \frac{1}{x!}$$

$$= \sum_{x=0}^{\infty} x \frac{\rho_{(1)}}{(\alpha+\rho+y+1)_{(1)}} \frac{(\alpha+y+1)_{(x)} \, 1_{(x)}}{(\alpha+\rho+y+2)_{(x)}} \frac{1}{x!} +$$

$$+ (y+1) \sum_{x=0}^{\infty} \frac{\rho_{(1)}}{(\alpha+\rho+y+1)_{(1)}} \frac{(\alpha+y+1)_{(x)} \, 1_{(x)}}{(\alpha+\rho+y+2)_{(x)}} \frac{1}{x!}$$

$$= \frac{\alpha+y+1}{\rho-1} + y + 1$$

$$= \frac{\alpha}{\rho-1} + (y+1)\,\frac{\rho}{\rho-1}$$

$$= \mu + (y+1)\mu_1$$

Αντίστροφο: Εστω ότι $E\left[X \mid X>y\right] = \mu + (y+1)\,\mu_1$. Η σχέση αυτή γράφεται ισοδύναμα,

$$\sum_{x=y+1}^{\infty} x\, P(X=x) = \mu\, P(X>y) + (y+1)\,\mu_1\, P(X>y) \qquad (3.5.5)$$

Εξειδικεύοντας τη σχέση (3.5.5) για y=r και y=r+1 και αφαιρώντας κατά μέλη τις εξισώσεις που προκύπτουν έχουμε ότι,

$$y\, P(X=y) = \mu\, P(X=y) + y\mu_1\left[P(X>y-1) - P(X>y)\right] - \mu_1\, P(X>Y)$$

ή

$$(y-\mu-y\mu_1)\, P(X=y) = -\mu_1\, P(X>y)$$

Εφόσον η E(X) υπάρχει, συνεπάγεται ότι $\rho > 1$ και επομένως η μ_1 υπάρχει και είναι μεγαλύτερη από τη μονάδα. Αρα η τελευταία σχέση γράφεται

$$P(X>y) = \left(\frac{\mu_1-1}{\mu_1}\, y + \frac{\mu}{\mu_1}\right) P(X=y), \quad y = 0,1,2,\ldots \qquad (3.5.6)$$

Αυτή όμως είναι χαρακτηριστική για την γενικευμένη κατανομή Waring $\left(\alpha = \frac{\mu}{\mu_1-1}\,,\ 1,\ \rho = \frac{\mu_1}{\mu_1-1}\right)$ σύμφωνα με το λήμμα 3.5.1.

Στη συνέχεια θα διατυπώσουμε ένα γενικό θεώρημα το οποίο θα μας επιτρέψει να αποδείξουμε ως πορίσματα γνωστά συμπεράσματα για συγκεκριμένες κατανομές. Το θεώρημα αυτό θα μπορούσε να θεωρηθεί ως το διακριτό ανάλογο του θεωρήματος του Kotlarski που αναφέραμε στην προηγούμενη ενότητα.

Θεώρημα 3.5.4 Εστω X_0 και X δυο διακριτές τυχαίες μεταβλητές που παίρνουν τιμές στους θετικούς ακέραιους και έχουν συναρτήσεις κατανομών F_0 και F_X αντίστοιχα. Εστω επίσης ότι $F_X(\alpha+0) = F_0(\alpha) = 0$ για κάποιο $\alpha>0$. Ας θεωρήσουμε επίσης μια πραγματική συνάρτηση $h(.)$ τέτοια ώστε $E\left[h(X_0)\right] < \infty$ όσο και $E\left[h(X)\right] < \infty$. Τότε οι τυχαίες μεταβλητές X_0 και X ακολουθούν την ίδια κατανομή αν και μόνον αν ικανοποιείται η σχέση

$$E\left[h(X)\,|\,X>y\right] = h(y+1)+\left[\overline{F}_0(y)\right]^{-1}\sum_{x=y+1}^{\infty}\overline{F}_0(x)\left[h(x+1)-h(x)\right] \qquad (3.5.7)$$

όπου $\overline{F}_0(.) = 1 - F_0(.)$

Απόδειξη

Ευθύ: Με βάση τον ορισμό της $E\left[h(X)\,|\,X>y\right]$ έχουμε διαδοχικά ότι,

$$E\left[h(X)\,|\,X>y\right] = \sum_{x=y+1}^{\infty} h(x)\; P(X=x\,|\,X>y)$$

$$= \sum_{x=y+1}^{\infty} h(x)\;\frac{P(X=x)}{P(X>y)}$$

$$= \sum_{x=y+1}^{\infty} h(x) \frac{P(X \geq x) - P(X \geq x+1)}{P(X>y)}$$

$$= \frac{1}{P(X>y)} \left[\sum_{x=y+1}^{\infty} h(x)\ P(X \geq x) - \sum_{x=y+1}^{\infty} h(x)\ P(X \geq x+1) \right]$$

$$= \frac{1}{P(X>y)} \left[\sum_{x=y}^{\infty} h(x+1)\ P(X \geq x+1) - \sum_{x=y+1}^{\infty} h(x)\ P(X \geq x+1) \right]$$

$$= \frac{1}{P(X>y)} \left[h(y+1)P(X \geq y+1) + \sum_{x=y+1}^{\infty} h(x+1)P(X \geq x+1) - \sum_{x=y+1}^{\infty} h(x)P(X \geq x+1) \right]$$

$$= h(y+1) + \frac{1}{P(X>y)} \sum_{x=y+1}^{\infty} \left[h(x+1)-h(x) \right] P(X>x)$$

$$= h(y+1) + \frac{1}{\overline{F}_X(y)} \sum_{x=y+1}^{\infty} \left[h(x+1)-h(x) \right] \overline{F}_X(x)$$

Δηλαδή,

$$E\left[h(X) \mid X>y \right] = h(y+1) + \left[\overline{F}_X(y) \right]^{-1} \sum_{x=y+1}^{\infty} \overline{F}_X(x) \left[h(x+1)-h(x) \right] \qquad (3.5.8)$$

Είναι προφανές ότι αν $\overline{F}_X = \overline{F}_0$ η σχέση (3.5.8) στην οποία καταλήξαμε ισοδυναμεί με την (3.5.7).

Αντίστροφο Εστω ότι ισχύει η σχέση (3.5.7). Οπως προηγουμένως, αποδεικνύεται ότι ,

$$E\left[h(X)\,|\,X>y\right] = h(y+1)+\left[\overline{F}_X(y)\right]^{-1}\sum_{x=y+1}^{\infty}\overline{F}_X(x)\left[h(x+1)-h(x)\right] \qquad (3.5.9)$$

Συγκρίνοντας τις σχέσεις (3.5.7) και (3.5.9) συμπεραίνουμε ότι,

$$\left[\overline{F}_0(y)\right]^{-1}\sum_{x=y+1}^{\infty}\overline{F}_0(x)\left[h(x+1)-h(x)\right]=\left[\overline{F}_X(y)\right]^{-1}\sum_{x=y+1}^{\infty}\overline{F}_X(x)\left[h(x+1)-h(x)\right] \qquad (3.5.10)$$

Εστω

$$K_0(y) = \sum_{x=y+1}^{\infty}\overline{F}_0(x)\left[h(x+1)-h(x)\right] \qquad (3.5.11)$$

και

$$K(y) = \sum_{x=y+1}^{\infty}\overline{F}_X(x)\left[h(x+1)-h(x)\right] . \qquad (3.5.12)$$

Από τη σχέση (3.5.11) προκύπτει ότι

$$K_0(y+1) = \sum_{x=y+2}^{\infty}\overline{F}_0(x)\left[h(x+1)-h(x)\right] . \qquad (3.5.13)$$

Αφαιρώντας κατά μέλη τις σχέσεις (3.5.11) και (3.5.13) προκύπτει ότι,

$$K_0(y) - K(y) = \overline{F}_0(y+1)\left[h(y+2) - h(y+1)\right]$$

δηλαδή,

$$\overline{F}_0(y+1) = \frac{K_0(y) - K_0(y+1)}{h(y+2)-h(y+1)} \quad .$$

Τελικά δηλαδή

$$\overline{F}_0(y) = \frac{K_0(y) - K_0(y-1)}{h(y) - h(y+1)} \quad . \qquad (3.5.14)$$

Με όμοιο τρόπο, ξεκινώντας από τη σχέση (3.5.12) καταλήγουμε στο

$$\overline{F}_x(y) = \frac{K(y) - K(y-1)}{h(y)-h(y+1)} \quad . \qquad (3.5.15)$$

Αντικαθιστώντας τις σχέσεις (3.5.11), (3.5.12), (3.5.14) και (3.5.15) στη σχέση (3.5.10) έχουμε ότι,

$$\frac{K_0(y)}{K_0(y) - K_0(y-1)} = \frac{K(y)}{K(y) - K(y-1)} \quad . \qquad (3.5.16)$$

Από τη σχέση (3.5.19) προκύπτει ότι,

$K(y) = C\ K_0(y)$, όπου C σταθερά.

Λόγω των (3.5.11) και (3.5.12) η τελευταία αυτή σχέση γράφεται,

$$\sum_{x=y+1}^{\infty} \overline{F}_x(x)\left[h(x+1)-h(x)\right] = C \sum_{x=y+1}^{\infty} \overline{F}_0(x)\left[h(x+1)-h(x)\right] \quad . \qquad (3.5.17)$$

Θέτοντας όπου y το y+1 η σχέση (3.5.17) γίνεται,

$$\sum_{x=y+2}^{\infty} \overline{F}_X(x)\left[h(x+1)-h(x)\right] = C \sum_{x=y+2}^{\infty} \overline{F}_0(x)\left[h(x+1)-h(x)\right]. \qquad (3.5.18)$$

Αφαιρώντας τις σχέσεις (3.5.17) και (3.5.18) κατά μέλη προκύπτει ότι,

$$\overline{F}_X(y+1)\left[h(y+2) - h(y+1)\right] = C\,\overline{F}_0(y+1)\left[h(y+1) - h(y+1)\right].$$

Η τελευταία αυτή σχέση είναι ισοδύναμη με την,

$$\overline{F}_X(y+1) = C\,\overline{F}_0(y+1)$$

δηλαδή,

$$\overline{F}_X(y) = C\,\overline{F}_X(y)$$

Απότην οποία τελικά προκύπτει ότι,

$$1 - F_X(y) = C - C\,F_0(y).$$

Επειδή όμως, $F_x(\alpha) = F_0(\alpha) = 0$, συνεπάγεται ότι $C = 1$, και άρα,

$$F_X(y) = F_0(y), \quad y \geq \alpha \ , \ \alpha > 0.$$

Στη συνέχεια αποδεικνύεται ότι γνωστοί χαρακτηρισμοί διακριτών κατανομών, μπορούν να θεωρηθούν ειδικές περιπτώσεις του θεωρήματος 3.5.4.

Πόρισμα 3.5.1 Εστω ότι ικανοποιούνται οι υποθέσεις του θεωρήματος 3.5.4 και έστω ότι,

$$\bar{F}_0(x) = (1-p)^x \quad \text{δηλαδή } X_0 \sim \text{γεωμετρική κατανομή} \tag{3.5.19}$$

και

$$h(x) = x \tag{3.5.20}$$

Ικανή και αναγκαία συνθήκη για να ακολουθεί η X γεωμετρική κατανομή είναι η

$$E\left[X|X>y\right] = y + \frac{1}{p} \; . \tag{3.5.21}$$

Απόδειξη Αντικαθιστώντας τις σχέσεις (3.5.19) και (3.5.20) στη σχέση (3.5.7) προκύπτει ότι,

$$E\left[X|X>y\right] = (y+1) + \frac{1}{(1-p)^y} \sum_{x=y+1}^{\infty} (1-p)^x$$

$$= (y+1) + \frac{1-p}{p}$$

$$= y + \frac{1}{p}$$

Η σχέση (3.5.21) όμως, είναι χαρακτηριστική ιδιότητα της γεωμετρικής κατανομής που απεδείχθει από τον Shanbhag (1970).
Το πόρισμα 3.5.2 καθορίζει την ικανή και αναγκαία συνθήκη για να ακολουθεί η τυχαία μεταβλητή X κατανομη Yule.

Πόρισμα 3.5.2 Εστω ότι ικανοποιούνται οι υποθέσεις του θεωρήματος 3.5.4 και έστω ότι,

$$\overline{F}_0(x) = \frac{x!}{(\rho+1)_{(x)}} \quad \text{δηλαδή} \quad X_0 \sim \text{μετατοπισμένη Yule } (\rho) \qquad (3.5.22)$$

και

$$h(x) = x \qquad (3.5.23)$$

Ικανή και αναγκαία συνθήκη για να ακολουθεί η X μετατοπισμένη Yule (ρ) είναι η

$$E\left[X|X>y\right] = (y+1)\,\frac{\rho}{\rho-1}\,. \qquad (3.5.24)$$

Απόδειξη Αντικαθιστώντας τις σχέσεις (3.5.22) και (3.5.23) στη σχέση (3.5.7) προκύπτει ότι,

$$E\left[X|X>y\right] = y+1 + \frac{\sum_{x=y+1}^{\infty} x!/(\rho+1)_{(x)}}{y!/(\rho+1)_{(y)}}$$

$$= y+1 + \frac{\sum_{x=0}^{\infty} 1_{(x+y+1)}/(\rho+1)_{(x)}}{y!/(\rho+1)_{(y)}}$$

$$= y+1 + \frac{(y+1)!}{(\rho+1)_{(y+1)}} \frac{\sum_{x=0}^{\infty} (y+2)_{(x)} / (\rho+y+2)_{(x)}}{y!/(\rho+1)_{(y)}}$$

$$= y + 1 + \frac{(y+1)!}{(\rho+1)_{(y)} \ (\rho+y+1)_{(1)}} \frac{(\rho+y+1)/(\rho-1)}{y!/(\rho+1)_{(y)}}$$

$$= y + 1 + \frac{y+1}{1} \frac{1}{\rho-1}$$

$$= (y+1) \frac{\rho}{\rho-1}$$

Η σχέση (3.5.24) όμως είναι η χαρακτηριστική ιδιότητα της κατανομής Yule που αποδεικνύεται από τις Dimaki and Xekalaki (1991) , και το πόρισμα 3.5.2 αποτελεί επομένως μια αναδιατύπωση του δικού τους σχετικού θεωρήματος.

Το τελευταίο πόρισμα προσδιορίζει την ικανή και αναγκαία συνθήκη για να ακολουθεί η τυχαία μεταβλητή X τη μονοδιάστατη γενικευμένη κατανομή Waring.

<u>Πόρισμα 3.5.3</u> Εστω ότι ικανοποιούνται οι υποθέσεις του θεωρήματος 3.5.4 και έστω ότι,

$$\overline{F}_0(x) = \frac{\alpha+x}{\rho} P(X_0=x) \qquad (3.5.25)$$

όπου

$$P(X_0=x) = p_x = \frac{\rho}{\alpha+\rho}\,\frac{\alpha_{(x)}\;1_{(x)}}{(\alpha+\rho+1)_{(x)}}\,\frac{1}{x!}, \qquad x=0,1,2,\ldots \qquad (3.5.26)$$

$$\alpha,\rho>0$$

δηλαδή, $X_0 \sim$ μονοδιάστατη γενικευμένη Waring $(\alpha,1;\rho)$

και

$$h(x) = x \qquad (3.5.27)$$

Ικανή και αναγκαία συνθήκη για να ακολουθεί η X μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha,1;\rho)$ είναι η

$$E\left[X|X>y\right] = \frac{\alpha}{\rho-1} + (y+1)\frac{\rho}{\rho-1}$$

Απόδειξη Αντικαθιστώντας τις σχέσεις (3.5.25) και (3.5.27) στη σχέση (3.5.7) προκύπτει ότι,

$$E\left[X|X>y\right] = (y+1) + \left[\frac{\alpha+y}{\rho}\,P(X_0=y)\right]^{-1}\sum_{x=y+1}^{\infty}\frac{\alpha+x}{\rho}\,P(X_0=x)$$

$$= (y+1) + \frac{\alpha}{\rho}\left[\frac{\alpha+y}{\rho}\,P(X_0=y)\right]^{-1}\sum_{x=y+1}^{\infty}P(X_0=x) +$$

$$+ \frac{1}{\rho}\left[\frac{\alpha+y}{\rho}\,P(X_0=y)\right]^{-1}\sum_{x=y+1}^{\infty}x\,P(X_0=x)$$

$$= (y+1) + \frac{\alpha}{\rho}\left[\frac{\alpha+y}{\rho}\,P(X_0=y)\right]^{-1}\sum_{x=y+1}^{\infty}P(X_0=x) +$$

$$+ \frac{1}{\rho}\left[P(X_0>y)\right]^{-1}\sum_{x=y+1}^{\infty}x\,P(X_0=x)$$

$$= (y+1) + \frac{\alpha}{\rho}\left[\frac{\alpha+y}{\rho}\,P(X_0=y)\right]^{-1}P(X_0>y) + \frac{1}{\rho}\sum_{x=y+1}^{\infty}x\,\frac{P(X_0=x)}{P(X_0>x)}\ .$$

Λαμβάνοντας υπόψη τη σχέση (3.5.27) της υπόθεσης καταλήγουμε στο ότι,

$$E\left[X|X>y\right] = (y+1) + \frac{\alpha}{\rho}\left[P(X_0>y)\right]^{-1}P(X_0>y) + \frac{1}{\rho}\,E\left[X_0|X_0>y\right]\ . \quad (3.5.28)$$

Από τις υποθέσεις όμως του πορίσματος η τυχαία μεταβλητή X_0 ακολουθεί μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha,1;\rho)$. Κατά συνέπεια με βάση το ευθύ του θεωρήματος 3.5.3 η σχέση (3.5.28) γίνεται,

$$E\left[X|X>y\right] = y + 1 + \frac{\alpha}{\rho} + \frac{1}{\rho}\left[\frac{\alpha}{\rho-1} + (y+1)\,\frac{\rho}{\rho-1}\right]$$

$$= \frac{\alpha}{\rho-1} + (y+1)\,\frac{\rho}{\rho-1}\ .$$

Κατ' αναλογία με τη συνεχή περίπτωση μπορούμε να διατυπώσουμε το θεώρημα 3.5.4 και σε μια άλλη εξίσου χρήσιμη μορφή. Η παραλλαγή του διακριτού αναλόγου του θεωρήματος του Kotlarski δίνεται από το θεώρημα που ακολουθεί.

Θεώρημα 3.5.5 Εστω X διακριτή τυχαία μεταβλητή που ορίζεται στο σύνολο $\{m,m+1,...\}$, $m \in I^+ \cup \{0\}$ και η οποία έχει συνάρτηση κατανομής $F_X(x)$. Εστω επίσης h(.) μια γνησίως αύξουσα διαφορίσιμη συνάρτηση. Τότε η συνάρτηση

$$E\left[h(X)|X>y\right] - h(y) = H(y), \quad y=m,m+1,... \quad m \in I^+ \cup \{0\} \qquad (3.5.29)$$

ορίζει μονοσήμαντα την κατανομή της X.

Απόδειξη Με βάση τον ορισμό της $E\left[h(X)|X>y\right]$ και ακολουθώντας την επιχειρηματολογία που ακολουθήθηκε για την απόδειξη του θεωρήματος 3.5.4 έχουμε

$$E\left[h(X)|X>y\right] = h(y+1) + \frac{1}{\overline{F}_X(y)} \sum_{x=y+1}^{\infty} \left[h(x+1)-h(x)\right] \overline{F}_X(x).$$

Επομένως η σχέση (3.5.29) γίνεται,

$$\frac{1}{\overline{F}_X(y)} \sum_{x=y+1}^{\infty} \left[h(x+1)-h(x)\right] \overline{F}_X(x) = H(y) + h(y) - h(y-1)$$

ή

$$\sum_{x=y+1}^{\infty} h(x+1)\bar{F}_X(x) - \sum_{x=y+1}^{\infty} h(x)\bar{F}_X(x) = \bar{F}_X(y)\left[H(y)+h(y)-h(y+1)\right] \qquad (3.5.30)$$

Εξειδικεύοντας τη σχέση (3.5.30) για y=r και y=r+1 και αφαιρώντας κατά μέλη τις εξισώσεις που προκύπτουν έχουμε ότι,

$$H(r+1)\ \bar{F}_X(r+1) - \left[H(r)+h(r)-h(r+1)\right]\ \bar{F}_X(x) = 0\ .$$

Επειδή η h(.) είναι γνησίως αύξουσα, έπεται από την (3.5.29) ότι H(r) > 0 για κάθε r. Αρα η τελευταία σχέση γράφεται

$$P(X\geq r+1) - \frac{H(r-1)+h(r-1)-h(r)}{H(r-1)}\ P(X\geq r) = 0.$$

Συνεπώς δεδομένου ότι P(X≥m) = 1 η μοναδική λύση της εξίσωσης αυτής διαφορών δίνεται από τη σχέση,

$$P(X\geq y) = \prod_{r=m}^{y-1} \frac{H(r-1)+h(r-1)-h(r)}{H(r-1)} \qquad (3.5.31)$$

Από τη σχέση (3.5.31) προκύπτει ότι η P(X≥y) ορίζεται μονοσήμαντα από τις συναρτήσεις h(y) και H(y).

Τα πορίσματα που ακολουθούν τεκμηριώνουν την πρακτική αξία του θεωρήματος 3.5.5.

Πόρισμα 3.5.4 Εστω X διακριτή τυχαία μεταβλητή με τιμές στο I^+ . Αν

$$E\left[X|X>y\right] - y = C \ , \qquad \text{για} \quad C > 1 \quad \text{και} \quad y = 1,2,...$$

Τότε η κατανομή της X ορίζεται μονοσήμαντα και είναι η γεωμετρική.

Απόδειξη Αντικαθιστώντας τις σχέσεις $h(x) = x$ και $H(y) = C$ στην (3.5.32) προκύπτει ότι,

$$P(X\geq y) = \left(1 - \frac{1}{C}\right)^{y-1}$$

Δηλαδή το ζητούμενο.

Πόρισμα 3.5.5 Εστω X διακριτή τυχαία μεταβλητή με τιμές στο $\{m, m+1,...\}$. Εστω επίσης ότι

$$E\left[X|X>y\right] - h(y) = \alpha + \beta y \ , \quad \text{για} \quad \alpha < 1, \ \beta < 0 \ , \quad \text{και} \quad y = m,m+1,...$$

Τότε η κατανομή της X ορίζεται μονοσήμαντα και είναι η $(m-1)$-περικομένη Yule (ρ) όπου $\rho = -\frac{1}{\beta} > 0$.

Απόδειξη Αντικαθιστώντας τις σχέσεις $h(x) = x$ και $H(y) = \alpha+\beta y$ στην (3.5.31) προκύπτει ότι,

$$P(X\geq y) = \prod_{r=m}^{y-1} \frac{(\alpha-\beta+1)+\beta r}{(\alpha-\beta)+\beta r}$$

ή ισοδύναμα,

$$P(X\geq y) = \left(\frac{\alpha-\beta+1}{\beta}+m\right)_{(y-m)} \Big/ \left(\frac{\alpha-\beta}{\beta}+m\right)_{(y-m)}, \quad y = m, m+1, \ldots \qquad (3.5.32)$$

Στην ειδική περίπτωση που $\alpha-\beta+1 = \beta$ η σχέση (3.5.32) γίνεται,

$$P(X\geq y) = \frac{(m+1)_{(y-m)}}{(1+\rho+m)_{(y-m)}}$$

$$= (\rho+1+y)\,\frac{(m+1)_{(y-m)}}{(1+\rho+m)_{(y-m+1)}}$$

$$= \left[1 + \frac{1}{\rho}\,(y+1)\right]\,\rho\,\frac{(m+1)_{(y-m)}}{(1+\rho+m)_{(y-m+1)}}\,.$$

Αλλά αυτή είναι η ικανή και αναγκαία συνθήκη για να ακολουθεί η X την (m-1)-περικομένη Yule (ρ), $\rho = -\frac{1}{\beta} > 0$ (Xekalaki (1984b)).

ΚΕΦΑΛΑΙΟ 4

ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΔΙΔΙΔΙΑΣΤΑΤΩΝ ΚΑΤΑΝΟΜΩΝ PARETO & YULE

4.1 ΕΙΣΑΓΩΓΗ

Οπως φαίνεται και από τη βιβλιογραφία που παραταθέσαμε στα δυο προηγούμενα κεφάλαια σχετικά με τους χαρακτηρισμούς της μονοδιάστατης κατανομής Pareto υπάρχει πλήθος δημοσιευμένων εργασιών που καλύπτουν ένα πολύ μεγάλο μέρος του θέματος. Δεν συμβαίνει όμως το ίδιο για την περίπτωση των διδιάστατων κατανομών Pareto. Δεν μπορέσαμε να εντοπίσουμε κάποια σχετική εργασία ούτε καν για την ήδη γνωστή στη βιβλιογραφία διδιάστατη κατανομή Pareto τύπου Mardia I.

Στο κεφάλαιο αυτό αποδεικνύονται ιδιότητες των διδιάστατων κατανομών Pareto που ορίσθηκαν στο Κεφάλαιο 2 μερικές από τις οποίες αποδεικνύεται ότι είναι χαρακτηριστικές. Διατυπώνονται επίσης ιδιότητες που αφορούν διδιάστατες κατανομές οι οποίες συνδέονται με τις διδιάστατες κατανομές Pareto καθώς επίσης ιδιότητες για το διακριτό ανάλογο της διδιάστατης Pareto τύπου Mardia I δηλαδή τη διδιάστατη κατανομή Yule.

Αναλυτικότερα, η ενότητα 2 που ακολουθεί περιέχει ιδιότητες της διδιάστατης κατανομής Pareto τύπου Marshall- Olkin ενώ η ενότητα 3 εξετάζει την περίπτωση της διδιάστατης κατανομής Pareto τύπου Mardia I. Τέλος, η ενότητα 4 ασχολείται με χαρακτηριστικές ιδιότητες άλλων διδιάστατων κατανομών.

4.2 ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΔΙΔΙΑΣΤΑΤΗΣ ΚΑΤΑΝΟΜΗΣ PARETO ΤΥΠΟΥ MARSHAL-OLKIN.

Τα θεωρήματα της ενότητας αυτής αναφέρονται σε ιδιότητες της διδιάστατης κατανομής Pareto τύπου Marshall-Olkin που ορίσθηκε στην ενότητα 2.3.2. Το πρώτο από τα θεωρήματα αυτά δηλαδή το θεώρημα 4.2.1 θα μπορούσε να θεωρηθεί εναλλακτικά σαν διαδικασία ορισμού της διδιάστατης αυτής κατανομής.

<u>Θεώρημα 4.2.1</u> Εστω U_i $(i = 1,2,3)$ ανεξάρτητες τυχαίες μεταβλητές που κατανέμονται σύμφωνα με την κατανομή Pareto $(1,\alpha k_i)$ $(i = 1,2,3)$ και έστω,

$$X = \min (U_1, U_3)$$

$$Y = \min (U_2, U_3).$$

Τότε το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin με συνάρτηση επιβίωσης την

$$\overline{F}_{X,Y}(x,y) = x^{-\alpha k_1}\, y^{-\alpha k_2} \left[\max (x^{\alpha}, y^{\alpha})\right]^{-k_3}.$$

<u>Απόδειξη</u> Με βάση τον ορισμό της συνάρτησης επιβίωσης προκύπτει ότι,

$$\overline{F}_{X,Y}(x,y) = P\,(X>x,\ Y>y)$$

$$= P\left(\min (U_1, U_3) > x,\ \min (U_2, U_3) > y\right)$$

Από τις υποθέσεις του θεωρήματος όμως έχουμε ότι οι τυχαίες μεταβλητές U_i $(i = 1,2,3)$ είναι ανεξάρτητες και ακολουθούν την

κατανομή Pareto, ειδικότερα δε

$$\bar{F}_{U_i}(u) = u^{\alpha k_i}, \quad i = 1,2,3.$$

Ετσι έχουμε διαδοχικά ότι,

$$\bar{F}_{X,Y}(x,y) = P\left(U_1>x,\ U_2>y,\ U_3>\max(x,y)\right)$$

$$= P\ (U_1>x)\ P\ (U_2>y)\ P\left(U_3>\max(x,y)\right)$$

$$= \bar{F}_{U_1}(x)\ \bar{F}_{U_2}(y)\ \bar{F}_{U_3}\left(\max(x,y)\right)$$

$$= x^{-\alpha k_1}\ y^{-\alpha k_2}\left[\max\ (x^{\alpha},y^{\alpha})\right]^{-k_3}.$$

Η Dimaki (1977) αποδεικνύει ότι αν X_i , i=1,2,...,n είναι ανεξάρτητες τυχαίες μεταβλητές που κατανέμονται σύμφωνα με την κατανομή Pareto (θ,α) τότε η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $X_{1:n} = \min_i\left(X_{i:n}\right)$ i=1,2,...,n είναι η Pareto $(\theta,n\alpha)$. Στη συνέχεια θα αποδείξουμε ότι όταν οι τυχαίες μεταβλητές είναι εξαρτημένες και n=2 η ιδιότητα διατηρείται στην περίπτωση που το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin.

<u>Θεώρημα 4.2.2</u> Εστω X,Y τυχαίες μεταβλητές με από κοινού κατανομή τη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin με

συνάρτηση επιβίωσης,

$$\bar{F}_{X,Y}(x,y) = x^{-\alpha k}\, y^{-\alpha m} \left[\max\,(x^{\alpha}, y^{\alpha})\right]^{-r}$$

τότε το min(X,Y) ακλουθεί την κατανομή Pareto $(1,\beta)$ με $\beta=\alpha(k+m+r)$.

<u>Απόδειξη</u> Με βάση τις υποθέσεις του θεωρήματος έπεται ότι,

$$P\left(\min(X,Y) > x\right) = P\,(X > x,\ Y > y)$$

$$= x^{-\alpha k}\, x^{-\alpha m}\, x^{-\alpha r}$$

$$= x^{\alpha(k+m+r)}.$$

Δηλαδή, $\min(X,Y) \sim \text{Pareto}\left(1,\alpha(k+m+r)\right)$.

Είναι γνωστό ότι η μονοδιάστατη κατανομή Pareto χαρακτηρίζεται από τη σχέση

$$\bar{F}_X(st) = \bar{F}_X(s)\,\bar{F}_X(t), \qquad s,t \geq 1.$$

Μια προφανής μεταφορά της παραπάνω σχέσης στο χώρο των δύο διαστάσεων είναι η

$$\bar{F}_{X,Y}(s_1t_1, s_2t_2) = \bar{F}_{X,Y}(s_1,s_2)\,\bar{F}_{X,Y}(t_1,t_2). \qquad (4.2.1)$$

Το θεώρημα που ακολουθεί αποδεικνύει ότι η σχέση (4.2.1) οδηγεί σε

διπλή Pareto.

Θεώρημα 4.2.3 Η μοναδική λύση της συναρτησιακής εξίσωσης

$$\bar{F}_{X,Y}(s_1t_1, s_2t_2) = \bar{F}_{X,Y}(s_1, s_2)\ \bar{F}_{X,Y}(t_1, t_2)\ ,\ s_i, t_i \geq 1,\ i=1,2$$

είναι η διπλή Pareto.

Απόδειξη Εστω η συναρτησιακή εξίσωση

$$\bar{F}_{X,Y}(s_1t_1, s_2t_2) = \bar{F}_{X,Y}(s_1, s_2)\ \bar{F}_{X,Y}(t_1, t_2). \qquad (4.2.2)$$

Θέτοντας $s_2 = t_2 = 1$ προκύπτει ότι,

$$\bar{F}_{X,Y}(s_1t_1, 1) = \bar{F}_{X,Y}(s_1, 1)\ \bar{F}_{X,Y}(t_1, 1)$$

ή

$$\bar{F}_X(s_1t_1) = \bar{F}_X(s_1)\ \bar{F}_X(t_1), \quad \text{για όλα τα } s_1, t_1 \geq 1. \qquad (4.2.3)$$

Η σχέση (4.2.3) όπως γνωρίζουμε είναι ισοδύναμη με τη σχέση

$$\bar{F}_X(x) = x^{-\alpha_1}. \qquad (4.2.4)$$

Με όμοιο τρόπο καταλήγουμε στο ότι,

$$\bar{F}_Y(y) = y^{-\alpha_2}. \qquad (4.2.5)$$

Θέτοντας όπου $s_2 = t_1 = 1$ στη σχέση (4.2.2) προκύπτει ότι

$$\bar{F}_{X,Y}(s_1,t_2) = \bar{F}_{X,Y}(s_1,1)\ \bar{F}_{X,Y}(1,t_2)$$

ή

$$\bar{F}_{X,Y}(s_1,t_2) = \bar{F}_X(s_1)\ \bar{F}_Y(t_2).$$

Λόγω της (4.2.4) και της (4.2.5) η τελευταία αυτή εξίσωση γίνεται,

$$\bar{F}_{X,Y}(s_1,t_2) = x^{-\alpha_1}\ y^{-\alpha_2}.$$

Δηλαδή, (X,Y) ~ διπλη Pareto.

Στη συνέχεια θα αποδείξουμε ότι στην ειδική περίπτωση που $t_1=t_2=t$ η σχέση (4.2.1) είναι δυνατόν να οδηγήσει σε μονοσήμαντο ορισμό της κατανομής του τυχαίου διανύσματος (X,Y) ως διδιάστατης Pareto τύπου Marshall-Olkin.

Θεώρημα 4.2.4 Η μοναδική λύση της συναρτησιακής εξίσωσης

$$\bar{F}_{X,Y}(s_1t,s_2t) = \bar{F}_{X,Y}(s_1,s_2)\ \bar{F}_{X,Y}(t,t) \qquad (4.2.6)$$

είναι η

$$\bar{F}_{X,Y}(x,y) = x^{-k}\ y^{-m}\ \left[\max\ (x,y)\right]^{-r}\ ,\ \text{όπου } k,m,r\geq 0$$

υπό την προυπόθεση ότι οι περιθώριες κατανομές των X και Y είναι μονοδιάστατες Pareto.

Απόδειξη Θέτοντας όπου $(s_1, s_2) = (s, s)$ στη συναρτησιακή εξίσωση (4.2.6) προκύπτει ότι,

$$\bar{F}_{X,Y}(st, st) = \bar{F}_{X,Y}(s, s)\ \bar{F}_{X,Y}(t, t).$$

Επομένως,

$$\bar{F}_{X,Y}(s, s) = s^{-\theta} \qquad \text{για } \theta > 0.$$

Θέτοντας όπου $s_2 = 1$, η συναρτησιακή εξίσωση (4.2.6) γίνεται,

$$\bar{F}_{X,Y}(s_1 t, t) = \bar{F}_{X,Y}(s_1, 1)\ \bar{F}_{X,Y}(t, t)$$

$$= \bar{F}_X(s_1)\ t^{-\theta}.$$

Δηλαδή,

$$\bar{F}_{X,Y}(s_1 t, t) = t^{-\theta}\ \bar{F}_X(s_1). \qquad (4.2.7)$$

Με ανάλογο τρόπο προκύπτει ότι,

$$\bar{F}_{X,Y}(t, s_2 t) = t^{-\theta}\ \bar{F}_Y(s_2). \qquad (4.2.8)$$

Από τις υποθέσεις του θεωρήματος έχουμε ότι οι περιθώριες κατανομές των τυχαίων μεταβλητών X και Y είναι μονοδιάστατες Pareto. Κατά συνέπεια οι σχέσεις (4.2.7) και (4.2.8) γίνονται,

$$\overline{F}_{X,Y}(st,t) = t^{-\theta}\, s^{-\theta_1} \qquad \text{για} \quad \theta_1 > 0 \qquad (4.2.9)$$

και

$$\overline{F}_{X,Y}(t,st) = t^{-\theta}\, s^{-\theta_2} \qquad \text{για} \quad \theta_2 > 0 \qquad (4.2.10)$$

Επειδή δε $s>1$, $t>1$ οι σχέσεις (4.2.9) και (4.2.10) γίνονται αντίστοιχα,

$$\overline{F}_{X,Y}(x,y) = y^{-\theta} \left(\frac{x}{y}\right)^{-\theta_1} \qquad \text{αν} \quad x \geq y \qquad (4.2.11)$$

και

$$\overline{F}_{X,Y}(x,y) = x^{-\theta} \left(\frac{y}{x}\right)^{-\theta_2} \qquad \text{αν} \quad x \leq y \qquad (4.2.12)$$

με $\theta, \theta_1, \theta_2 > 0$.

Οι σχέσεις (4.2.11) και (4.2.12) ορίζουν από κοινού μια διδιάστατη κατανομή για όλα τα $x,y \geq 1$. Η $\overline{F}_{X,Y}(x,y)$ είναι φθίνουσα ως προς y όταν $\theta \geq \theta_1$, κατά συνέπεια όταν η διαφορά $\kappa = \theta - \theta_1$ είναι μη αρνητική. Ομοια, η $\overline{F}_{X,Y}(x,y)$ είναι φθίνουσα ως προς x όταν ότι $\theta \geq \theta_2$, επομένως όταν η $m = \theta - \theta_2$ είναι επίσης μη αρνητική. Εστω τώρα $r = \theta_1+\theta_2-\theta$. για να δείξουμε ότι $r \geq 0$ αρκεί να δείξουμε ότι $\theta_1+\theta_2 \geq \theta$. Για το σκοπό αυτό παρατηρώ ότι η διδιάστατη κατανομή έτσι όπως περιγράφεται από τις (4.2.11) και (4.2.12) οδηγεί σε μονοδιάστατη κατανομή της οποίας η συνάρτηση κατανομής είναι

$$F_X(x) = F_{X,Y}(x,x)$$

$$= 1 + \bar{F}_{X,Y}(x,x) - \bar{F}_X(x) - \bar{F}_Y(x)$$

όπου $\bar{F}_X(x) = x^{-\theta_1}$, $\bar{F}_Y(x) = x^{-\theta_2}$ και η $\bar{F}_{X,Y}(x,x)$ δίνεται από τις σχέσεις (4.2.11) και (4.2.12) με $y = x$. Επομένως,

$$f_X(x) = \frac{\partial}{\partial x} F_{X,Y}(x,x)$$

$$= \frac{\partial}{\partial x} \left[1 + x^{-\theta} - x^{-\theta_1} - x^{-\theta_2}\right]$$

$$= -\theta x^{-(\theta+1)} + \theta_1 x^{-(\theta_1+1)} + \theta_2 x^{-(\theta_2+1)}$$

$$\geq 0 \qquad \text{για όλα τα } x.$$

Εστω ότι $x \to 1$. Παρατηρούμε ότι $\theta_1+\theta_2 \geq \theta$ από το οποίο συνεπάγεται ότι το $r = \theta_1+\theta_2-\theta$ είναι μη αρνητικό.

Αντικαθιστώντας όπου $\theta = m+k+r$, $\theta_1 = k+r$ και $\theta_2 = m+r$ στις σχέσεις (4.2.11) και (4.2.12) προκύπτει ότι η μοναδική λύση της δοθείσας συναρτησιακής εξίσωσης είναι η

$$\bar{F}_{X,Y}(x,y) = y^{-m} x^{-k} \left[\max(x,y)\right]^{-r}, \quad x,y > 1$$
$$k,m,r \geq 0.$$

4.3 ΧΑΡΑΚΤΗΡΙΣΜΟΙ ΤΗΣ ΔΙΔΙΑΣΤΑΤΗΣ ΚΑΤΑΝΟΜΗΣ PARETO ΤΥΠΟΥ MARDIA I

Τα δυο θεωρήματα που περιέχονται στην ενότητα αυτή αποτελούν χαρακτηριστικές ιδιότητες της διδιάστατης κατανομής Pareto τύπου Mardia I.

Θεώρημα 4.3.1 Εστω (X_1,X_2) τυχαίο διάνυσμα του οποίου οι συνιστώσες ορίζονται στο $(0,+\infty)$. Τότε η σχέση

$$P(X_1>x,\ X_2>y) = \frac{1}{\rho\ (\rho+1)}\ (x+y+1)^2\ f_{X_1,X_2}(x,y)\ ,\quad x,y>0\ ,\ \rho>0$$

προσδιορίζει μονοσήμαντα την κατανομή του (X_1,X_2) ως διδιάστατη κατανομή Pareto τύπου Mardia I, με

$$f_{X_1,X_2}(x,y) = \rho\ (\rho+1)\ (x+y+1)^{-(\rho+2)},\qquad \rho>0.$$

Απόδειξη

Ευθύ: Εστω ότι το τυχαίο διάνυσμα (X_1,X_2) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I, τότε

$$P(X_1>x,\ X_2>y) = F_{X_1,X_2}(x,y) - F_{X_1}(x) - F_{X_2}(y) + 1$$

$$= (x+y+1)^{-\rho}$$

$$= \frac{1}{\rho\ (\rho+1)}\ (x+y+1)^2\ \left[\rho(\rho+1)\ (x+y+1)^{-(\rho+2)}\ \right]$$

$$= C\ (x+y+1)^2\ f_{X_1,X_2}(x,y).$$

Αντίστροφο: Εστω ότι ισχύει

$$P(X>x,\ Y>y) = C\ (x+y+1)^2\ f_{X,Y}(x,y)$$

ή, ισοδύναμα

$$\overline{F}_{X,Y}(x,y) = C\ (x+y+1)^2\ \frac{\partial^2\ \overline{F}_{X,Y}(x,y)}{\partial x\ \partial y}\ .$$

Γράφοντας την παραπάνω εξίσωση με την μορφή

$$\frac{\partial^2\ \overline{F}_{X,Y}(x,y)}{\partial x\ \partial y} - \frac{1}{C(x+y+1)^2}\ \overline{F}_{X,Y}(x,y) = 0,$$

παρατηρούμε ότι η μοναδική λύση της υπό τον περιορισμό $\overline{F}_{X,Y}(1,1) = 1$, δίνεται από την

$$\overline{F}_{X,Y}(x,y) = (x+y+1)^{-\rho}.$$

Θεώρημα 4.3.2 Εστω (X_1,X_2) τυχαίο διάνυσμα του οποίου οι συνιστώσες ορίζονται στο $(0,+\infty)$. Τότε η σχέση

$$\int_r^{\infty} f_{X_1,X_2}(x,m)\,dx = \int_m^{\infty} f_{X_1,X_2}(r,y)\,dy = \frac{1}{\rho+1}\,(1+r+m)\,f_{X_1,X_2}(r,m)$$

προσδιορίζει μονοσήμαντα την κατανομή του (X_1,X_2) ως διδιάστατη Pareto τύπου Mardia I με

$$f_{X_1,X_2}(x,y) = \rho\,(\rho+1)\,(x+y+1)^{-(\rho+2)}, \qquad \rho > 0.$$

Απόδειξη

Ευθύ: Εστω ότι το τυχαίο διάνυσμα (X_1,X_2) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I. Τότε,

$$\int_r^{\infty} f_{X_1,X_2}(x,m)\,dx = \int_r^{\infty} \rho\,(\rho+1)\,(1+x+m)^{-(\rho+2)}\,dx$$

$$= \rho\,(\rho+1)\int_r^{\infty}(1+x+m)^{-(\rho+2)}\,dx$$

$$= \rho\,(\rho+1)\,\frac{(1+r+m)^{-(\rho+1)}}{\rho+1}$$

$$= \frac{1}{\rho+1}\,(1+r+m)\,f_{X_1,X_2}(r,m)$$

$$= C\ (1+r+m)\ f_{X_1,X_2}(r,m)$$

όπου $C = \frac{1}{\rho+1}$.

Αντίστροφο: Εστω ότι

$$\int_r^{\infty} f_{X_1,X_2}(x,m)\ dx = C\ (1+r+m)\ f_{X_1,X_2}(r,m).$$

Παραγωγίζοντας ως προς r προκύπτει ότι,

$$-\ f_{X_1,X_2}(r,m) = C\ f_{X_1,X_2}(r,m) + C\ (1+r+m)\ \frac{\partial}{\partial r}\ f_{X_1,X_2}(r,m) \Rightarrow$$

$$C\ (1+r+m)\ \frac{\partial}{\partial r}\ f_{X_1,X_2}(r,m) + (C+1)\ f_{X_1,X_2}(r,m) = 0 \Rightarrow$$

$$\frac{\partial}{\partial r}\ f_{X_1,X_2}(r,m) + \frac{(C+1)}{C\ (1+r+m)}\ f_{X_1,X_2}(r,m) = 0 .$$

Λύση της διαφορικής αυτής εξίσωσης είναι η

$$f_{X_1,X_2}(r,m) = C_1 \exp\left\{ - \int \frac{(C+1)}{C\ (1+r+m)}\ dr \right\}$$

$$= C_1 \exp\left\{ - \frac{C+1}{C} \int \frac{1}{1+r+m}\ dr \right\}$$

$$= C_1 \exp\left\{ \ln\ (1+r+m)^{-\frac{C+1}{C}} \right\} .$$

Αρα τελικά προκύπτει ότι,

$$f_{X_1,X_2}(r,m) = C_1\ (1+r+m)^{-\frac{C+1}{C}} .$$

Αλλά,

$$\int_0^\infty \int_0^\infty f_{X_1,X_2}(r,m)\ dr\ dm = 1.$$

Κατά συνέπεια,

$$C_1 \int_0^\infty \int_0^\infty (1+r+m)^{-\frac{C+1}{C}}\ dr\ dm = 1$$

$$C\,C_1 \int_0^{\infty} (1+m)^{-1/C}\, dm = 1 \ .$$

Από την τελευταία αυτή σχέση προκύπτει ότι,

$$C_1 = \frac{1}{C}\,\frac{1-C}{C} \ .$$

Αρα καταλήγουμε στο ότι,

$$f_{X_1,X_2}(r,m) = \frac{1-C}{C}\,\frac{1}{C}\,(1+r+m)^{-\frac{C+1}{C}} \ .$$

Δηλαδή, το τυχαίο διάνυσμα (X_1,X_2) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho = \frac{1}{C} - 1$.

4.4 ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΔΙΔΙΑΣΤΑΤΗΣ ΚΑΤΑΝΟΜΗΣ YULE

Στην ενότητα αυτή περιέχεται μια χαρακτηριστική ιδιότητα της διδιάστατης κατανομής Yule.

Θεώρημα 4.4.1 Εστω το τυχαίο διάνυσμα (X_1,X_2) το οποίο έχει μη αρνητικές συνιστώσες οι οποίες παίρνουν ακέραιες τιμές. Τότε η κατανομή του τυχαίου διανύσματος (X_1,X_2) είναι η διδιάστατη Yule (ρ), $\rho>0$ με από κοινού συνάρτηση πιθανότητας

$$p_{r,m} = \frac{\rho_{(2)}\,(r+m)!}{(\rho+1)_{(r+m+2)}} \qquad (4.4.1)$$

αν και μόνον αν

$$P(X_1>r,\ X_2>m) = \frac{1}{\rho\ (\rho+1)}\ (r+m+1)\ (r+m+2)\ p_{r,m}$$

Απόδειξη

Ευθύ: Εστω ότι το τυχαίο διάνυσμα (X_1,X_2) ακολουθεί τη διδιάστατη κατανομή Yule. Τότε,

$$P(X_1>r,\ X_2>m) = \sum_{x=r+1}^{\infty}\ \sum_{y=m+1}^{\infty} \frac{\rho_{(2)}\,(x+y)!}{(\rho+1)_{(x+y+2)}}$$

$$= \rho_{(2)} \sum_{x=0}^{\infty}\ \sum_{y=0}^{\infty} \frac{1_{(x+y+r+m+2)}}{(\rho+1)_{(r+m+4)}\,(\rho+r+m+5)_{(x+y)}}$$

$$= \frac{\rho_{(2)}\ 1_{(r+m+2)}}{(\rho+1)_{(r+m+4)}} \sum_{x=0}^{\infty}\ \sum_{y=0}^{\infty} \frac{(r+m+3)_{(x+y)}}{(\rho+r+m+5)_{(x+y)}}$$

$$= \frac{\rho_{(2)}\ 1_{(r+m+2)}}{(\rho+1)_{(r+m+4)}}\ \frac{(\rho+r+m+3)_{(2)}}{\rho_{(2)}}\ \sum_{x=0}^{\infty}\ \sum_{y=0}^{\infty} \frac{\rho_{(2)}}{(\rho+r+m+3)_{(2)}}\ \frac{(r+m+3)_{(x+y)}}{(\rho+r+m+5)_{(x+y)}}$$

Δηλαδή,

$$P(X_1>r,\ X_2>m) = \frac{\rho_{(2)}\ 1_{(r+m)}\ (1+r+m)_{(2)}\ (\rho+r+m+3)_{(2)}}{(\rho+1)_{(r+m+4)}\ \rho_{(2)}}\ .$$

Λόγω δε της (4.4.1) τελικά προκύπτει ότι,

$$P(X_1>r,\ X_2>m) = \frac{1}{\rho_{(2)}}\ (1+r+m)_{(2)}\ p_{r,m}$$

$$= C\ (r+m+1)(r+m+2)\ p_{r,m}$$

Αντίστροφο: Εστω ότι,

$$P(X>r,\ Y>m) = C\ (r+m+1)(r+m+2)\ p_{r,m}$$

ή

$$\sum_{x=r+1}^{\infty}\ \sum_{y=m+1}^{\infty} p_{x,y} = C\ (r+m+1)(r+m+2)\ p_{r,m}$$

Από τη σχέση αυτή για $r,m = 0,1,\ldots$ προκύπτουν διαδοχικά οι παρακάτω σχέσεις,=

$$C(r+m+1)(r+m+2)\ p_{r,m} - C(r+m+3)(r+m+2)p_{r+1,m} = \sum_{x=m+1}^{\infty} p_{r+1,x}$$

$$C(r+m+2)(r+m+1)\ p_{r,m} - C(r+m+3)(r+m+2)p_{r,m+1} = \sum_{x=r+1}^{\infty} p_{x,m+1}$$

$$C(r+m+2)(r+m+1)\ p_{r,m} - C(r+m+4)(r+m+3)p_{r+1,m+1} =$$

$$= \sum_{x=m+1}^{\infty} p_{r+1,x} + \sum_{x=r+1}^{\infty} p_{x,m+1} - p_{r+1,m+1}.$$

Από τις τρεις αυτές τελευταίες σχέσεις προκύπτει ότι,

$$(r+m+4)(r+m+3)p_{r+1,m+1} + (r+m+2)(r+m+1)p_{r,m} - (r+m+3)(r+m+2)p_{r+1,m}$$

$$- (r+m+3)(r+m+2)p_{r,m+1} = \frac{1}{C}\ p_{r+1,m+1}$$

ή

$$\left[(r+m+4)(r+m+3) - \frac{1}{C}\right] p_{r+1,m+1} + (r+m+2)(r+m+1)p_{r,m} -$$

$$- (r+m+3)(r+m+2)p_{r+1,m} - (r+m+3)(r+m+2)p_{r,m+1} = 0$$

Λύση της εξίσωσης αυτής διαφορών είναι η $p_{r,m}$ έτσι όπως αυτή δίνεται από την (4.4.1) . Οπου $C = 1/\rho(\rho+1)$.

Κ Ε Φ Α Λ Α Ι Ο 5

ΟΙ ΚΑΤΑΝΟΜΕΣ

PARETO ΚΑΙ YULE

ΣΤΟ ΠΛΑΙΣΙΟ ΤΗΣ ΘΕΩΡΙΑΣ ΑΞΙΟΠΙΣΤΙΑΣ

5.1 ΕΙΣΑΓΩΓΗ

Η θεωρία αξιοπιστίας διατυπώνεται ως η θεωρία πρόβλεψης, εκτίμησης ή βελτιστοποίησης της πιθανότητας επιβίωσης *(probability of survival)*, της μέσης ζωής *(mean life)*, ή πιο γενικά της κατανομής ζωής *(life distribution)* των συνιστωσών ενός συστήματος ή και συστημάτων. Η ανάπτυξη της θεωρίας αξιοπιστίας υπήρξε ραγδαία και τούτο γιατί η γνώση της συμπεριφοράς αποτυχίας μιας συνιστώσας ενός συστήματος *(failure behaviour of a component)* μπορεί να οδηγήσει σε μικρότερο κόστος παραγωγής ή συντήρησης και σε μερικές περιπτώσεις στη προστασία της ανθρώπινης ζωής. Επίσης διαπιστώθηκε η συμβολή της θεωρίας σε τομείς άλλους από εκείνους οι οποίοι συνδέονται με μηχανές και τα εξαρτήματα αυτών και οι οποίοι έδωσαν το έναυσμα για το ξεκίνημα της θεωρίας. Σήμερα δείχνουν αυξημένο ενδιαφέρον για τη θεωρία αυτή μαθηματικοί, οικονομολόγοι καθώς και εκείνοι που ασχολούνται με θέματα μελέτης του περιβάλλοντος. Με απλή παρατήρηση καταλήγει κανείς στο συμπέρασμα ότι οι μηχανικές αποτυχίες συμβαίνουν τυχαία. Αυτή η παρατήρηση οδηγεί στο συμπέρασμα ότι μελέτες που αφορούν την αξιοπιστία πρέπει να γίνονται με βάση τη θεωρία πιθανοτήτων και τη στατιστική θεωρία. Στο πλαίσιο της θεωρίας αξιοπιστίας η προσέγγιση αυτή είναι γενικά αποδεκτή. Ετσι καθοριστικής σημασίας είναι η αλληλεπίδραση που υπάρχει μεταξύ θεωρίας αξιοπιστίας και στατιστικής. Αξιοσημείωτο επίσης είναι ότι ενώ η θεωρία αξιοπιστίας θεωρείται σαν εφαρμογή των πιθανοτήτων και της στατιστικής σε μια ειδική κατηγορία προβλημάτων, η θεωρία αξιοπιστίας συνέβαλε στο να αναπτυχθούν καινούριες περιοχές της

στατιστικής. Σαν τέτοιες μπορούν να θεωρηθούν ο μη παραμετρικός χαρακτηρισμός συναρτήσεων κατανομών στο πλάισιο της θεωρίας ανανέωσης ή της θεωρίας ταξινόμησης και επιλογής κατανομών.

Οπως αντιλαμβανόμαστε η εμβέλεια της θεωρίας αξιοπιστίας είναι τεράστια. Η παρούσα διατριβή αντιμετωπίζει το θέμα από συγκεκριμένη οπτική γωνία αφιερώνοντας σε αυτό ένα μόνο κεφάλαιο.

Είναι γνωστό ότι κατανομές όπως η εκθετική, η Weibull, η Gamma και η λογαριθμοκανονική έχουν χαρακτηρισθεί, άλλες περισσότερο και άλλες λιγώτερο, σαν κατανομές χρόνου ζωής (life time distributions). Στο κεφάλαιο αυτό εξετάζονται αφενός μεν οι κατανομές Pareto αφετέρου δε οι κατανομές Yule στο πλαίσιο της θεωρίας αξιοπιστίας. Ετσι αποδεικνύεται ότι οι κατανομές που προαναφέραμε ικανοποιούν ένα σημαντικό αριθμό ιδιοτήτων και κατά συνέπεια θα μπορούσαν να χρησιμοποιηθούν με επιτυχία για την αντιμετώπιση προβλημάτων στο πλαίσιο της θεωρίας αυτής.

Στην ενότητα που ακολουθεί γίνεται μια μικρή εισαγωγή σε βασικές έννοιες της θεωρίας αξιοπιστίας. Πέρα από την ανάλυση της ήδη χρησιμοποιούμενης ορολογίας γίνεται και επέκταση υπαρχουσών εννοιών. Ορισμένα συμπεράσματα μεταφέρονται επίσης και στο χώρο των δύο διαστάσεων.

Στην τρίτη ενότητα αποδεικνύονται ιδιότητες χαρακτηριστικές ή μη των μονοδιάστατων κατανομών Pareto και Yule οι οποίες συνδέονται με έννοιες της θεωρίας αξιοπιστίας. Ορισμένες από τις ιδιότητες αυτές αποδεικνύεται ότι είναι ισοδύναμες με άλλες ήδη γνωστές χαρακτηριστικές ιδιότητες των κατανομών αυτών.

Τέλος, στην τέταρτη ενότητα αποδεικνύονται χαρακτηριστικές ή μη ιδιότητες τόσο για διμεταβλητές μορφές της κατανομής Pareto (συνεχής περίπτωση) όσο και για διμεταβλητές μορφές της κατανομής Yule (διακριτή περίπτωση).

Θα τελειώσουμε την εισαγωγική αυτή στο θέμα ενότητα με μια σύντομη αναφορά στην υπάρχουσα βιβλιογραφία. Αξίζει να αναφέρουμε ότι η μόνη εργασία στην οποία εντοπίσαμε σαφή αναφορά στις μονοδιάστατες κατανομές Pareto και Yule ως κατανομές χρόνου ζωής είναι εκείνη της Xekalaki (1983b). Η υπόλοιπη από τη βιβλιογραφία που παραθέτουμε χρησίμευσε στο να αποκτήσουμε στοιχειώδες υπόβαθρο σε θέματα της θεωρίας αξιοπιστίας τα οποία αντιμετωπίζονται με πιθανοθεωρητική και στατιστική μεθοδολογία και στα οποία οι κατανομές Pareto και Yule μπορούν να παίξουν σημαντικό ρόλο. Ετσι αναφέρουμε το ειδικό στο θέμα βιβλίο των Mann, Schafer and Singpurwalla (1974), καθώς και εκείνο των Kalbfleisch and Prentice (1980). Από τις δημοσιευμένες εργασίες θα ξεχωρίσουμε ως πιο χαρακτηριστικές τις παρακάτω τις οποίες παραθέτουμε ακολουθώντας τη χρονική στιγμή της δημοσίευσής τους: Epstein and Sobel (1953, 1954, 1961), Reinhardt (1968), Bryson and Siddiqui (1969), Harris (1970), Basu (1971), Brindley and Thompson (1972), Laurent (1974, 1975), Marshall (1975), Block (1977a, 1977b), Friday and Patil (1977), Shimi and Tsokos (1977), Muth (1977), Morrison (1978), Esary and Marshall (1979), Block and Savits (1980a, 1980b, 1981a, 1981b), Gupta (1980), Thompson (1981), Roy and Mukherjee (1986).

5.2 ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΑΞΙΟΠΙΣΤΙΑΣ

Η πιθανή συμπεριφορά αποτυχίας μιας συνιστώσας ενός συστήματος προσδιορίζεται γενικά από μια συνάρτηση, έστω f(t), η οποία ονομάζεται **πυκνότητα αποτυχίας** *(failure density)*. Στην πραγματικότητα η f(t) είναι η συνάρτηση πυκνότητας πιθανότητας η οποία ορίζεται στον ημιάξονα των θετικών πραγματικών αριθμών (αποτυχία δεν μπορεί να συμβεί πριν από την έναρξη της χρήσης), και η οποία καθορίζει την πιθανότητα στιγμιαίας αποτυχίας τη χρονική στιγμή t. Το **μοντέλο αποτυχίας** *(failure model)* περιλαμβάνει την εξειδικευμένη συναρτησιακή μορφή της f(t) καθώς και τις τιμές των παραμέτρων της. Αντίστοιχη προς την πυκνότητα f(t) είναι η συνάρτηση κατανομής

$$F(t) = \int_0^t f(x)\, dx$$

η οποία ορίζεται ως η αθροιστική πιθανότητα αποτυχίας μέχρι τη χρονική στιγμή t. Εμείς όμως, ενδιαφερόμαστε κυρίως για την πιθανότητα η συνιστώσα του συστήματος να μην αποτύχει τη χρονική στιγμή t. Η συνάρτηση η οποία καθορίζει αυτή την πιθανότητα δίνεται από την

$$\overline{F}(t) = 1 - F(t),$$

και είναι γνωστή ως **συνάρτηση αξιοπιστίας** *(reliability function)*.

Επιπλέον, για δοσμένη συνάρτηση κατανομής $F(t) = P(X \leq t)$ ορίζουμε την **κατανομή υπολειπόμενης ζωής κατά τη χρονική στιγμή t**

(residual life distribution at time t), έστω $F^x(t)$, με βάση τον παρακάτω ορισμό,

$$F^x(t) = P(x < X \leq x+t | X > x) \quad , \; t \geq 0. \tag{5.2.1}$$

Η συνάρτηση υπολειπόμενης ζωής ορίζεται για εκείνα τα x για τα οποία $P(X>x) > 0$.

Άμεση συνέπεια του ορισμού που περιγράφεται από τη σχέση (5.2.1) είναι και η παρακάτω ισοδύναμη έκφραση για τη συνάρτηση υπολειπόμενης ζωής κατά τη χρονική στιγμή t,

$$F^x(t) = 1 - \frac{\bar{F}(x+t)}{\bar{F}(x)} \quad , \qquad t \geq 0. \tag{5.2.2}$$

Ο λόγος

$$r^x(t) = \frac{P(X>x+t)}{P(X>x)} = \frac{\bar{F}(x+t)}{\bar{F}(x)} \quad , \; t \geq 0 \tag{5.2.3}$$

εκφράζει την πιθανότητα ένα σύστημα να επιζήσει για μια διάρκεια t δεδομένου ότι επέζησε μέχρι τη χρονική στιγμή x, είναι δε γνωστός στη διεθνή βιβλιογραφία ως **συνάρτηση βαθμού αποτυχίας** *(failure rate function)* . Ενας μεγάλος αριθμός χαρακτηρισμών της εκθετικής και της γεωμετρικής κατανομής στηρίζεται σε αυτόν. Στην παρούσα διατριβή αυτή προτείνεται η εισαγωγή μιας νέας έννοιας, της έννοιας της **συνάρτησης πολλαπλασιαστικού βαθμού αποτυχίας** *(multiplicative failure rate function)* η οποία ορίζεται από τον λόγο

$$r_x(t) = \frac{P(X>x\ t)}{P(X>x)} \quad , \quad t \geq 1 \qquad (5.2.4)$$

ενώ προτείνεται ο όρος συνάρτηση βαθμού προσθετικής αποτυχίας *(additive failure rate function)* για τον λόγο

$$r^x(t) = \frac{P(X>x+t)}{P(X>x)} \quad , \quad t \geq 0 \ . \qquad (5.2.5)$$

Μια έννοια καθοριστικής σημασίας για τη μελέτη φαινομένων στο πλαίσιο της θεωρίας αξιοπιστίας είναι εκείνη της συνάρτησης βαθμού κινδύνου *(hazard rate function)*. συμβολίζεται με h(t) και εκφράζει την πιθανότητα αποτυχίας του συστήματος στο χρόνο t δοθέντος ότι αυτό επέζησε μέχρι τη χρονική στιγμή t, δηλαδή

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P(\text{σύστημα ηλικίας } t \text{ θα αποτύχει στο διάστημα } (t, t+\Delta t) \mid \text{επέζησε μέχρι τη χρονική στιγμή } t)}{\Delta t} \right]$$

$$= \lim_{\Delta t \to 0} \left[\frac{P\left(X \leq t + \Delta t \mid X > t \right)}{\Delta t} \right]$$

Ορίζεται επομένως, για μεν την περίπτωση συνεχούς κατανομής από τη σχέση

$$h(t) = \frac{f_X(t)}{\bar{F}_X(t)} \quad , \qquad t \geq 0 \qquad (5.2.6)$$

για δε την περίπτωση διακριτής κατανομής από την

$$h(t) = \frac{P(X=t)}{P(X\geq t)}, \quad t = 0,1,2,... \tag{5.2.7}$$

Η μέση υπολειπόμενη ζωή κατά τη χρονική στιγμή t *(mean residual life at time* t*)* συμβολίζεται με $\mu^x(t)$ και ορίζεται για μεν την περίπτωση συνεχούς κατανομής ως

$$\mu^x(t) = \frac{1}{\overline{F}_X(t)} \int_t^\infty \overline{F}_X(x)\, dx \tag{5.2.8}$$

$$= \frac{1}{\overline{F}_X(t)} \int_t^\infty x\, dF_X(x) - t$$

$$= E\left(X-t \mid X>t\right), \quad t \geq 0,$$

για δε την περίπτωση διακριτής κατανομής,

$$\mu^x(t) = \frac{1}{\overline{F}_X(t)} \sum_{x=t}^{\infty} \overline{F}_X(x). \tag{5.2.9}$$

Εστω ότι θέλουμε να εκφράσουμε τη συνάρτηση αξιοπιστίας $\bar{F}_X(t)$ συναρτήσει της $\mu^x(t)$. Θα αποδείξουμε ότι ισχύει το παρακάτω θεώρημα.

Θεώρημα 5.2.1 Εστω X τυχαία μεταβλητή η οποία έχει συνάρτηση αξιοπιστίας $\bar{F}_X(t)$. Η κατανομή της X ορίζεται μονοσήμαντα

είτε α) από τη συνάρτηση βαθμού κινδύνου h(t).

είτε β) από τη μέση υπολειπόμενη ζωή $\mu^x(t)$.

Απόδειξη

Α' ΣΥΝΕΧΗΣ ΠΕΡΙΠΤΩΣΗ.

Από τη σχέση (5.2.8) προκύπτει ότι,

$$\mu^x(t) = \frac{1}{\bar{F}_X(t)} \left[\mu - \int_0^t \bar{F}_X(x)\, dx \right] \qquad (5.2.10)$$

όπου μ είναι η αναμενόμενη τιμή της τυχαίας μεταβλητής X η οποία υποτίθεται ότι είναι πεπερασμένη. Προφανώς, $\mu^x(0) = \mu$. Πολλαπλασιάζοντας και τα δύο μέλη της (5.2.10) επί $\bar{F}_X(t)$ και παραγωγίζοντας ως προς t προκύπτει ότι,

$$\mu^{x\prime}(t)\, \bar{F}_X(t) + \mu^x(t)\, \bar{F}'_X(t) = -\bar{F}_X(t) \qquad (5.2.11)$$

δηλαδή,

$$\frac{1 + \mu^{x}{}'(t)}{\mu^{x}(t)} = h(t) \qquad (5.2.12)$$

όπου

$$h(t) = - \frac{\bar{F}_X'(t)}{\bar{F}_X(t)}$$

είναι η συνάρτηση κινδύνου. Επειδή τόσο η $\mu^{x}(t)$ όσο και η $h(t)$ ειναι μη αρνητικές για όλα τα t από τη σχέση (5.2.12) προκύπτει ότι,

$\mu^{x}{}'(t) \geq -1$.

Παρατηρούμε λοιπόν ότι μπορούμε να εκφράσουμε τη συνάρτηση αξιοπιστίας $\bar{F}_X(t)$ συναρτήσει της $h(t)$, δηλαδή,

$$\bar{F}_X(t) = \exp\left[-\int_0^t h(x)\, dx\right]. \qquad (5.2.13)$$

Αν αντικαταστήσουμε στη σχέση (5.2.13) την (5.2.12) προκύπτει ότι,

$$\bar{F}_X(t) = \exp\left[-\int_0^t \frac{1 + \mu^{x}{}'(x)}{\mu^{x}(x)}\, dx\right].$$

Επιπλέον,

$$\int_0^t \frac{1 + \mu^{x}{}'(x)}{\mu^{x}(x)}\, dx = \ln \mu^{x}(x) - \ln \mu^{x}(0) + \int_0^t \frac{dx}{\mu^{x}(x)}$$

επομένως, τελικά προκύπτει ότι,

$$\bar{F}_X(t) = \frac{\mu^x(0)}{\mu^x(x)} \exp\left[- \int_0^t \frac{dx}{\mu^x(x)} \right] . \qquad (5.2.14)$$

Β' ΔΙΑΚΡΙΤΗ ΠΕΡΙΠΤΩΣΗ.

Από τη σχέση (5.2.7) προκύπτει ότι,

$$P(X \geq t) = \frac{P(X=t)}{h(t)} \qquad (5.2.15)$$

Εξειδικεύοντας τη σχέση (5.2.15) για t=r και t=r+1 και αφαιρώντας τις εξισώσεις που προκύπτουν έχουμε ότι,

$$P(X=r) = \frac{P(X=r)}{h(r)} - \frac{P(X=r+1)}{h(r+1)}$$

δηλαδή,

$$P(X=r+1) - \frac{\left[1-h(r)\right] h(r+1)}{h(r)} P(X=r) = 0.$$

Η μοναδική λύση της εξίσωσης αυτής διαφορών δίνεται από τη σχέση

$$P(X=r) = P(X=0) \prod_{i=0}^{r-1} \frac{\left[1-h(i)\right] h(i+1)}{h(i)} , \quad i = 1,2,\ldots . \qquad (5.2.16)$$

Από τη σχέση (5.2.9) προκύπτει ότι,

$$\mu^{x}(t)\ \overline{F}_{X}(t) = \mu - \sum_{x=0}^{t-1} \overline{F}_{X}(x) \qquad (5.2.17)$$

όπου μ είναι η αναμενόμενη τιμή της τυχαίας μεταβλητής X η οποία υποτίθεται ότι είναι πεπερασμένη. Προφανώς, $\mu^{x}(0) = \mu$. Εξειδικεύοντας τη σχέση (5.2.17) για t=r και t=r+1 αφαιρώντας τις εξισώσεις που προκύπτουν έχουμε ότι,

$$\mu^{x}(r+1)\ \overline{F}_{X}(r+1) - \mu^{x}(r)\ \overline{F}_{X}(r) = -\ \overline{F}_{X}(r)$$

δηλαδή,

$$\overline{F}_{X}(r+1) - \frac{\mu^{x}(r)-1}{\mu^{x}(r+1)}\ \overline{F}_{X}(r) = 0$$

Η μοναδική λύση της παραπάνω εξίσωσης διαφορών υπό τον περιορισμό P(X>0) = 1 δίνεται από τη σχέση,

$$P(X>r) = \prod_{i=0}^{r} \frac{\mu^{x}(r)-1}{\mu^{x}(r+1)} \quad , \quad r = 1,2,\ldots . \qquad (5.2.18)$$

Από τις σχέσεις (5.2.13), (5.2.17) και (5.2.14),(5.2.18) προκύπτει ότι η συνάρτηση αξιοπιστίας $\overline{F}_{X}(t)$ ορίζεται μονοσήμαντα τόσο από την συνάρτηση κινδύνου h(t) όσο και από τη συνάρτηση μέσης υπολειπόμενης ζωής $\mu^{x}(t)$. Επειδή δε η συνάρτηση αξιοπιστίας $\overline{F}_{X}(t)$

ορίζει μονοσήμαντα το νόμο πιθανότητας τον οποίο ακολουθεί η τυχαία μεταβλητή X, συνεπάγεται ότι η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X προσδιορίζεται μονοσήμαντα είτε από τη συνάρτηση κινδύνου h(t) είτε από τη συνάρτηση μέσης υπολειπόμενης ζωής $\mu^{x}(t)$.

Θα ολοκληρώσουμε την ενότητα μεταφέροντας στο χώρο των δυο διαστάσεων μερικές από τις έννοιες που ήδη ορίσαμε. Εστω το διδιάστατο μοντέλο αποτυχίας το οποίο περιλαμβάνει τη διδιάστατη πυκνότητα αποτυχίας $f(t_1,t_2)$. Η διδιάστατη συνάρτηση αξιοπιστίας ορίζεται ως

$$\bar{F}_{X,Y}(x,y) = F_{X,Y}(x,y) - F_X(x) - F_Y(y) + 1 \ . \qquad (5.2.19)$$

<u>Ορισμός 5.2.1</u> Εστω το τυχαίο διάστημα (X,Y) . Θα ονομάζουμε **διδιάστατη συνάρτηση βαθμού πολλαπλασιαστικής αποτυχίας** την

$$r_{x,y}(t,t) = \frac{\bar{F}_{X,Y}(xt,yt)}{\bar{F}_X(x)} \ , \ t \geq 1 \ . \qquad (5.2.20)$$

<u>Ορισμός 5.2.2</u> Εστω το τυχαίο διάστημα (X,Y) . Θα ονομάζουμε **διδιάστατη συνάρτηση βαθμού κινδύνου** για μεν την περίπτωση συνεχούς κατανομής την

$$h(x,y) = \frac{f_{X,Y}(x,y)}{\bar{F}_{X,Y}(x,y)} \qquad (5.2.21)$$

για δε την περίπτωση διακριτής κατανομής την

$$h(x,y) = \frac{P(X=x,Y=y)}{P(X>x,Y>y)} \ . \qquad (5.2.22)$$

5.3 ΟΙ ΜΟΝΟΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ PARETO ΚΑΙ YULE ΣΤΟ ΠΛΑΙΣΙΟ ΤΗΣ ΘΕΩΡΙΑΣ ΑΞΙΟΠΙΣΤΙΑΣ.

Είναι γνωστό ότι η εκθετική κατανομή χαρακτηρίζεται από τη σχέση

$$P\left(X \geq x+t \mid X \geq t\right) = P\left(X \geq x\right) \qquad (5.3.1)$$

η οποία είναι γνωστή στη βιβλιογραφία ως **ιδιότητα έλλειψης μνήμης** *(lack of memory property)*. Στη μελέτη αυτή θα αναφερόμαστε στην ιδιότητα που περιγράφεται από τη σχέση (5.3.1) ως **ιδιότητα έλλειψης προσθετικής** μνήμης και τούτο γιατί θα εισάγουμε και μια άλλη άμεσα συνδεόμενη έννοια αυτήν της πολλαπλασιαστικής μνήμης. Συγκεκριμένα θα αποδείξουμε ότι η μονοδιάστατη κατανομή Pareto χαρακτηρίζεται από την **ιδιότητα έλλειψης πολλαπλασιαστικής μνήμης** έτσι όπως αυτή περιγράφεται από τη σχέση (5.3.2).

<u>Θεώρημα 5.3.1</u> Εστω X συνεχής τυχαία μεταβλητή. Τότε η σχέση

$$P\left(X \geq x\ t \mid X \geq t\right) = P\left(X \geq x\right)\ ,\ t \geq 1 \qquad (5.3.2)$$

ορίζει μονοσήμαντα την κατανομή της τυχαίας μεταβλητής X ως Pareto $(1,\alpha)$.

<u>Απόδειξη</u>

Ευθύ: Εστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $(1,\alpha)$, $\alpha>0$. Τότε,

$$P\left(X \geq x\ t \mid X \geq t\right) = \frac{P(X \geq x\ t,\ X \geq t)}{P(X \geq t)} \qquad x, t \geq 1.$$

Επομένως,

$$P\left(X \geq x\ t \mid X \geq t\right) = \frac{P(X \geq x\ t)}{P(X \geq t)}$$

$$= x^{-\alpha}$$

$$= P(X \geq x).$$

Αντίστροφο: Εστω ότι η σχέση (5.3.2) ισχύει. Τότε,

$$\frac{P(X \geq x\ t,\ X \geq t)}{P(X \geq t)} = P(X \geq x).$$

Δηλαδή,

$$\frac{P(X \geq x\ t)}{P(X \geq t)} = P(X \geq x).$$

Κατά συνέπεια,

$$\overline{F}_X(x\ t) = \overline{F}_X(x)\ \overline{F}_X(t)$$

Η γενική λύση της συναρτησιακής αυτής εξίσωσης είναι η, $\overline{F}_X(x) = x^c$.

Επειδή δε $\overline{F}_X(\infty) = 0 \Rightarrow c = -\alpha$, όπου $\alpha > 0$.

Το θεώρημα που ακολουθεί αποδεικνύει ότι η μορφή της συνάρτησης πολλαπλασιαστικού βαθμού αποτυχίας ορίζει μονοσήμαντα την κατανομή Pareto.

Θεώρημα 5.3.2 Εστω X συνεχής τυχαία μεταβλητή η οποία παίρνει τιμές μεγαλύτερες ή ίσες από τη μονάδα. Τότε η X ακολουθεί τη κατανομή Pareto $(1,\alpha)$, $\alpha>0$ αν και μόνον αν η συνάρτηση πολλαπλασιαστικού βαθμού αποτυχίας έτσι όπως αυτή ορίζεται από τη σχέση (5.2.4) είναι σταθερή ως προς x για κάθε $t \geq 1$.

Απόδειξη

Ευθύ: Εστω X συνεχής τυχαία μεταβλητή η οποία ακολουθεί την κατανομή Pareto $(1,\alpha)$, $\alpha > 0$. Τότε,

$$\frac{\bar{F}_X(x\ t)}{\bar{F}_X(x)} = t^{-\alpha}$$

$$= g(t) , \quad t \geq 1.$$

Αντίστροφο: Εστω X συνεχής τυχαία μεταβλητή της οποίας η συνάρτηση πολλαπλασιαστικού βαθμού αποτυχίας είναι σταθερή ως προς x για κάθε $t \geq 1$. Τότε,

$$\frac{\bar{F}_X(x\ t)}{\bar{F}_X(x)} = g(t), \quad t \geq 1.$$

Δηλαδή,

$$\bar{F}_X(x\ t) = g(t)\ \bar{F}_X(x).$$

Η γενική λύση της συναρτησιακής αυτής εξίσωσης δίνεται από τις σχέσεις:

$$\bar{F}_X(t) = b\ t^c \quad \text{και} \quad g(t) = t^c.$$

Επειδή δε αφενός μεν $\bar{F}_X(1) = 1$ αφετέρου $\bar{F}_X(\infty) = 0$ προκύπτει ότι $b = 1$ και $c = -\alpha$ όπου $\alpha > 0$.

Συνδυάζοντας τα συμπεράσματα των θεωρημάτων 5.3.1 και 5.3.2 προκύπτει το θεώρημα που ακολουθεί.

Θεώρημα 5.3.3 Εστω X συνεχής τυχαία μεταβλητή η οποία παίρνει τιμές μεγαλύτερες ή ίσες από τη μονάδα. Τότε η πρόταση: η συνάρτηση πολλαπλασιαστικού βαθμού αποτυχίας είναι σταθερή ως προς x για κάθε t, είναι ισοδύναμη με την πρόταση : η τυχαία μεταβλητή X έχει την ιδιότητα της έλλειψης πολλαπλασιαστικής μνήμης.

Απόδειξη Με βάση τους ορισμούς που έχουν προηγηθεί προκύπτει ότι τελικά θέλουμε να αποδείξουμε την ισοδυναμία των προτάσεων:

$$\alpha)\ P\left(X \geq x\ t \middle| X \geq t\right) = P\left(X \geq x\right),\ t \geq 1 \qquad (5.3.3)$$

και

$$\beta)\ \frac{\bar{F}_X(x\ t)}{\bar{F}_X(x)} = g(t), \quad t \geq 1. \qquad (5.3.4)$$

Η απόδειξη είναι προφανής, αφού με βάση αφενός μεν το θεώρημα 5.3.1 ισχύει ότι,

$$P\left(X \geq x\ t \mid X \geq t\right) = P\left(X \geq x\right)\ ,\ t \geq 1 \quad \Leftrightarrow \quad X \sim \text{Pareto}\ (1,\alpha)$$

αφέτερου δε με βάση το θεώρημα 5.3.2 ισχύει ότι,

$$\frac{\overline{F}_X(x\ t)}{\overline{F}_X(x)} = g(t), \quad t \geq 1 \quad \Leftrightarrow \quad X \sim \text{Pareto}\ (1,\alpha).$$

Το θεώρημα που ακολουθεί είναι ουσιαστικά πόρισμα του θεωρήματος 5.2.1. (Η Xekalaki (1983b) αποδεικνύει το συμπέρασμα αυτό ακολουθώντας όμως άλλη αποδεικτική διαδικασία).

Θεώρημα 5.3.4 Εστω X συνεχής τυχαία μεταβλητή που ορίζεται στο $[\theta,\infty)$, $\theta > 0$. Τότε η X ακολουθεί την κατανομή Pareto (θ,α), $\theta,\alpha > 0$ αν και μόνον αν η συνάρτηση βαθμού κινδύνου κατά τη χρονική στιγμή t είναι αντιστρόφως ανάλογη ως προς t.

Απόδειξη Με βάση το θεώρημα 5.2.1 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση βαθμού κινδύνου h(t). Επομένως αρκεί να υπολογίσουμε τη μορφή της h(t) στην περίπτωση που η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto (θ,α).

$$h(t) = \frac{f_X(t)}{\overline{F}_X(t)}$$

$$= \frac{\alpha\,\theta^{\alpha}\,t^{-(\alpha+1)}}{\theta^{\alpha}\,t^{-\alpha}} \ .$$

Δηλαδή,

$$h(t) = \frac{\alpha}{t} \ , \ t \geq \theta, \ \alpha > 0. \qquad (5.3.5)$$

Το θεώρημα 5.3.5 θα μπορούσε και αυτό να θεωρηθεί ως πόρισμα του θεωρήματος 5.2.1. Για λόγους διάρθρωσης της ύλης προτιμήθηκε να περιληφθεί ως ανεξάρτητο συμπέρασμα στην ενότητα αυτή.

Θεώρημα 5.3.5 Εστω X συνεχής τυχαία μεταβλητή που ορίζεται στο $[\theta,\infty)$, $\theta > 0$. Τότε η X ακολουθεί την κατανομή Pareto (θ,α), $\theta,\alpha > 0$ αν και μόνον αν η μέση υπολειπόμενη ζωή κατα τη χρονική στιγμή t είναι γραμμική συνάρτηση του t.

Απόδειξη Το θεώρημα 5.2.1 αποδεικνύει ότι η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση μέσης υπολειπόμενης ζωής $\mu^{x}(t)$. Ετσι αρκεί να υπολογίσουμε την $\mu^{x}(t)$ στην περίπτωση που η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto (θ,α).

$$\mu^{x}(t) = \frac{1}{\overline{F}_X(t)} \int_t^{\infty} \overline{F}_X(x)\,dx$$

$$= \frac{1}{\theta^{\alpha}\,t^{-\alpha}} \int_t^{\infty} \theta^{\alpha}\,x^{-\alpha}\,dx.$$

Δηλαδή,

$$\mu^x(t) = \frac{1}{\alpha-1}\, t \ , \ t \geq \theta \ , \ \alpha > 1. \qquad (5.3.6)$$

Το θεώρημα που ακολουθεί αποτελεί έναν πολύ απλό αλλά ταυτόχρονα σημαντικό χαρακτηρισμό της κατανομής Pareto (θ,α). Συγκεκριμένα θα αποδείξουμε ότι η κατανομή της X ορίζεται μονοσήμαντα ως Pareto (θ,α), $\theta > 0$, $\alpha > 0$ αν και μόνον αν η συνάρτηση βαθμού κινδύνου τη χρονική στιγμή t είναι πολλαπλάσιο του αντιστρόφου της μέσης υπολειπόμενης ζωής κατά τη χρονική στιγμή t.

Θεώρημα 5.3.6 Εστω X συνεχής τυχαία μεταβλητή που ορίζεται στο $[\theta,\infty)$, $\theta > 0$. Τότε η X ακολουθεί την κατανομή Pareto (θ,α), $\theta > 0$, $\alpha > 1$ αν και μόνον αν

$$h(t) = C\,\frac{1}{\mu^x(t)} \ , \text{ όπου C σταθερά } , \ t \geq \theta. \qquad (5.3.7)$$

Απόδειξη

Ευθύ: Εστω X συνεχής τυχαία μεταβλητή η οποία ακολουθεί την κατανομή Pareto (θ,α), $\theta > 0$, $\alpha > 1$. Τότε από μεν το θεώρημα 5.3.4 προκύπτει ότι,

$$h(t) = \frac{\alpha}{t} \ , \ t \geq \theta \ , \ \alpha > 0,$$

από δε το θεώρημα 5.3.5 ότι,

$$\mu^x(t) = \frac{t}{\alpha-1} \quad , \; t \geq \theta, \; \alpha > 1.$$

Κατά συνέπεια,

$$h(t)\, \mu^x(t) = \frac{\alpha}{\alpha-1} \quad , \quad t \geq \theta, \; \theta > 0, \; \alpha > 1.$$

Αντίστροφο: Εστω ότι ισχύει η σχέση (5.3.7). Δηλαδή,

$$h(t)\, \mu^x(t) = C \quad , \quad t \geq \theta, \; \theta > 0, \; C > 0. \qquad (5.3.8)$$

Επίσης, από τον ορισμό της μέσης υπολειπόμενης ζωής προκύπτει ότι,

$$\mu^x(t) = \frac{1}{\bar{F}_X(t)} \left[\mu - \int_0^t \bar{F}_X(x)\, dx \right] \qquad (6.3.11)$$

όπου μ είναι η αναμενόμενη τιμή της τυχαίας μεταβλητής X η οποία υποτίθεται ότι είναι πεπερασμένη. Προφανώς, $\mu^x(0) = \mu$. Πολλαπλασιάζοντας και τα δύο μέλη της (5.3.11) επί $\bar{F}_X(t)$ και παραγωγίζοντας ως προς t προκύπτει ότι,

$$\mu^{x\prime}(t)\, \bar{F}_X(t) + \mu^x(t)\, \bar{F}'_X(t) = -\, \bar{F}_X(t) \qquad (6.3.12)$$

δηλαδή,

$$\frac{1 + \mu^{x\prime}(t)}{\mu^x(t)} = h(t) \qquad (5.3.13)$$

και άρα,

$$h(t)\ \mu^{x}(t) = 1 + \mu^{x}{}'(t)\ . \qquad (5.3.14)$$

Από τις σχέσεις (5.3.8) και (5.3.14) προκύπτει ότι,

$$1 + \mu^{x}{}'(t) = C\ , \qquad t \geq \theta,\ \ \theta > 0,\ C > 0.$$

Από τη λύση της διαφορικής αυτής εξίσωσης προκύπτει ότι

$$\mu^{x}(t) = (C-1)x\ . \qquad (5.3.15)$$

Με βάση το θεώρημα 5.3.5 η σχέση (5.3.15) στην οποία καταλήξαμε ισοδυναμεί με την πρόταση ότι η κατανομή της X είναι η Pareto (θ,α) όπου $\theta > 0$ και $\alpha = C/(C-1) > 1$. Άρα το θεώρημα απεδείχθη.

Στη συνέχεια θα ασχοληθούμε με συμπεράσματα που αφορούν την κατανομή Yule. Θα υπολογίσουμε τόσο τη συνάρτηση βαθμού πολλαπλασιαστικής αποτυχίας όσο και τη συνάρτηση βαθμού αθροιστικής αποτυχίας. Δυστυχώς, σε αντίθεση με την περίπτωση της κατανομής Pareto τα αποτελέσματα στα οποία καταλήξαμε δεν φαίνεται να επιδέχονται κάποια φυσική ερμηνεία.

<u>Θεώρημα 5.3.7</u> Εστω X διακριτή τυχαία μεταβλητή η οποία ακολουθεί τη μετατοπισμένη κατανομή Yule (ρ) με συνάρτηση πιθανότητας

$$p_x = P(X=x) = \frac{\rho\,(x-1)!}{(\rho+1)_{(x)}}\,, \qquad x = 1,2,\ldots \tag{5.3.16}$$

Τότε,

α) Η συνάρτηση βαθμού αθροιστικής αποτυχίας δίνεται από τη σχέση,

$$\frac{\overline{F}_X(x+t)}{\overline{F}_X(x)} = \frac{(x+1)_{(t)}}{(\rho+x+1)_{(t)}} = \frac{(x+1)_{(\rho)}}{(t+x+1)_{(\rho)}}$$

β) Η συνάρτηση βαθμού πολλαπλασιαστικής αποτυχίας δίνεται από τη σχέση,

$$\frac{\overline{F}_X(x\ t)}{\overline{F}_X(x)} = \frac{(x+1)_{(\rho)}}{(x\ t+1)_{(\rho)}}$$

<u>Απόδειξη</u> Εστω ότι η τυχαία μεταβλητή X ακολουθεί τη μετατοπισμένη κατανομή Yule (ρ) με συνάρτηση πιθανότητας την (5.3.16). Τότε με βάση το συμπέρασμα του λήμματος 4.5.1 στην ειδική περίπτωση που α=1, προκύπτει ότι,

$$r^x(t) = \frac{\overline{F}_X(x+t)}{\overline{F}_X(x)}$$

$$= \frac{P(X>x+t)}{P(X>x)}$$

$$= \frac{\frac{x+t}{\rho}\, P(X=x+t)}{\frac{x}{\rho}\, P(X=x)}$$

$$= \frac{x+t}{x}\, \frac{(x+t-1)!\ (\rho+1)_{(x)}}{(x-1)!\ (\rho+1)_{(x+t)}}$$

$$= \frac{(x+t)!}{x!}\, \frac{(\rho+1)_{(x)}}{(\rho+1)_{(x)}\ (\rho+1+x)_{(t)}}$$

$$= (x+1)(x+2)\ldots(x+t)\, \frac{1}{(\rho+x+1)_{(t)}}$$

$$= \frac{(x+1)_{(t)}}{(\rho+x+1)_{(t)}}$$

$$= \frac{(x+1)_{(\rho)}}{(t+x+1)_{(\rho)}}\ .$$

β) Με όμοιο τρόπο καταλήγουμε στο ότι,

$$r_x(t) = \frac{\overline{F}_X(x\ t)}{\overline{F}_X(x)}$$

$$= \frac{P(X>x\ t)}{P(X>x)}$$

$$= \frac{(x+1)_{(\rho)}}{(t\ x+1)_{(\rho)}}\ .$$

Το θεώρημα το οποίο ακολουθεί είναι ουσιαστικά πόρισμα του θεωρήματος 5.2.1. Και πάλι για λόγους παρουσίασης της ύλης περιλαμβάνεται στην ενότητα αυτή. (Η Xekalaki (1983b) αποδεικνύει και αυτό το συμπέρασμα ακολουθώντας όμως διαφορετική αποδεικτική διαδικασία).

Θεώρημα 5.3.8 Εστω X διακριτή τυχαία μεταβλητή που παίρνει ακέραιες και μη αρνητικές τιμές. Τότε η X ακολουθεί τη κατανομή Yule (ρ), $\rho > 0$ αν και μόνον αν η συνάρτηση βαθμού κινδύνου κατά τη χρονική στιγμή t είναι αντιστρόφως ανάλογη ως προς t.

Απόδειξη Με βάση το θεώρημα 5.2.1 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση βαθμού κινδύνου h(t). Επομένως αρκεί να υπολογίσουμε τη μορφή της h(t) στην περίπτωση που η τυχαία μεταβλητή X ακολουθεί τη κατανομή Yule (ρ).

$$h(t) = \frac{P(X = t)}{P(X \geq t)} .$$

Με βάση το συμπέρασμα του λήμματος 4.5.1 στην ειδική περίπτωση που $\alpha = 1$ προκύπτει ότι,

$$h(t) = \frac{P(X = t)}{P(X>t)+P(X=t)}$$

$$= \frac{P(X = t)}{\frac{t+1}{\rho} P(X=t)+P(X=t)} .$$

Δηλαδή,

$$h(t) = \frac{\rho}{\rho+t+1} \quad . \tag{6.3.17}$$

Το θεώρημα 5.3.9 θα μπορούσε επίσης να θεωρηθεί ως πόρισμα του θεωρήματος 5.2.1.

Θεώρημα 5.3.9 Εστω X διακριτή τυχαία μεταβλητή που παίρνει ακέραιες και μη αρνητικές τιμές. Τότε η X ακολουθεί τη κατανομή Yule (ρ), $\rho > 0$, αν και μόνον αν η μέση υπολειπόμενη ζωή κατά τη χρονική στιγμή t είναι γραμμική συνάρτηση του t.

Απόδειξη Με βάση το θεώρημα 5.2.1 η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα από τη συνάρτηση μέσης υπολειπόμενης ζωής $\mu^x(t)$. Επομένως αρκεί να υπολογίσουμε τη μορφή της $\mu^x(t)$ στην περίπτωση που η τυχαία μεταβλητή X ακολουθεί την κατανομή Yule (ρ).

$$\mu^x(t) = \sum_{x=0}^{\infty} \overline{F}^x(y)$$

$$= \sum_{x=0}^{\infty} \frac{P(X>x+t)}{P(X>t)}$$

$$= \frac{1}{\frac{t+1}{\rho}\, P(X=t)} \sum_{x=0}^{\infty} \frac{x+t+1}{\rho}\, P(X=x+t)$$

$$= \frac{1}{\frac{t+1}{\rho}\, P(X=t)} \sum_{x=0}^{\infty} \frac{x+t+1}{\rho} \frac{\rho\,(x+t)!}{(\rho+1)_{(x+t+1)}}$$

$$= \frac{(\rho+1)_{(t+1)}}{1_{(t+1)}} \sum_{x=0}^{\infty} \frac{1_{(x+t+1)}\, 1_{(x)}}{(\rho+1)_{(x+t+1)}} \frac{1}{x!}$$

$$= \frac{\Gamma(\rho+t+2)\,\Gamma(\rho-1)}{\Gamma(\rho)\,\Gamma(\rho+t+1)} .$$

Δηλαδή,

$$\mu^{x}(t) = \frac{\rho+t+1}{\rho-1} . \qquad (6.3.18)$$

Το θεώρημα 5.3.10 που ακολουθεί είναι αντίστοιχο του θεωρήματος 5.3.6 που αποδείχθηκε για την κατανομή Pareto. Έτσι θα αποδείξουμε ότι η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα ως Yule (ρ), $\rho>0$ αν και μόνον αν η συνάρτηση βαθμού κινδύνου τη χρονική στιγμή t είναι πολλαπλάσιο του αντιστρόφου της μέσης υπολειπόμενης ζωής κατά τη χρονική στιγμή t.

Θεώρημα 5.3.10 Εστω X διακριτή τυχαία μεταβλητή που παίρνει ακέραιες και μη αρνητικές τιμές. Τότε η X ακολουθεί την κατανομή Yule (ρ), $\rho>0$ αν και μόνον αν

$$h(t) = C \frac{1}{\mu^x(t)}, \text{ όπου } C \text{ σταθερά, } t \geq 0. \qquad (5.3.19)$$

Απόδειξη Έστω X διακριτή τυχαία μεταβλητή η οποία ακολουθεί την κατανομή Yule (ρ), $\rho > 0$, $x = 0,1,\ldots$. Τότε με βάση το θεώρημα 5.3.8 προκύπτει ότι,

$$h(t) = \frac{\rho}{\rho+t+1}, \quad t \geq 0, \ \rho > 0.$$

Επίσης, λόγω του θεωρήματος 5.3.9 ισχύει ότι,

$$\mu^x(t) = \frac{\rho+t+1}{\rho-1}, \quad t \geq 0, \ \rho > 0.$$

Κατά συνέπεια,

$$h(t)\ \mu^x(t) = \frac{\rho}{\rho-1}, \quad t \geq 0, \ \rho > 0.$$

Αντίστροφο: Έστω ότι ισχύει η σχέση (5.3.19). Δηλαδή,

$$h(t)\ \mu^x(t) = C, \quad t \geq 0. \qquad (5.3.20)$$

Επίσης, από τον ορισμό της μέσης υπολειπόμενης ζωής προκύπτει ότι,

$$\mu^x(t) = \sum_{x=0}^{\infty} \overline{F}^x(y)$$

$$= \sum_{x=0}^{\infty} \frac{P(X>x+t)}{P(X>t)} .$$

Κατά συνέπεια,

$$\mu^x(t)\ P(X>t) = \sum_{y=t}^{\infty} P(X>y)$$

ή

$$\mu^x(t)\ P(X>t) = \sum_{y=0}^{\infty} P(X>y) - \sum_{y=0}^{t-1} P(X>y) . \qquad (6.3.21)$$

Εξειδικεύοντας τη σχέση (5.3.21) για t=r και t=r+1 και αφαιρώντας τις εξισώσεις που προκύπτουν έχουμε ότι,

$$\mu^x(r+1)\ P(X>r+1) - \mu^x(r)\ P(X>r) = -\ P(X>r)$$

ή

$$\mu^x(r+1)\ P(X>r+1) - \left[\mu^x(r) - 1\right] P(X>r) = 0 \Rightarrow$$

$$\mu^x(r+1)\ \left[P(X\geq r+1) - P(X=r+1)\right] - \left[\mu^x(r) - 1\right] P(X>r) = 0 \Rightarrow$$

$$\left[\mu^x(r+1) - \mu^x(r) + 1\right] P(X\geq r+1) - \mu^x(r+1)\ P(X=r+1) = 0 \Rightarrow$$

$$\frac{\mu^x(r+1) - \mu^x(r) + 1}{\mu^x(r+1)} = \frac{P(X=r+1)}{P(X\geq r+1)} . \qquad (6.3.22)$$

Αντικαθιστώντας στη σχέση (5.3.22) τη σχέση (5.2.7) προκύπτει ότι,

$$\frac{\mu^x(r+1) - \mu^x(r) + 1}{\mu^x(r+1)} = h(r+1) \Rightarrow$$

$$\mu^x(r+1) - \mu^x(r) + 1 = \mu^x(r+1)\, h(r+1) \qquad (6.3.23)$$

Λογω της σχέσης (5.3.20) η σχέση (5.3.23) γίνεται,

$$\mu^x(r+1) - \mu^x(r) + 1 = C \ ,$$

και τελικά,

$$\mu^x(r+1) - \mu^x(r) = C-1, \quad r \geq 0.$$

Η λύση της εξίσωσης αυτής διαφορών δίνεται από τη σχέση,

$$\mu^x(r) = k + (C-1)\, r \ , \quad r \geq 0, \text{ όπου } k \text{ σταθερά} \qquad (5.3.24)$$

και C σταθερά η οποία

ορίζεται από τη σχέση (5.3.19).

Με βάση το συμπέρασμα του θεωρήματος (5.3.9) η σχέση (5.3.24) είναι η ικανή και αναγκαία συνθήκη για να ακολουθεί η X την κατανομή Yule (ρ) όπου $\rho = C/(C-1) = (k+1)/(k-1) > 0$.

5.4 ΟΙ ΔΙΔΙΑΣΤΑΤΕΣ ΚΑΤΑΝΟΜΕΣ PARETO ΚΑΙ YULE ΣΤΟ ΠΛΑΙΣΙΟ ΤΗΣ ΘΕΩΡΙΑΣ ΑΞΙΟΠΙΣΤΙΑΣ

Στην ενότητα αυτή γίνεται προσπάθεια μεταφοράς στο χώρο των δυο διαστάσεων ορισμένων συμπερασμάτων που αποδείχθηκαν στην προηγούμενη ενότητα. Με το θεώρημα 5.3.3 αποδείξαμε ότι η μονοδιάστατη κατανομή Pareto έχει σταθερή συνάρτηση βαθμού πολλαπλασιαστικής αποτυχίας. Με το θεώρημα 5.4.1 αποδεικνύουμε ότι από τις διδιάστατες κατανομές Pareto η τύπου Marshall-Olkin διατηρεί την αντίστοιχη ιδιότητα στο χώρο των δυο διαστάσεων.

<u>Θεώρημα 5.4.1</u> Εστω το τυχαίο διάνυσμα $\underset{\sim}{X} = (X_1, X_2)$ και έστω ότι ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin τότε η διδιάστατη συνάρτηση βαθμού πολλαπλασιαστικής αποτυχίας είναι σταθερή ως προς $\underset{\sim}{x} = (x_1, x_2)$ για κάθε $t \geq 1$.

<u>Απόδειξη</u> Εστω $\underset{\sim}{X} = (X_1, X_2)$ το τυχαίο διάνυσμα το οποίο ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin. Τότε, με βάση τον ορισμό της διδιάστατης συνάρτησης βαθμού πολλαπλασιαστικής αποτυχίας που δίνεται από τη σχέση (5.2.10) προκύπτει ότι,

$$\frac{\bar{F}_{\underset{\sim}{X}}\left(\underset{\sim}{x}\, t\right)}{\bar{F}_{\underset{\sim}{X}}\left(\underset{\sim}{x}\right)} = \frac{(x_1 t)^{-\alpha k}\,(x_2 t)^{-\alpha l}\left\{\max\left[(x_1 t)^{-\alpha}, (x_2 t)^{-\alpha}\right]\right\}^{-r}}{x_1^{-\alpha k}\, x_2^{-\alpha l}\left\{\max\left[x_1^{-\alpha}, x_2^{-\alpha}\right]\right\}^{-r}}$$

$$= t^{-\alpha\,(k+1+r)} \quad .$$

Με το θεώρημα 5.3.1 αποδείξαμε ότι η μονοδιάστατη κατανομή Pareto χαρακτηρίζεται από την έλλειψη πολλαπλασιαστικής μνήμης της οποίας ο ορισμός περιγράφεται από τη σχέση (5.3.2). Μια προφανής μεταφορά του ορισμού αυτού στο χώρο των δυο διαστάσεων αποτελεί ο ορισμός που ακολουθεί

<u>Ορισμός 5.4.1</u> Εστω το τυχαίο διάνυσμα $\underset{\sim}{X} = (X_1, X_2)$. Η διδιάστατη κατανομή την οποία ακολουθεί το τυχαίο αυτό διάνυσμα θα χαρακτηρίζεται από έλλειψη πολλαπλασιαστικής μνήμης αν και μόνον αν ισχύει η σχέση:

$$P\left(X_1 > x_1 t_1, X_2 > x_2 t_2 \mid X_1 > t_1, X_2 > t_2\right) = P\left(X_1 > x_1, X_2 > x_2\right) \qquad (5.4.1)$$

όπου $x_1, x_2, t_1, t_2 \geq 1$.

Παρατηρούμε ότι η σχέση (5.4.1) είναι ισοδύναμη προς την

$$\overline{F}_{X_1, X_2}\left(x_1 t_1, x_2 t_2\right) = \overline{F}_{X_1, X_2}\left(x_1, x_2\right)\, \overline{F}_{X_1, X_2}\left(t_1, t_2\right) \qquad (5.4.2)$$

η οποία με βάση το θεώρημα 5.2.3 ορίζει μονοσήμαντα την κατανομή του (X_1, X_2) ως διπλή Pareto. Ετσι αποδεικνύεται το παρακάτω θεώρημα:

Θεώρημα 5.2.2 Εστω το τυχαίο διάνυσμα (X_1,X_2) του οποίου οι συνιστώσες είναι συνεχείς και μεγαλύτερες από τη μονάδα. Τότε η σχέση

$$P\left(X_1>x_1t_1,X_2>x_2t_2|X_1>t_1,X_2>t_2\right) = P\left(X_1>x_1,X_2>x_2\right)$$

όπου $x_1,x_2,t \geq 1$.

ορίζει μονοσήμαντα την κατανομή του (X_1,X_2) ως διπλή Pareto.

Εν τούτοις, με το θεώρημα 4.2.4 αποδεικνύουμε ότι στην ειδική περίπτωση που $t_1=t_2=t$ οπότε η (5.4.2) γράφεται

$$\overline{F}_{X_1,X_2}\left(x_1t,x_2t\right) = \overline{F}_{X_1,X_2}\left(x_1,x_2\right)\overline{F}_{X_1,X_2}\left(t,t\right) \qquad (5.4.3)$$

η κατανομή του τυχαίου διανύσματος (X_1,X_2) ορίζεται μονοσήμαντα ως δισδιάστατη Pareto τύπου Marshall-Olkin. Ετσι οδηγούμεθα σε ένα πολύ σημαντικό συμπέρασμα για την κατανομή αυτή που, αν και η απόδειξη του είναι σχεδόν προφανής, θα διατυπώσουμε με το παρακάτω θεώρημα.

Θεώρημα 5.4.2 Εστω το τυχαίο διάνυσμα (X_1,X_2) του οποίου οι συνιστώσες είναι συνεχείς και μεγαλύτερες από τη μονάδα. Τότε η σχέση

$$P\left(X_1>x_1t,X_2>x_2t|X_1>t,X_2>t\right) = P\left(X_1>x_1,X_2>x_2\right) \qquad (5.4.4)$$

όπου $x_1,x_2,t \geq 1$.

ορίζει μονοσήμαντα την κατανομή του (X_1,X_2) ως διδιάστατη Pareto τύπου Marshall-Olkin αν οι περιθώριες κατανομές κατανομές των X_1 και X_2 είναι μονοδιάστατες Pareto.

<u>Απόδειξη</u> Η σχέση (5.4.4) είναι προφανώς ισοδύναμη προς τη συναρτησιακή εξίσωση

$$\overline{F}_{X_1,X_2}\left(x_1t,x_2t\right) = \overline{F}_{X_1,X_2}\left(x_1,x_2\right)\overline{F}_{X_1,X_2}\left(t,t\right) . \qquad (5.4.5)$$

Μοναδική δε λύση της (5.4.5) με βάση το θεώρημα 4.2.4 είναι η

$$\overline{F}_{X_1,X_2}\left(x_1,x_2\right) = x_1^{-\alpha k}\, x_2^{-\alpha l}\left\{\max\left[x_1^{-\alpha}, x_2^{-\alpha}\right]\right\}^{-r}, \quad x_1,x_2 \geq 1.$$

Το τυχαίο διάνυσμα (X_1,X_2) ακολουθεί επομένως τη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin.

Με το θεώρημα 5.4.2 αποδείχθηκε ότι η τύπου Marshall-Olkin είναι η μόνη από τις διδιάστατες Pareto που χαρακτηρίζεται από την έλλειψη πολλαπλασιαστικής μνήμης μεταφέροντας έτσι στο χώρο των δυο διαστάσεων τη γνωστή ιδιότητα της μονοδιάστατης κατανομής Pareto. Στο συμπέρασμα αυτό καθώς και σε εκείνο του θεωρήματος 5.4.1 στηρίζεται το θεώρημα το οποίο ακολουθεί.

<u>Θεώρημα 5.4.3</u> Το τυχαίο διάνυσμα $\underset{\sim}{X} = (X_1,X_2)$ ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Marshall-Olkin αν από την

πρόταση

$$\bar{F}_{\underset{\sim}{X}}\left(x_1 t\ ,\ x_2 t\right) = \bar{F}_{\underset{\sim}{X}}\left(x_1\ ,x_2\right)\ \bar{F}_{\underset{\sim}{X}}\left(t\ ,\ t\right) \qquad (5.4.6)$$

προκύπτει ότι η διδιάστατη συνάρτηση βαθμού πολλαπλασιαστικής αποτυχίας είναι σταθερή ως προς $\underset{\sim}{x} = (x_1, x_2)$ για κάθε $t \geq 1$.

<u>Απόδειξη</u> Εστω $\underset{\sim}{X}$ τυχαίο διάνυσμα το οποίο ικανοποιεί τη σχέση (5.4.6). Τότε με βάση το θεώρημα 4.2.4 προκύπτει ότι το $\underset{\sim}{X}$ ακολουθεί τη διδιάστατη Pareto τύπου Marshall-Olkin. Κατά συνέπεια με βάση το θεώρημα 5.4.1 η διδιάστατη συνάρτηση βαθμού πολλαπλασιαστικής αποτυχίας είναι σταθερή ως προς $\underset{\sim}{x} = (x_1, x_2)$ για κάθε $t \geq 1$.

Το θεώρημα το οποίο ακολουθεί αποτελεί χαρακτηρισμό της διδιάστατης κατανομής Pareto τύπου Mardia I . Είναι μεταφορά στο χώρο των δυο διαστάσεων του συμπεράσματος το οποίο περιγράφεται από το θεώρημα 5.3.4 και το οποίο αφορά τη μονοδιάστατη κατανομή Pareto και βασίζεται στη διδιάστατη συνάρτηση βαθμού κινδύνου έτσι όπως αυτή ορίζεται από τη σχέση (5.2.21).

<u>Θεώρημα 5.4.4</u> Το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\alpha > 0$ αν και μόνον αν η διδιάστατη συνάρτηση βαθμού κινδύνου ικανοποιεί τη σχέση

$$h(x,y) = \frac{\alpha(\alpha+1)}{(x+y+1)^2} \quad x,y \geq 0. \qquad (5.4.7)$$

Απόδειξη

Ευθύ: Εστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I. Τότε με βάση τον ορισμό της διδιάστατης συνάρτησης βαθμού κινδύνου προκύπτουν διαδοχικά,

$$h(x,y) = \frac{f_{X,Y}(x,y)}{\bar{F}_{X,Y}(x,y)}$$

$$= \frac{\alpha\ (\alpha+1)\ (x+y-1)^{-(\alpha+2)}}{(x+y+1)^{-\alpha}}$$

$$= \frac{\alpha(\alpha+1)}{(x+y+1)^2} .$$

Αντίστροφο: Εστω ότι,

$$h(x,y) = \frac{\alpha(\alpha+1)}{(x+y+1)^2} .$$

Δηλαδή,

$$\frac{f_{X,Y}(x,y)}{\bar{F}_{X,Y}(x,y)} = \frac{a(a+1)}{(x+y+1)^2} ,$$

ή

$$\frac{\dfrac{\partial^2 \bar{F}_{X,Y}(x,y)}{\partial x\, \partial y}}{\bar{F}_{X,Y}(x,y)} = \frac{a(a+1)}{(x+y+1)^2}$$

ή

$$\frac{\partial^2 \bar{F}_{X,Y}(x,y)}{\partial x\, \partial y} - \frac{a(a+1)}{(x+y+1)^2} \bar{F}_{X,Y}(x,y) = 0$$

Η λύση της διαφορικής αυτής εξίσωσης υπό τον περιορισμό ότι $\bar{F}_{X,Y}(1,1) = 1$ δίνεται από τη σχέση,

$\bar{F}_{X,Y}(x,y) = (x+y+1)^{-a}$. Αρα το θεώρημα απεδείχθη.

Στη συνέχεια θα χαρακτηρίσουμε τη διδιάστατη κατανομή Yule μεταφέροντας στο χώρο των δυο διαστάσεων το συμπέρασμα του θεωρήματος 5.3.8 το οποίο αφορά τη μονοδιάστατη κατανομή Yule. Συγκεκριμένα θα αποδείξουμε ότι η διδιάστατη κατανομή Yule ορίζεται μονοσήμαντα από τη μορφή της διδιάστατης συνάρτησης βαθμού κινδύνου έτσι όπως αυτή ορίζεται από τη σχέση (6.2.22). Πριν από την απόδειξη του κυρίως συμπεράσματος το οποίο

συνοψίζεται από το θεώρημα 5.4.5 κρίνεται απαραίτητη η παρουσίαση με τη μορφή λήμματος του παρακάτω συμπεράσματος.

<u>Λήμμα 5.4.1</u> Αν το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Yule (ρ), $\rho > 0$, με συνάρτηση πιθανότητας την

$$p_{r,l} = \frac{\rho_{(2)}\ (r+1)!}{(\rho+1)_{(r+l+2)}}\ , \qquad r,l = 0,1,2,\ldots,\ \rho>0. \tag{5.4.8}$$

Τότε,

$$P(X>x\ ,\ Y>y) = \frac{(x+y+1)(x+y+2)}{\rho(\rho+1)}\ p_{x,y}\ . \tag{5.4.9}$$

<u>Απόδειξη</u> Εστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Yule (ρ), $\rho > 0$ με συνάρτηση πιθανότητας εκείνην που δίνεται από τη σχέση (5.4.8). Τότε,

$$P(X>x\ ,\ Y>y) = \sum_{u=x+1}^{\infty}\ \sum_{v=y+1}^{\infty} \frac{\rho_{(2)}\ (u+v)!}{(\rho+1)_{(u+v+2)}}$$

$$= \rho_{(2)} \sum_{u=0}^{\infty}\ \sum_{v=0}^{\infty} \frac{(u+v+x+y+2)!}{(\rho+1)_{(u+v+x+y+4)}}$$

$$= \frac{\rho_{(2)}\,\Gamma(x+y+3)}{(\rho+1)_{(x+y+4)}} \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \frac{\dfrac{\Gamma(u+v+x+y+3)}{\Gamma(x+y+3)}}{(\rho+x+y+5)_{(u+v)}}$$

$$= \frac{\rho_{(2)}\,(x+y+2)!}{(\rho+1)_{(x+y+4)}} \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \frac{(x+y+3)_{(u+v)}}{(\rho+x+y+5)_{(u+v)}}$$

$$= \frac{\rho_{(2)}\,(x+y+2)!}{(\rho+1)_{(x+y+4)}} \frac{(x+y+3+\rho)_{(2)}}{\rho_{(2)}} \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \frac{\rho_{(2)}}{(x+y+3+\rho)_{(2)}} \frac{(x+y+3)_{(u+v)}}{(\rho+x+y+5)_{(u+v)}}$$

Αρα τελικά,

$$P(X>x\ ,\ Y>y) = \frac{(x+y+2)!\ (x+y+3+\rho)_{(2)}}{(\rho+1)_{(x+y+4)}}$$

$$= \frac{(x+y+2)!\ (x+y+3+\rho)_{(2)}}{(\rho+1)_{(x+y+2)}(\rho+1+x+y+2)_{(2)}}$$

$$= \frac{(x+y+2)(x+y+1)}{\rho_{(2)}} \frac{(x+y)!\ \rho_{(2)}}{(\rho+1)_{(x+y+2)}}$$

$$= \frac{(x+y+1)(x+y+2)}{\rho(\rho+1)}\ p_{x,y}\ .$$

Θεώρημα 5.4.5 Το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Yule (ρ), $\rho > 0$ με συνάρτηση πιθανότητας την (5.4.8) αν και μόνον αν η διδιάστατη συνάρτηση βαθμού κινδύνου ικανοποιεί τη σχέση

$$h(x,y) = \frac{\rho(\rho+1)}{(x+y+1)(x+y+2)} .$$

Απόδειξη

Ευθύ: Εστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Yule (ρ), $\rho > 0$ με συνάρτηση πιθανότητας εκείνην που δίνεται από τη σχέση (5.4.8). Τότε, λόγω και του λήμματος 5.4.1 προκύπτει ότι,

$$h(x,y) = \frac{P(X=x,Y=y)}{P(X>x,Y>y)}$$

$$= \frac{p_{x,y}}{\frac{(x+y+1)(x+y+2)}{\rho(\rho+1)}\, p_{x,y}}$$

$$= \frac{\rho(\rho+1)}{(x+y+1)(x+y+2)} .$$

Αντίστροφο: Εστω ότι,

$$h(x,y) = \frac{\rho(\rho+1)}{(x+y+1)(x+y+2)} .$$

Αλλά,

$$h(x,y) = \frac{p_{x,y}}{\sum_{u=x+1}^{\infty}\sum_{v=y+1}^{\infty} p_{u,v}}$$

Επομένως,

$$\frac{p_{x,y}}{\sum_{u=x+1}^{\infty}\sum_{v=y+1}^{\infty} p_{u,v}} = \frac{\rho(\rho+1)}{(x+y+1)(x+y+2)} .$$

Δηλαδή,

$$p_{r,l} = \frac{\rho(\rho+1)}{(x+y+1)(x+y+2)} \sum_{x=r+1}^{\infty}\sum_{y=l+1}^{\infty} p_{x,y} . \qquad (5.4.10)$$

Από τη σχέση (5.4.10) για $r,l = 0,1,\ldots$ προκύπτουν διαδοχικά οι παρακάτω τρεις σχέσεις,

$$(r+l+2)(r+l+1)p_{r,l} - (r+l+3)(r+l+2)p_{r+1,l} = \rho(\rho+1)\sum_{x=l+1}^{\infty} p_{r+1,x} \qquad (5.4.11)$$

$$(r+l+2)(r+l+1)p_{r,l} - (r+l+3)(r+l+2)p_{r,l+1} = \rho(\rho+1)\sum_{x=r+1}^{\infty} p_{x,l+1} \qquad (5.4.12)$$

$$(r+l+2)(r+l+1)p_{r,l}-(r+l+4)(r+l+3)p_{r+1,l+1} =$$

$$=\rho(\rho+1)\left[\sum_{x=l+1}^{\infty} p_{r+1,x} + \sum_{x=r+1}^{\infty} p_{x,l+1} + p_{r+1,l+1}\right]. \quad (5.4.13)$$

Με βάση τις σχέσεις (5.4.11), (5.4.12) και (5.4.13) προκύπτει ότι,

$$(r+l+4)(r+l+3)p_{r+1,l+1}+(r+l+2)(r+l+1)p_{r,l}-(r+l+3)(r+l+2)p_{r+1,l}-$$

$$-(r+l+3)(r+l+2)p_{r,l+1} = \rho(\rho+1)p_{r+1,l+1} ,$$

ή

$$\left[(r+l+4)(r+l+3)-\rho(\rho+1)\right]p_{r+1,l+1}+(r+l+2)(r+l+1)p_{r,l}-$$

$$-(r+l+3)(r+l+2)p_{r+1,l}-(r+l+3)(r+l+2)p_{r,l+1} = 0.$$

Λύση της εξίσωσης αυτής διαφορών είναι η,

$$p_{r,l} = \frac{\rho_{(2)}\ (r+l)!}{(\rho+1)_{(r+l+2)}} , \quad r,l = 0,1,2,..., \ \rho>0.$$

ΚΕΦΑΛΑΙΟ 6

ΜΟΝΟΔΙΑΣΤΑΤΑ

ΣΥΣΤΗΜΑΤΑ ΚΑΤΑΝΟΜΩΝ

6.1 ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό μελετώνται οι μονοδιάστατες σταθμισμένες κατανομές *(univariate weighted distributions)* και οι μονοδιάστατες κατανομές που βασίζονται σε μερικά αθροίσματα *(univariate distributions based on partial sums)*. Αντικειμενικός σκοπός του κεφαλαίου είναι η διατύπωση και απόδειξη χαρακτηριστικών ιδιοτήτων γνωστών κατανομών μέσω ιδιοτήτων των σταθμισμένων κατανομών αφενός, και των κατανομών των βασισμένων στα μερικά αθροίσματα αφετέρου. Λόγω του μεγάλου ενδιαφέροντος που παρουσιάζουν οι σταθμισμένες κατανομές ως αυτοτελές θέμα μελέτης, θα ήταν χρήσιμη μια ενημερωτική εισαγωγή στο αντικείμενο αυτό. Θα ασχοληθούμε τόσο με τη συνεχή όσο και με τη διακριτή περίπτωση.

6.2 ΜΟΝΟΔΙΑΣΤΑΤΕΣ ΣΤΑΘΜΙΣΜΕΝΕΣ ΚΑΤΑΝΟΜΕΣ

Η έννοια των σταθμισμένων κατανομών μπορεί να ανιχνευτεί στη μελέτη του R.A. Fisher (1934) ο οποίος διατύπωσε, μεταξύ άλλων, την άποψη ότι το πρόβλημα της στατιστικής συμπερασματολογίας μπορεί να αναλυθεί στα εξής επιμέρους προβλήματα: εξειδίκευση του πιθανοθεωρητικού μοντέλου, εκτίμηση των παραμέτρων, έλεγχος των υποθέσεων και τέλος ερμηνεία των αποτελεσμάτων προκειμένου να ληφθεί κάποια απόφαση. Εδωσε δε, μεγάλη σημασία στην εξειδίκευση του μοντέλου η οποία ισοδυναμεί ουσιαστικά με την επιλογή μιας οικογένειας μέτρων πιθανότητας σε ένα δειγματικό χώρο. Είναι προφανές ότι λανθασμένη επιλογή μοντέλου θα οδηγήσει σε λανθασμένο συμπέρασμα. Ο C.R. Rao (1965) στο περίφημο άρθρο του *"on discrete*

distributions arising out of methods of ascertainement" πρώτος διαπιστώνει τη χρησιμότητα των σταθμισμένων κατανομών και διατυπώνει σαφώς την έννοια τους. Υπογραμμίζει ότι υπάρχουν καταστάσεις για την περιγραφή των οποίων το καταλληλότερο μοντέλο είναι εκείνο των σταθμισμένων κατανομών. Οι καταστάσεις αυτές αναφέρονται στις περιπτώσεις εκείνες που οι καταγραφόμενες παρατηρήσεις (recorded observations) δεν μπορούν να θεωρηθούν ως τυχαίο δείγμα από την πραγματική κατανομή (original distribution). Κατι τέτοιο είναι δυνατό να συμβεί είτε γιατί κάποια ενδεχόμενα είναι μη παρατηρήσιμα (non-observable) είτε γιατί οι τιμές της αρχικής παρατήρησης υφίστανται μερική καταστροφή (αλλοίωση,στρέβλωση partial destruction) είτε τέλος γιατί η δειγματοληπτική τεχνική που χρησιμοποιείται δίνει άνισες πιθανότητες επιλογής στις δειγματοληπτικές μονάδες του δειγματοληπτικού πλαισίου.

Στα 25 χρόνια που μεσολάβησαν από τη διατύπωση του C.R. Rao (1965) μέχρι σήμερα οι σταθμισμένες κατανομές ανάπτύχθηκαν ως ένα χρήσιμο εργαλείο επιλογής του κατάλληλου μοντέλου για τα παρατηρούμενα δεδομένα, δεν μπορούμε όμως να πούμε ότι το πλήθος των δημοσιευμένων εργασιών στο αντικείμενο αυτό είναι ανάλογο του μεγάλου ενδιαφέροντος που παρουσιάζουν. Σημαντική προσφορά στο αντικείμενο αποτελούν οι εργασίες των Patil and Ord (1975), Patil and Rao (1977, 1978), Mahfoud and Patil (1981,1982), Rao (1985), Patil, Rao and Ratnaparkhi (1986).

Το φάσμα των εφαρμογών των σταθμισμένων κατανομών είναι ευρύτατο και μεταξύ άλλων περιλαμβάνει εφαρμογές στην περιοχή μελέτης ανθρώπινων ομάδων (οικογενειών κ.λ.π.) (Haldane (1938),

Rao (1965), Neel and Schull (1966), Stene (1981)], όπως επίσης και στην μελέτη κοινωνικών προβλημάτων (αλκοολισμός, κ.λ.π.) [Smart (1963)], οικολογικών προβλημάτων [Cook and Martin (1974)], και διακριτών ανανεωτικών συστηματων [Feller (1966), Uppuluri and Patil (1976)]. Το πρόβλημα της παραγωγής συνεχών σταθμισμένων κατανομών συνδέεται με την παραγωγή τυχαίων μεταβλητών με τη χρήση ηλεκτρονικού υπολογιστή [Patil, Bonwell and Friday (1975)], θεωρία ανανέωσης [Cox (1962,1969)], παράδοξο συνεχούς χρόνου αναμονής [Feller (1966)], κινητικότητα κυττάρων και έγκαιρη διάγνωση ασθενειών [Takahashi (1966, 1968), Zelen (1969, 1971, 1974) Schotz and Zelen (1971), Rosenzweig et al (1978)], δασικές μελέτες [Warren (1974, 1975)], και διάφορα άλλα προβλήματα [Kendall and Moran (1963), Watson (1971),Brown (1972)]. Οι moment distributions που χρησιμοποιούνται ευρύτατα στην οικονομική θεωρία είναι και αυτές ειδικές περιπτώσεις σταθμικών κατανομών [Aitchison and Brown (1957), Klein (1962), Hart (1975 Ord (1975)] . Οι moment distributions χρησιμοποιούνται επίσης σε προβλήματα αξιοπιστίας [Wallach and Sichel (1963), Schaeffer (1972), Vardi et al (1981)], στην ιατρική [Zelen (1971), Simon (1975, 1980)], στη μελέτη

οικολογικών προβλημάτων (Pielou (1969), Ord (1972), Belle and Schneidermann (1973), Scultz (1975), Patil (1981)) τέλος στο πλαίσιο του μοντέλου ανανέωσης αποθέματος τόσο στη διακριτή όσο και στη συνεχή περίπτωση (Xekalaki (1988a, 1988b)). Στην ειδική αυτή μορφή των σταθμισμένων κατανομών θα επανέλθουμε αργότερα.

Ας θεωρήσουμε ένα μηχανισμό ο οποίος παράγει μια τυχαία μεταβλητή X με συνάρτηση (πυκνότητας) πιθανότητας $f_X(x)$. Εστω επίσης w(x) μια μη αρνητική συνάρτηση βάρους (weight function) και ας υποθέσουμε ότι η E[w(X)] υπάρχει. Ορίζουμε μια καινούρια συνάρτηση (πυκνότητας) πιθανότητας με βάση τον τύπο,

$$f_X^w(x) = \frac{w(x)\, f(x)}{E[w(X)]} \qquad (6.2.1)$$

και συμβολίζουμε την αντίστοιχη τυχαία μεταβλητή με X^w. Θα ονομάζουμε τις X και X^w **αρχική τυχαία μεταβλητή και σταθμισμένη τυχαία μεταβλητή** αντίστοιχα και τις $f_X(x)$ και $f_X^w(x)$ **αρχική συνάρτηση (πυκνότητας) πιθανότητας** και **σταθμισμένη συνάρτηση (πυκνότητας) πιθανότητας** αντίστοιχα. Μερικές από τις συναρτήσεις βάρους που χρησιμοποιούνται στην εφαρμοσμένη στατιστική ανάλυση περιέχονται στον Πίνακα 6.2.1.

Π Ι Ν Α Κ Α Σ 6.2.1

ΣΥΝΑΡΤΗΣΕΙΣ ΒΑΡΟΥΣ

w(x)	Τυχαία μεταβλητή
x^{α}	διακριτή ή συνεχής
x+1	διακριτή
x(x-1)...(x-r+1)	διακριτή
$e^{\alpha x}$	συνεχής

Στην Οικονομική Επιστήμη όπως ήδη αναφέραμε χρησιμοποιούνται οι moment distributions οι οποίες είναι σταθμισμένες κατανομές με συναρτήσεις βάρους $w(x) = x^j$. Ενδεικτική εφαρμογή αναφέρουμε τη μέτρηση της ανισότητας, με τη βοήθεια της καμπύλης Lorenz, των έντονα ασσύμετρων κατανομών εισοδήματος, πλούτου, μεγέθους επιχειρήσεων κ.λ.π.

Στο πλαίσιο της στατιστικής ανάλυσης και στην περίπτωση που η συνάρτηση βάρους είναι η $w(x) = x^a$, $a>0$ η τυχαία μεταβλητή X^w ονομαζεται size-biased version της X τάξης a και συμβολίζεται με X^{*a}. Στην ειδική περίπτωση που a=1 η X^w συμβολίζεται με $X^*=X^{*1}$ και ονομάζεται απλά size-biased version της X και η αντίστοιχη size-biased συνάρτηση (πυκνότητας) πιθανότητας δίνεται απο τον τύπο,

$$f_X^*(x) = \frac{x\, f(x)}{E(X)} \qquad (6.2.2)$$

Η ονομασία αυτή οφείλεται στο ότι η στάθμιση της f(x) μέσω της (6.2.2) ή της γενικότερης μορφής προκύπτει σε περιπτώσεις που η πιθανότητα να παρατηρηθεί μία τιμή x της μεταβλητής X αυξάνει αναλόγως με το μέγεθος της x. Για παράδειγμα στην ανάλυση εισοδήματος είναι περισσότερο πιθανό να παρατηρηθεί ένα υψηλό εισόδημα από ένα χαμηλότερο, όπως επίσης σε οικολογικές μελέτες είναι περισσότερο πιθανό να παρατηρηθεί ένα μεγάλο κοπάδι ζώων από ένα μικρότερο.

Για τις σταθμισμένες κατανομές με w(x) = x αποδεικνύεται ότι ισχύουν μερικές πολύ ενδιαφέρουσες ιδιότητες. Θα συνοψίσουμε τέσσερις από αυτές στο παρακάτω θεώρημα του οποίου η απόδειξη είναι εύκολη.

Θεώρημα 6.2.1 Εστω η μη αρνητική τυχαία μεταβλητή X με συνάρτηση (πυκνότητας) πιθανότητας $f_X(x)$. Εστω επίσης ότι οι τρεις πρώτες ροπές υπάρχουν τόσο για την αρχική μεταβλητή X όσο και για τη size biased version αυτής. Τότε;

a. $\mu^* = \mu + \frac{\sigma^2}{\mu}$ όπου $\mu^* = E(X^*)$

$\mu = E(X)$

$\mu_2 = E(X-\mu)^2 = \sigma^2$

b. $\mu^* = \mu_2 + \left(\frac{\mu_2}{\mu}\right)^2 \left(\gamma_1/c-1\right)$ όπου $\gamma_1 = \mu_3/\sigma^3$

$c = \sigma/\mu$

$\mu_3 = E(X^3)$

c. $\left[\mu^*_{-1}\right]^{-1} = \mu$ όπου $\mu^*_{-1} = E\left(X^{*-1}\right)$

d. $\mu_2 = \left[\mu^*_{-1}\right]^{-1} \left[\mu^* - \left(\mu'^*_{-1}\right)^{-1}\right]$

Στις πρακτικές εφαρμογές παρουσιάζει ενδιαφέρον να γνωρίζουμε πότε οι συναρτήσεις (πυκνότητας) πιθανότητας $f_X(x)$ και $f^*_X(x)$ ή και γενικότερα οι $f_X(x)$ και $f^w_X(x)$, αντιστοιχούν σε κατανομές του ιδίου τύπου, με διαφορετικές ενδεχόμενα τιμές παραμέτρων μόνο. Ετσι, πληροφορίες που συλλεχτηκαν με δειγματοληψία size-biased θα μπορούν να συνδεθούν με τον αρχικό πληθυσμό και να οδηγήσουν σε συμπεράσματα για αυτόν. Σε θεωρητικό επίπεδο, παρουσιάζει ενδιαφέρον να ξέρουμε κατά πόσο η ιδιότητα του αμετάβλητου της μορφής *(form invariance property)* είναι χαρακτηριστική για κάποιες οικογένειες κατανομών. Ο όρος αμετάβλητο της μορφής *(form invariance)* χρειάζεται πιό αυστηρή περιγραφή. Αυτό επιτυγχάνεται με τον παρακάτω ορισμό.

<u>Ορισμός 6.2.1</u> Εστω η τυχαία μεταβλητή X με συνάρτηση (πυκνότητας) πιθανότητας $f_X(x;\theta)$. Θα λέμε ότι η $f_X(x,\theta)$ είναι αμετάβλητη ως προς τη μορφή *(form invariant)* υπό τη συνάρτηση βάρους $w(x) = x^a$ αν η τυχαία μεταβλητή X^w έχει συνάρτηση (πυκνότητας) πιθανότητας την $f_X(x,\eta)$, όπου η είναι η νέα παράμετρος η οποία εξαρτάται από την παλιά παράμετρο θ και τη σταθερά a.

Με άλλα λόγια, από τον παραπάνω ορισμό προκύπτει ότι ότι η παρατηρούμενη μεταβλητή $\left(X(\theta)\right)^{*a}$ είναι ισοδύναμη ως προς την κατανομή *(equivalent in distribution)* με τη μεταβλητή $X(\eta)$, την αρχική δηλαδή μεταβλητή με παράμετρο η. Συμβολικά μπορούμε να γράψουμε ότι

$$\left(X(\theta)\right)^{*a} = X(\eta) \quad \Rightarrow \quad f_X^{*a}(x;\theta) = f_X(x;\eta).$$

Στους Πίνακες 6.2.2, και 6.2.3 συνοψίζονται πληροφορίες σχετικές με τις size-biased μορφές γνωστών κατανομών. Τα στοιχεία των πινακων αυτών βασίζονται είτε στην υπάρχουσα βιβλιογραφία είτε σε συμπεράσματα της ενότητας 6.3 που ακολουθεί. Παρατηρούμε ότι με εξαίρεση την περίπτωση της κατανομής Yule οι υπόλοιπες διακριτές κατανομές πέρα από τη μετατόπιση προς τα δεξιά, διατηρούν τη μορφή τους υπό τη συνάρτηση βάρους $w(x) = x$. Αξίζει να αναφέρουμε ότι αν ορίσουμε σταθμική κατανομή με βάση τον τύπο

$$f_X^{*(j)} = \frac{x^{(j)} f_X(x)}{E\left[X^{(j)}\right]} \qquad ,x = j,\ j+1,\ldots \qquad (6.2.3)$$

όπου $E\left[X^{(j)}\right]$ είναι η παραγοντική ροπή τάξης j και $x^{(j)} = x(x-1)\ldots(x-j+1)$, τότε οι κατανομές του Πινακα 6.2.2 διατηρούν την αρχική τους μορφή με μία μετατόπιση κάτω από την συνάρτηση βάρους παραγοντικής τάξης $j=1$. Σημειώνουμε ότι στην περίπτωση αυτή τα συμπεράσματα είναι γενικότερα αφού αποδεικνύεται ότι ισχύουν για τάξη j. Ορισμένα συμπεράσματα συνοψίζονται στον Πίνακα 6.2.4. Από αυτά θα διατυπώσουμε και θα αποδείξουμε υπό μορφή θεωρήματος την περίπτωση της κατανομής Waring. Με ανάλογο τρόπο αποδεικνύονται και τα υπόλοιπα συμπεράσματα.

Π Ι Ν Α Κ Α Σ 6.2.2

ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ ΚΑΙ Η SIZE-BIASED ΜΟΡΦΗ ΤΟΥΣ

Αρχική τυχαία μεταβλητή	*Συνάρτηση Πιθανότητας*	***Size-Biased** Μορφή (τάξης 1)*
Διωνυνική		
$B(x\|n,p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$B(x-1\|n-1,p)$
Αρνητική Διωνυμική		
$NB(x\|k,p)$	$\binom{k+x-1}{x} p^k (1-p)^x$	$NB(x-1\|k+1,p)$
Poisson		
$P(x\|\lambda)$	$\frac{e^{-\lambda}\,\lambda^x}{x!}$	$P(x-1\|\lambda)$
Υπεργεωμετρική		
$H(x\|n,M,N)$	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$H(x-1\|n-1,M-1,N-1)$
Waring		
$UGWD(x\|a,k;\rho)$	$\frac{\rho_{(k)}}{(a+\rho)_{(k)}} \frac{a_{(x)}\,k_{(x)}}{(a+k+\rho)_{(x)}} \frac{1}{x!}$	$UGWD(x-1\|a+1,k+1;\rho-1)$
Yule		
$Yule(x\|\rho)$	$\frac{\rho\, x!}{(\rho+1)_{(x+1)}}$	$UGWD(x-1\|2,2;\rho-1)$

Π Ι Ν Α Κ Α Σ 6.2.3

ΣΥΝΕΧΕΙΣ ΚΑΤΑΝΟΜΕΣ ΚΑΙ Η SIZE-BIASED ΜΟΡΦΗ ΤΟΥΣ

Αρχική τυχαία μεταβλητή	*Συνάρτηση Πυκνότητας Πιθανότητας*	***Size-Biased Μορφή (τάξης 1)***
Γάμμα G(a,k)	$\frac{a^k x^{k-1}}{\Gamma(k)} e^{-ax}$	G(a,k+1)
Εκθετική E(α)	$\alpha e^{-\alpha x}$	E(α)
Beta I $B_1(\alpha,\beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$B_1(\alpha+1,\beta)$
Beta II $B_2(\alpha,\beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1+x)^{-(\alpha+\beta)}$	$B_2(\alpha+1,\beta-1)$
Pareto Pa(θ,α)	$\alpha \theta^{\alpha} x^{-(\alpha+1)}$	Pa(θ,α-1)
Lognormal $LN(\mu,\sigma^2)$	$\frac{(2\pi)^{-1/2}}{x\sigma} \exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right]$	$LN(\mu+\sigma^2,\sigma^2)$
Power Function PF(β,ρ)	$\rho \beta^{-\rho} x^{\rho-1}$	PF(β,ρ+1)

ΠΙΝΑΚΑΣ 6.2.4

ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ

ΚΑΙ

Η SIZE-BIASED ΜΟΡΦΗ ΤΟΥΣ ΠΑΡΑΓΟΝΤΙΚΗΣ ΤΑΞΗΣ j

Αρχική τυχαία μεταβλητή	*Συνάρτηση Πιθανότητας*	*Size-Biased Μορφή παραγοντικής τάξης j*
Διωνυνική		
$B(x\|n,p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$B(x-j\|n-j,p)$
Αρνητική Διωνυμική		
$NB(x\|k,p)$	$\binom{k+x-1}{x} p^k (1-p)^x$	$NB(x-j\|k+j,p)$
Poisson		
$P(x\|\lambda)$	$\frac{e^{-\lambda}\,\lambda^x}{x!}$	$P(x-j\|\lambda)$
Υπερχεωμετρική		
$H(x\|n,M,N)$	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$H(x-j\|n-j,M-j,N-j)$
Waring		
$UGWD(x\|a,k;\rho)$	$\frac{\rho_{(k)}}{(a+\rho)_{(k)}} \frac{a_{(x)}\,k_{(x)}}{(a+k+\rho)_{(x)}} \frac{1}{x!}$	$UGWD(x-j\|a+j,k+j;\rho-j)$
Yule		
$Yule(x\|\rho)$	$\frac{\rho\, x!}{(\rho+1)_{(x+1)}}$	$UGWD(x-j\|2,2;\rho-j)$

Θεώρημα 6.2.2 Εστω οι τυχαίες μεταβλητές X και $X^{*(j)}$ οι οποίες παίρνουν ακέραιες τιμές. Ας υποθέσουμε επίσης ότι $E\left(X^{(j)}\right) < \infty$ και ότι οι συναρτήσεις πιθανότητας των X και $X^{*(j)}$ είναι p και $q^{*(j)}$ αντίστοιχα και ικανοποιούν τη σχέση

$$q_x^{*(j)} = \frac{x^{(j)}\, p_x}{E\left[X^{(j)}\right]}, \quad x = j, j+1, \ldots .$$

Τότε η κατανομή της $X^{*(j)}$ είναι η μετατοπισμένη κατά j μονάδες προς τα δεξιά μονοδιάστατη κατανομή Waring $(\alpha+j, \kappa+j; \rho-j)$ αν και μόνον αν η κατανομή της X είναι μία παραλλαγμένη μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha, \kappa; \rho)$, με συνάρτηση πιθανότητας:

$$p_x = \begin{cases} d_x & x = 0, 1, 2, \ldots, j-1 \\ C_j \dfrac{\rho_{(\kappa)}}{(\alpha+\rho)_{(\kappa)}} \dfrac{\alpha_{(x)}\, \kappa_{(x)}}{(\alpha+\kappa+\rho)_{(x)}} \dfrac{1}{x!}, & x = j, j+1, \ldots \end{cases} \tag{6.2.4}$$

όπου C_j και d_x είναι αυθαίρετες σταθερές τέτοιες ώστε $\sum_{x=0}^{\infty} p_x = 1$.

Απόδειξη

Ευθύ: Εστω ότι η τυχαία μεταβλητή X ακολουθεί την παραλλαγμένη μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha, \kappa; \rho)$ με συνάρτηση πιθανότητας την (6.2.4).

Προφανώς,

και $\mu^{(j)} = E\left[X^{(j)}\right] = \frac{\alpha_{(j)}\,\kappa_{(j)}}{(\rho-1)(\rho-2)\ldots(\rho-j)}$. $j=0,1,2,\ldots$

Επειδή δε,

$$x^{(j)} = x(x-1)\ldots(x-j+1) = \frac{x!}{(x-j)!}\ , \quad x = j,\ j+1,\ldots$$

συνεπάγεται ότι

$$q_x^{*(j)} = \frac{\frac{x!}{(x-j)!}\ \frac{\rho_{(\kappa)}}{(\alpha+\rho)_{(\kappa)}}\ \frac{\alpha_{(x)}\,\kappa_{(x)}}{(\alpha+\kappa+\rho)_{(x)}}\ \frac{1}{x!}}{\frac{\alpha_{(j)}\,\kappa_{(j)}}{(\rho-1)(\rho-2)\ldots(\rho-j)}}$$

$$= \frac{(\rho-j)_{(j)}\ \rho_{(\kappa)}\ \alpha_{(x)}\,\kappa_{(x)}}{(\alpha+\rho)_{(\kappa)}\ \alpha_{(j)}\kappa_{(j)}(\alpha+\kappa+\rho)_{(x)}}\ \frac{1}{(x-j)!}.$$

Αλλά

$$\frac{\alpha_{(x)}}{\alpha_{(j)}} = (\alpha+j)_{(x-j)}\ , \qquad \frac{\kappa_{(x)}}{\kappa_{(j)}} = (\kappa+j)_{(x-j)} \tag{6.2.5}$$

και $(\alpha+\kappa+\rho)_{(x)} = (\alpha+\kappa+\rho+j)_{(x-j)}\ (\alpha+\kappa+\rho)_{(j)}$. (6.2.6)

Τότε η τελευταία σχέση γίνεται

$$q^{*(j)} = \frac{(\rho-j)_{(j)}\ \rho_{(\kappa)}}{(\alpha+\rho)_{(\kappa)}}\ \frac{(\alpha+j)_{(x-j)}\ (\kappa+j)_{(x-j)}}{(\alpha+\kappa+\rho)_{(j)}\ (\alpha+\kappa+\rho+j)_{(x-j)}}\ \frac{1}{(x-j)!}\ . \qquad (6.2.7)$$

$$x = j,\ j+1,\ \ldots$$

Αλλά,

$$(\rho-j)_{(j)}\ \rho_{(\kappa)} = (\rho-j)_{(\kappa+j)} \qquad (6.2.8)$$

και

$$(\alpha+\rho)_{(j)}\ (\alpha+\kappa+\rho)_{(j)} = (\alpha+\rho)_{(\kappa+j)}. \qquad (6.2.9)$$

Επομένως η σχέση (6.2.5) τελικά γράφεται,

$$q^{*(j)} = \frac{(\rho-j)_{(\kappa+j)}}{(\alpha+\rho)_{(\kappa+j)}}\ \frac{(\alpha+j)_{(x-j)}\ (\kappa+j)_{(x-j)}}{(\alpha+\kappa+\rho+j)_{(x-j)}}\ \frac{1}{(x-j)!}\ .\quad x = j,\ j+1,\ldots$$

Δηλαδή η τυχαία μεταβλητή Z_1 ακολουθεί τη μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha+j,\kappa+j;\rho-j)$ μετατοπισμένη j μονάδες προς τα δεξιά.

Αντίστροφο: Εστω ότι η τυχαία μεταβλητή Z_1 ακολουθεί τη μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha+j,\kappa+j;\rho-j)$ μετατοπισμένη j μονάδες προς τα δεξιά, δηλαδή,

$$q^{*(j)} = \frac{(\rho-j)_{(\kappa+j)}}{(\alpha+\rho)_{(\kappa+j)}} \frac{(\alpha+j)_{(x-j)}\ (\kappa+j)_{(x-j)}}{(\alpha+\kappa+\rho+j)_{(x-j)}} \frac{1}{(x-j)!}\ ,\quad x = j,\ j+1,\ldots$$

Ισχύει όμως ότι,

$$q_x^{*(j)} = \frac{x^{(j)}\ p_x}{E\left[X^{(j)}\right]}\ ,\ x = j,\ j+1,\ldots$$

Επομένως,

$$\frac{x^{(j)}\ p_x}{E\left[X^{(j)}\right]} = \frac{(\rho-j)_{(\kappa+j)}}{(\alpha+\rho)_{(\kappa+j)}} \frac{(\alpha+j)_{(x-j)}\ (\kappa+j)_{(x-j)}}{(\alpha+\kappa+\rho+j)_{(x-j)}} \frac{1}{(x-j)!}\ .$$

Αρα,

$$p_x = \frac{E\left[X^{(j)}\right]}{x^{(j)}} \frac{(\rho-j)_{(\kappa+j)}}{(\alpha+\rho)_{(\kappa+j)}} \frac{(\alpha+j)_{(x-j)}\ (\kappa+j)_{(x-j)}}{(\alpha+\kappa+\rho+j)_{(x-j)}} \frac{1}{(x-j)!}$$

$$= \frac{\mu^{(j)}}{\frac{x!}{(x-j)!}} \frac{(\rho-j)_{(\kappa+j)}}{(\alpha+\rho)_{(\kappa+j)}} \frac{(\alpha+j)_{(x-j)}\ (\kappa+j)_{(x-j)}}{(\alpha+\kappa+\rho+j)_{(x-j)}} \frac{1}{(x-j)!}\ . \qquad (6.2.10)$$

Αντικαθιστώντας τις σχέσεις (6.2.8) και (6.2.9) στη σχέση (6.2.10) προκύπτει ότι,

$$p_x = \mu^{(j)} \frac{(\rho-j)_{(j)}\ \rho_{(\kappa)}}{(\alpha+\rho)_{(\kappa)}\ (\alpha+\kappa+\rho)_{(j)}} \frac{(\alpha+j)_{(x-j)}\ (\kappa+j)_{(x-j)}}{(\alpha+\kappa+\rho+j)_{(x-j)}} \frac{1}{x!}\ . \qquad (6.2.11)$$

$$x = j,\ j+1,\ldots$$

Λόγω των σχέσεων (6.2.5) και (6.2.6) η σχέση (6.2.11) γράφεται,

$$p_x = \mu^{(j)} \frac{(\rho-j)_{(j)}\, \rho_{(\kappa)}\; \alpha_{(x)}\, \kappa_{(x)}}{(\alpha+\rho)_{(\kappa)}\, \alpha_{(j)} \kappa_{(j)} (\alpha+\kappa+\rho)_{(x)}} \frac{1}{x!}, \quad x = j, j+1, \ldots$$

Επομένως η τυχαία μεταβλητή X ακολουθεί μια παραλλαγμένη μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha,\kappa;\rho)$, $\alpha,\kappa > 0$, $\rho > 0$, με συνάρτηση πιθανότητας της μορφής (6.2.4) όπου

$$C_j = \mu^{(j)}(\rho-j)_{(j)}/\alpha_{(j)}\, \kappa_{(j)}.$$

Θα ολοκληρώσουμε την ενότητα με αναφορά στην επίδραση του size-biased sampling στα μίγματα κατανομών. Θα αναφέρουμε παραδείγματα τόσο συνεχών όσο και διακριτών κατανομών. Τα συμπεράσματα προκύπτουν με βάση τα στοιχεία των πινάκων 6.2.2 και τα 6.2.3.

<u>Παράδειγμα 1</u> Εστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Poisson $P(\lambda)$ και έστω ότι η παράμετρός της λ είναι και αυτή τυχαία μεταβλητή με κατανομή Gamma. Τότε $(X\hat{}\lambda)^* \stackrel{d}{=} (X^*\hat{}\lambda^*)$.

<u>Παράδειγμα 2</u> Εστω ότι η τυχαία μεταβλητή X ακολουθεί την υπεργεωμετρική κατανομή H(n,M,N) και έστω ότι η παράμετρός της M είναι και αυτή τυχαία μεταβλητή που ακολουθεί τη διωνυμική κατανομή B(N,p). Τότε $(X\hat{}M)^* \stackrel{d}{=} (X^*\hat{}M^*)$.

<u>Παράδειγμα 3</u> Εστω ότι η τυχαία μεταβλητή X ακολουθεί την αρνητική διωνυμική κατανομή NB(k,q) όπου q = 1/(1+P) και έστω ότι

η παράμετρος P είναι και αυτή τυχαία μεταβλητή με κατανομή Beta II. Τότε $(X \hat{} P)^* \stackrel{d}{=} (X^* \hat{} P^*)$.

<u>Παράδειγμα 4</u> Εστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Gamma $(\alpha,1)$ και έστω ότι η παράμετρος α είναι και αυτή τυχαία μεταβλητή που ακολουθεί την Gamma $(1,\kappa)$. Τότε $(X \hat{} \alpha)^w \stackrel{d}{=} (X^w \hat{} \alpha^w)$ με $w(x) = x$ και $w(\alpha) = \alpha^{-1}$.

6.3 ΜΟΝΟΔΙΑΣΤΑΤΑ ΣΥΣΤΗΜΑΤΑ ΚΑΤΑΝΟΜΩΝ

Στην ενότητα αυτή εξετάζεται για ποιά γεννήτορα κατανομή το σύστημα των σταθμισμένων κατανομών και τα συστήματα των κατανομών των βασισμένων σε μερικά αθροίσματα οδηγούν στην ίδια κατανομή. Η παράγραφος 6.3.1 απαντά στο ερώτημα για την περίπτωση που η γεννήτωρ κατανομή είναι διακριτή ενώ η παράγραφος 6.3.2 αντιμετωπίζει τη συνεχή περίπτωση. Ετσι, θεμελιώνονται χαρακτηριστικές ιδιότητες για γνωστές μονοδιάστατες κατανομές όπως η διακριτή ομοιόμορφη, η Yule, η Pareto, η power function που θα προκύψουν από την ισότητα των συστημάτων κατανομών.

6.3.1 Α' Περίπτωση: Διακριτή Γεννήτωρ Κατανομή

Από μια δοσμένη γεννήτορα κατανομή με συνάρτηση πιθανότητας $p_x \equiv P(X=x)$ μπορούν να παραχθούν κατανομές με ένα από τους παρακάτω τρεις τρόπους:

Ορισμός 6.3.1 $q'_r \equiv P(Z_1=r) = \dfrac{w(r)p_r}{E[w(X)]}$ $\qquad r=1,2,\ldots$

$$\left(\text{με την προυπόθεση ότι } E[w(X)] < \infty\right)$$

Ορισμός 6.3.2 $q^+_r \equiv P(Z_2=r) = C_1 \dfrac{\sum_{j=r+1}^{\infty} p_j}{E(X)}$ $\qquad r=0,1,\ldots$

$$\left(\text{με την προυπόθεση ότι } E(X) < \infty\right)$$

Ορισμός 6.3.3 $q^*_r \equiv P(Z_3=r) = C_2 \sum_{j=r+1}^{\infty} \dfrac{p_j}{j}$ $\qquad r=0,1,\ldots$

Με βάση τους τρεις αυτούς ορισμούς θα διατυπώσουμε χαρακτηριστικές ιδιότητες για γνωστές διακριτές κατανομές. Παρατηρώντας τον ορισμό 6.3.1 διαπιστώνουμε ότι η συνάρτηση πιθανότητας της τυχαίας μεταβλητής Z_1 συμπίπτει με εκείνην της σχέσης (6.2.1) κατα συνέπεια $Z_1 \equiv X^w$. Χαρακτηριστικές ιδιότητες για γνωστές διακριτές κατανομές όπως η διωνυμική, αρνητική διωνυμική, υπεργεωμετρική μπορούν να αποδειχθούν με βάση τον ορισμό της σταθμισμένης κατανομής. Τα συμπεράσματα αυτά είναι συγκεντρωμένα στον Πίνακα 6.2.2. Στη συνέχεια θα αποδείξουμε μια χαρακτηριστική ιδιότητα που δεν συναντιέται στη βιβλιογραφία και η οποία αφορά τη μονοδιάστατη γενικευμένη κατανομή Waring. Σαν πόρισμα θα προκύψει η περίπτωση της κατανομής Yule.

Θεώρημα 6.3.1 Εστω X, Z_1 δύο τυχαίες μεταβλητές με τιμές στο σύνολο {1,2,3,...}. Εστω ότι $E(X) < \infty$ και έστω ότι οι συναρτήσεις πιθανότητας των X και Z_1 ικανοποιούν τη σχέση

$$q'_r \equiv P(Z_1=r) = \frac{rp_r}{E(X)} \quad , \quad r = 1,2,\ldots$$

όπου $q'_r = P(Z_1=r)$, $p_r = P(X=r)$.

Τότε η κατανομή της Z_1 είναι η μετατοπισμένη κατά μία μονάδα δεξιά μονοδιάστατη κατανομή Waring $(\alpha+1, \kappa+1; \rho-1)$ τότε και μόνο τότε αν η κατανομή της X είναι η μονοδιάστατη γενικευμένη κατανομή Waring $(\alpha, \kappa; \rho)$ περικομένη στο 0.

<u>Απόδειξη</u> Από το θεώρημα 6.2.2 προκύπτει ότι Z_1 ~ μετατοπισμένη μονοδιάστατη Waring $(\alpha+1, \kappa+1; \rho-1)$ ⇔ αν η X ακολουθεί την παραλλαγμένη κατανομή Waring με συνάρτηση πιθανότητας την 6.2.4 για $d_0 = 0$, $j=1$, $\kappa=1$.

Δηλαδή,

$$p_x = C_1 \rho \frac{\alpha_{(x)}}{(\alpha+\rho)_{(x+1)}} \quad , \quad x = 1,2,\ldots .$$

Αλλά $\sum_{x=1}^{\infty} p_x = 1.$

Αρα,

$$C_1\rho \sum_{x=1}^{\infty} \frac{\alpha_{(x)}}{(\alpha+\rho)_{(x+1)}} = 1 \quad \Leftrightarrow$$

$$C_1 \frac{\rho}{\alpha+\rho} \sum_{x=0}^{\infty} \frac{\alpha_{(x+1)}}{(\alpha+\rho+1)_{(x+1)}} = 1 \quad \Leftrightarrow$$

$$C_1 \frac{\rho}{\alpha+\rho} \frac{\alpha}{\alpha+\rho+1} \sum_{x=0}^{\infty} \frac{(\alpha+1)_{(x)}\, 1_{(x)}}{(\alpha+\rho+2)_{(x)}} \frac{1}{x!} = 1. \quad \Leftrightarrow$$

$$C_1 \frac{\rho}{\alpha+\rho} \frac{\alpha}{\alpha+\rho+1} \frac{\alpha+\rho+1}{\alpha} = 1 \quad \Leftrightarrow$$

$$C_1 = \frac{\alpha+\rho}{\alpha} .$$

Κατά συνέπεια

$$p_x = \frac{\rho}{\alpha} \frac{\alpha_{(x)}}{(\alpha+\rho+1)_{(x)}} , \quad x = 1,2,\ldots .$$

Αυτή όμως είναι η συνάρτηση πιθανότητας της 0-περικομένης κατανομής Waring (α,ρ). Αρα το θεώρημα απεδείχθη.

Στην ειδική περίπτωση που $\alpha=1$ προκύπτει το πόρισμα που ακολουθεί.

<u>Πόρισμα 6.3.1</u> Εστω X, Z_1 δύο τυχαίες μεταβλητές με τιμές στο σύνολο $\{1,2,3,\ldots\}$. Εστω ότι $E(X) < \infty$ και έστω ότι οι συναρτήσεις πιθανότητας των X και Z_1 ικανοποιούν τη σχέση

$$q'_r \equiv P(Z_1=r) = \frac{r p_r}{E(X)} , \quad r = 1,2,\ldots$$

όπου $q'_r = P(Z_1=r)$, $p_r = P(X=r)$.

Τότε η κατανομή της Z_1 είναι η μετατοπισμένη κατά μία μονάδα δεξιά μονοδιάστατη κατανομή Waring $(2, 2; \rho-1)$ τότε και μόνο τότε αν η κατανομή της X είναι η μονοδιάστατη Yule (ρ)

Απόδειξη Από το θεώρημα 6.3.1 για $\alpha = 1$ προκύπτει το συμπέρασμα του πορίσματος.

Πόρισμα 6.3.2 Εστω οι τυχαίες μεταβλητές X και Z_1 οι οποίες παίρνουν ακέραιες και μη αρνητικές τιμές . Ας υποθέσουμε επίσης ότι οι συναρτήσεις πιθανότητας των X και Z_1 είναι p και q' αντίστοιχα και ικανοποιούν τη συνθήκη:

$$q'_r \equiv P(Z_1=r) = \frac{rp_r}{E(X)}$$

Τότε η κατανομή της Z_1 είναι η μετατοπισμένη μονοδιαστατη γενικευμένη κατανομή Waring $(1,2;\rho)$ αν και μόνον αν η κατανομή της X είναι η μετατοπισμένη μονοδιάστατη Yule $(\rho+1)$.

Απόδειξη

Ευθυ: Εστω ότι η X ακολουθεί τη μετατοπισμένη μονοδιάστατη Yule $(\rho+1)$, δηλαδή

$$p_x = \frac{(\rho+1)\ (x-1)!}{(\rho+2)_{(x)}} \qquad x = 1,2,\ldots$$

και $E(X) = \frac{\rho+1}{\rho}$.

Τότε,

$$q'_r = P(Z_1=r) = \frac{r\,\frac{(\rho+1)\,(r-1)!}{(\rho+2)_{(r)}}}{\frac{\rho+1}{\rho}}$$

$$= \frac{\rho(\rho+1)r!}{(\rho+2)_{(r)}}$$

$$= \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}} .$$

Δηλαδή, η τυχαία μεταβλητή Z_1 ακολουθεί τη μετατοπισμένη γενικευμένη μονοδιάστατη κατανομή Waring $(1,2;\rho)$.

Αντίστροφο: Εστω ότι η Z_1 ακολουθεί τη μετατοπισμένη γενικευμένη μονοδιάστατη κατανομή Waring $(1,2;\rho)$, $\rho>0$. Εστω επίσης ότι

$$q'_r = \frac{rp_r}{\mu} , \quad r=1,2,\ldots .$$

Τότε,

$$\frac{rp_r}{\mu} = \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}} ,$$

δηλαδή,

$$p_r = \mu \frac{\rho}{\rho+2} \frac{1_{(r-1)}}{(\rho+3)_{(r-1)}} .$$

Αλλά,

$$\sum_{r=1}^{\infty} p_r = 1.$$

Επομένως,

$$\mu \frac{\rho}{\rho+2} \sum_{r=1}^{\infty} \frac{1_{(r-1)}}{(\rho+3)_{(r-1)}} = 1.$$

Δηλαδή,

$$\mu \frac{\rho}{\rho+2} \sum_{r=0}^{\infty} \frac{1_{(r)} 1_{(r)}}{(\rho+3)_{(r)}} \frac{1}{r!} = 1.$$

Από τη σχέση αυτή τελικά προκύπτει ότι

$$\mu = \frac{\rho+1}{\rho} .$$

Αρα,

$$p_r = \frac{(\rho+1)\ (r-1)!}{(\rho+2)_{(r)}} .$$

Δηλαδή, η X ακολουθεί τη μετατοπισμένη μονοδιάστατη Yule $(\rho+1)$.

Με βάση τον ορισμό 6.3.2 μπορούμε να οδηγηθούμε σε χαρακτηρισμό της μετατοπισμένης μονοδιάστατης κατανομής Yule $(\rho+1)$. Συγκεκριμένα αποδεικνύεται ότι ισχύει το παρακάτω θεώρημα 6.3.2.

Θεώρημα 6.3.3 Εστω οι τυχαίες μεταβλητές X και Z_2 οι οποίες παίρνουν ακέραιες και μη αρνητικές τιμές. Υποθέτουμε επίσης ότι $E(X)<\infty$ και οτι οι συναρτήσεις πιθανότητας των X και Z_2, p και q^+ αντίστοιχα, ικανοποιούν την παρακάτω συνθήκη:

$$q_r^+ = P(Z_2=r) = \frac{\sum_{j=r+1}^{\infty} p_j}{E(X)} \quad , \quad r=0,1,2,\ldots$$

Τότε η κατανομή της Z_2 περικομένη στο μηδέν είναι η μετατοπισμένη γενικευμένη μονοδιάστατη Waring κατανομή $(1,2;\rho)$, $\rho>0$, αν και μόνον αν η κατανομή της X είναι η μετατοπισμένη Yule $(\rho+1)$, $\rho>0$.

Απόδειξη

Ευθύ. Εστω ότι η τυχαία μεταβλητή X ακολουθεί την μετατοπισμένη Yule $(\rho+1)$ με συνάρτηση πιθανότητας την

$$p_x = \frac{(\rho+1)(x-1)!}{(\rho+2)_{(x)}} \qquad x=1,2,\ldots$$

και $$E(X) = \frac{\rho+1}{\rho}$$

Τότε η κατανομή της τυχαίας μεταβλητής Z_2 περικομένη στο μηδέν έχει συνάρτηση πιθανότητας

$$P(Z_2=r|Z_2>0) = \frac{q_r^+}{1-q_0^+} = \frac{\left(\sum_{j=r+1}^{\infty} p_j\right)\Big/E(X)}{1-\left(\sum_{j=1}^{\infty} p_j\right)\Big/E(X)} = \frac{\sum_{j=r+1}^{\infty} p_j}{E(X)-1} .$$

Εχει αποδειχθεί όμως (Xekalaki (1984)) ότι

$$\sum_{j=r+1}^{\infty} p_j = P(X > r) = \frac{r}{\rho+1} P(X = r).$$

Αρα,

$$\frac{q_r^+}{1-q_0^+} = \frac{\frac{r}{\rho+1} \frac{(\rho+1)(r-1)!}{(\rho+2)_{(r)}}}{\frac{\rho+1}{\rho} - 1} = \frac{\rho\, r!}{(\rho+2)_{(r)}} = \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+1)_{(r-1)}}.$$

Δηλαδή, η κατανομή της Z_2 περικομένη στο 0 είναι η μετατοπισμένη γενικευμένη μονοδιάστατη κατανομή Waring $(1,2;\rho)$, $\rho>0$.

Αντίστροφο. Εστω ότι η τυχαία μεταβλητή Z_2 περικομένη από τα αριστερά στο μηδέν ακολουθεί την μετατοπισμένη γενικευμένη μονοδιάστατη κατανομή Waring $(1,2;\rho)$, $\rho>0$. Δηλαδή,

$$P(Z_2=r \mid Z_2>0) = \frac{q_r^+}{1-q_0^+} = \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}}$$

Αλλά,

$$\frac{q_r^+}{1-q_0^+} = \frac{\sum_{j=r+1}^{\infty} p_j}{E(X) - 1}, \quad r=1,2,\ldots$$

Επομένως,

$$\frac{\sum_{j=r+1}^{\infty} p_j}{\mu-1} = \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}} \Rightarrow$$

$$\sum_{j=r+1}^{\infty} p_j = (\mu-1) \frac{\rho_{(2)} \quad 2_{(r-1)}}{(\rho+1)_{(2)}(\rho+3)_{(r-1)}} . \qquad (6.3.1)$$

Εξειδικεύοντας την παραπάνω σχέση για r+1 και αφαιρώντας την προκύπτουσα σχέση από την (6.3.1) κατά μέλη , προκύπτει ότι

$$p_{r+1} = (\mu-1) \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}} - (\mu-1) \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r)}}{(\rho+3)_{(r)}} ,$$

δηλαδή,

$$p_r = (\mu-1) \frac{\rho(\rho+1)}{(\rho+1)(\rho+2)} \left[\frac{2_{(r-2)}}{(\rho+3)_{(r-2)}} - \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}} \right], \quad r=1,2,\ldots .$$

Επειδή δε, $\sum_{r=1}^{\infty} p_r = 1$ συνεπάγεται ότι:

$$(\mu-1) \frac{\rho(\rho+1)}{(\rho+1)(\rho+2)} \sum_{r=1}^{\infty} \left[\frac{1_{(r-1)}}{\frac{(\rho+2)_{(r-1)}}{\rho+2}} - \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}} \right] = 1 \Rightarrow$$

$$(\mu-1)\ \frac{\rho(\rho+1)}{(\rho+1)(\rho+2)}\ \left[(\rho+2)\sum_{r=1}^{\infty}\frac{1_{(r-1)}}{(\rho+2)_{(r-1)}}-\sum_{r=1}^{\infty}\frac{2_{(r-1)}}{(\rho+3)_{(r-1)}}\right]=1\ \Rightarrow$$

$$(\mu-1)\ \frac{\rho(\rho+1)}{(\rho+1)(\rho+2)}\ \left[(\rho+2)\ \frac{\Gamma(\rho+2)\Gamma(\rho)}{\Gamma(\rho+1)\Gamma(\rho+1)}-\frac{\Gamma(\rho+3)\Gamma(\rho)}{\Gamma(\rho+2)\Gamma(\rho+1)}\right]=1\ \Rightarrow$$

$$(\mu-1)\ \frac{\rho(\rho+1)}{(\rho+1)(\rho+2)}\ \left[(\rho+2)\ \frac{(\rho+1)}{\rho}-\frac{\rho+2}{\rho}\right]=1\ \Rightarrow$$

$$(\mu-1)\rho=1\ \Rightarrow\ \mu=\frac{\rho+1}{\rho}$$

Αρα,

$$p_r=\frac{1}{\rho}\left[\frac{\rho_{(2)}}{(\rho+1)_{(2)}}\ \frac{2_{(r-2)}}{(\rho+3)_{(r-2)}}-\frac{\rho_{(2)}}{(\rho+1)_{(2)}}\ \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}}\right]$$

$$=\frac{(r-1)!}{(\rho+2)_{(r-1)}}-\frac{r!}{(\rho+2)_{(r)}}$$

$$=(r-1)!\ \frac{(\rho+1)\Gamma(\rho+2)}{\cdot\ \Gamma(\rho+r+2)}$$

$$=\frac{(\rho+1)(r-1)!}{(\rho+2)_{(r)}}\ ,\qquad r=1,2,\ldots\ .$$

Δηλαδή, η τυχαία μεταβλητή X ακολουθεί μετατοπισμένη κατανομή Yule $(\rho+1)$, $\rho>0$.

Το συμπέρασμα που ακολουθεί αποτελεί χαρακτηρισμό της μετατοπισμένης μονοδιάστατης κατανομής Yule (ρ+1) και βασίζεται στην ισότητα των κατανομών που προκύπτουν από τους ορισμούς 6.3.1 και 6.3.2.

Θεώρημα 6.3.3 Εστω η τυχαία μεταβλητή X με μη εκφυλισμένη συνάρτηση πιθανότητας $p_x = P(X=x)$, $x = 1,2,...$ και $E(X) = \mu < \infty$. Εστω επίσης οι τυχαίες μεταβλητές Z_1 και Z_2 των οποίων οι συναρτήσεις πιθανότητας είναι αντίστοιχα οι:

$$q'_r \equiv P(Z_1=r) = \frac{rp_r}{E(X)} \qquad r=1,2,...$$

$$q^+_r \equiv P(Z_2=r) = \frac{\sum_{j=r+1}^{\infty} p_j}{E(X)} \qquad r=0,1,2,...$$

Τότε οι τυχαίες μεταβλητές Z_1 και $Z_2|(Z_2>0)$ ακολουθούν την ίδια κατανομή αν και μόνον αν η τυχαία μεταβλητή X ακολουθεί την μετατοπισμένη κατά μία μονάδα προς τα δεξιά Yule (ρ), $\rho>0$.

Απόδειξη

Ευθύ: Εστω ότι η τυχαία μεταβλητή X ακολουθεί μετατοπισμένη Yule (ρ+1). Τότε λογω των θεωρημάτων 6.1.1 και 6.1.2 έχουμε αντίστοιχα ότι:

$$q'_r = \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}}$$

και

$$P(Z_2=r \mid Z_2>0) = \frac{q_r^+}{1-q_0^+} = \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \frac{2_{(r-1)}}{(\rho+3)_{(r-1)}} .$$

Επομένως $Z_1 \overset{d}{=} Z_2 \mid (Z_2>0)$.

Αντίστροφο: Εστω ότι $q'_r = \frac{q_r^+}{1-q_0^+}$,

με $q_0^+ \equiv P(Z_2=0) = \frac{\sum_{j=1}^{\infty} p_j}{E(X)} = \frac{1}{\mu}$,

δηλαδή, έστω ότι $q'_r = \frac{\mu}{\mu-1} q_r^+$.

Αλλά, $q'_r = \frac{rp_r}{\mu}$

και

$$q_r^+ = \frac{\sum_{j=r+1}^{\infty} p_j}{\mu} .$$

Επομένως,

$$\frac{rp_r}{\mu} = \frac{\mu}{\mu-1} \sum_{j=r+1}^{\infty} p_j \quad \Rightarrow$$

$$\frac{\mu-1}{\mu}\, r\, p_r = \sum_{j=r+1}^{\infty} p_j \ . \qquad (6.3.2)$$

Θέτοντας όπου r το r+1 , και αφαιρώντας την προκύπτουσα σχέση από την (7.3.2) κατά μέλη προκύπτει ότι

$$\frac{\mu-1}{\mu}\, r\, p_r - \frac{\mu-1}{\mu}\,(r+1)\, p_{r+1} = p_{r+1} \ .$$

Προφανώς $\mu > 1$. Διαφορετικά, εφόσον $X \geq 1$ θα ισχυε οτι $\sum_{x=1}^{\infty} x p_x = 1 = \sum_{x=1}^{\infty} p_x \Leftrightarrow \sum_{x=2}^{\infty} (x-1)p_x \Leftrightarrow p_x = 0,\ x=2,3,\ldots \Leftrightarrow P(X=1) = 1$. Αυτό όμως αντίκειται στις υποθέσεις του θεωρήματος. Συνεπώς από την παραπάνω σχέση έχουμε

$$p_{r+1} - \frac{r}{r + \frac{2\mu-1}{\mu-1}}\, p_r = 0,$$

Οπότε,

$$p_r = p_1 \prod_{i=1}^{r-1} \frac{i}{\frac{2\mu-1}{\mu-1} + i}$$

$$= p_1 \frac{(r-1)!}{\left(\frac{2\mu-1}{\mu-1} + 1\right)_{(r-1)}}$$

$$= p_1 \frac{(r-1)!}{\left(\frac{3\mu-2}{\mu-1}\right)_{(r-1)}} .$$

Ομως, $\sum_{r=1}^{\infty} p_r = 1.$

Αρα,

$$p_1 \sum_{r=1}^{\infty} \frac{(r-1)!}{\left(\frac{3\mu-2}{\mu-1}\right)_{(r-1)}} = 1 \quad \Leftrightarrow$$

$$p_1 \sum_{r=0}^{\infty} \frac{1_{(r)}\, 1'_{(r)}}{c_{(r)}} \frac{1}{r!} = 1 \qquad \left(\text{όπου} \quad c = \frac{3\mu-2}{\mu-1} \right) \Leftrightarrow$$

$$p_1 \frac{\Gamma(c)\Gamma(c-1)}{\Gamma(c-1)\Gamma(c-1)} = 1 \quad \Leftrightarrow$$

$$p_1 = \frac{c-2}{c-1}$$

ή

$$p_1 = \frac{\frac{3\mu-2}{\mu-1} - 2}{\frac{3\mu-2}{\mu-1} - 1} = \frac{\mu/(\mu-1)}{\left(\mu/(\mu-1)\right) + 1} .$$

Αρα τελικά,

$$p_r = \frac{\frac{\mu}{\mu-1}\,(r-1)!}{\left(\frac{\mu}{\mu-1} + 1\right)_{(r)}} , \quad r = 1,2,\ldots$$

Δηλαδή, η τυχαία μεταβλητή X ακολουθεί τη μετατοπισμένη κατανομή Yule $\left(\mu/(\mu-1)\right)$. Αρα το θεώρημα απεδείχθη.

Το θεώρημα που ακολουθεί διατυπώθηκε και αποδείχθηκε από την Xekalaki (1983a). Αποτελεί χαρακτηρισμό της μονοδιάστατης κατανομής Yule (ρ) και βασίζεται στον ορισμό 6.3.3.

<u>Θεώρημα 6.3.4</u> (Xekalaki,1983a) Εστω X τυχαία μεταβλητή με θετικές και ακέραιες τιμές της οποίας η συνάρτηση πιθανότητας είναι $p_r = P(X=r)$. Εστω επίσης, Z_3 μια άλλη τυχαία μεταβλητή με μη αρνητικές ακέραιες τιμές και συνάρτηση πιθανότητας q_r^* που ικανοποιεί τη συνθήκη:

$$q_r^* = P(Z_3=r) = \sum_{j=r+1}^{\infty} \frac{p_j}{j} .$$

Τότε η κατανομή της Z_3 περικομένη στο μηδέν συμπίπτει με την κατανομή της X αν και μόνον αν η X ακολουθεί κατανομή Yule.

Το συμπέρασμα αυτό χρησιμοποιήθηκε από την ίδια συγγραφέα (Xekalaki (1988b)) στο πλαίσιο ενός μοντέλου ανανέωσης αποθέματος προκειμένου, αφενός μεν να απλουστευθεί ο αλγόριθμος προσδιορισμού του "κρισίμου" αποθέματος, αφετέρου δε να δοθεί η δυνατότητα ακριβούς καθορισμού του κρισίμου αποθέματος.

6.3.2 Β' Περίπτωση: Συνεχής Γεννήτωρ Κατανομή

Στην ενότητα αυτή αντιμετωπίζεται η διαδικασία γένεσης συνεχών κατανομών από δοσμένη γεννήτορα κατανομή που έχει συνάρτηση πυκνότητας πιθανότητας $f_X(x)$. Η γένεση μπορεί να επιτευχθεί με έναν από τους παρακάτω τέσσερις τρόπους:

<u>Ορισμός 6.3.4</u> $$f_Y(y) = \frac{1}{y^2}\int_1^y x\, f_X(x)\, dx$$

<u>Ορισμός 6.3.5</u> $$f_{Z_1}(r) = \frac{w(r)f_X(r)}{E[w(X)]} \quad , \quad \left(E(w(X)) < \infty\right)$$

<u>Ορισμός 6.3.6</u> $$f_{Z_2}(r) = C_1\frac{\bar{F}_X(r)}{E(X)} \quad , \quad \left(E(X) < \infty\right)$$

<u>Ορισμός 6.3.7</u> $$f_{Z_3}(r) = C_2\int_r^\infty \frac{f_X(x)}{x}\, dx$$

Με βάση τους τέσσερις αυτούς ορισμούς θα διατυπώσουμε

χαρακτηριστικές ιδιότητες για γνωστές συνεχείς κατανομές. Ο πρώτος χαρακτηρισμός αφορά τη κατανομή power function και βασίζεται στον ορισμό 6.3.4. Ειδικότερα μπορούμε να αποδείξουμε το παρακάτω θεώρημα

Θεώρημα 6.3.5 Η τυχαία μεταβλητή Y με συνάρτηση πυκνότητας πιθανότητας $f_Y(y)$ που ορίζεται από την

$$f_Y(y) = \frac{1}{y^2} \int_0^y x\, f_X(x)\, dx$$

ακολουθεί την κατανομή power function $(\rho+1,\rho)$ αν και μόνον αν η τυχαία μεταβλητή X ακολουθεί την κατανομή power function $(1,\rho)$.

Απόδειξη

Ευθύ: Εστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή power function $(1,\rho)$, τότε

$$f_Y(y) = y^{-2} \int_0^y x\, \rho\, x^{\rho-1}\, dx$$

$$= y^{-2} \int_0^y \rho\, x^{\rho}\, dx$$

$$= \frac{1}{\rho+1}\, \rho\, y^{\rho-1}.$$

Δηλαδή, η τυχαία μεταβλητή Y ακολουθεί την κατανομή power function $(\rho+1,\rho)$.

Αντίστροφο: Εστω ότι,

$$f_Y(y) = \frac{1}{y^2} \int_0^y x\, f_X(x)\, dx$$

με

$$f_Y(y) = \frac{1}{\rho+1}\, \rho\, y^{\rho-1}.$$

Δηλαδή,

$$\frac{1}{\rho+1}\, \rho\, y^{\rho-1} = \frac{1}{y^2} \int_0^y x\, f_X(x)\, dx$$

ή

$$\frac{1}{\rho+1}\, y^{\rho+1} = \int_0^y x\, f_X(x)\, dx.$$

Παραγωγίζοντας ως προς y προκύπτει ότι,

$$\rho\, y^{\rho} = y\, f_X(y).$$

Δηλαδή,

$$f_X(y) = \rho\, y^{\rho-1}.$$

Αρα το θεώρημα απεδείχθη.

Παρατήρηση Στην ειδική περίπτωση που $\rho = 1$ το θεώρημα 6.3.4 οδηγεί σε χαρακτηρισμό της συνεχούς ομοιόμορφης κατανομής.

Παρατηρώντας τον ορισμό 6.3.5 διαπιστώνουμε ότι η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Z_1 συμπίπτει με εκείνην της σχέσης (6.2.1) κατά συνέπεια $Z_1 \equiv X^w$. Χαρακτηριστικές ιδιότητες για γνωστές συνεχείς κατανομές όπως η Γάμμα, εκθετική, Beta I, Beta II, Pareto, λογαριθμοκανονική μπορούν να αποδειχθούν με βάση τον ορισμό της σταθμισμένης κατανομής. Τα συμπεράσματα αυτά είναι συγκεντρωμένα στον Πίνακα 6.2.3. Στη συνέχεια ενδεικτικά θα αποδείξουμε τη σχετική χαρακτηριστική ιδιότητα που αφορά την κατανομή Pareto.

Θεώρημα 6.3.6 Εστω X τυχαία μεταβλητή με τιμές στο διάστημα $(1,+\infty)$ και συνάρτηση πυκνότητας πιθανότητας $f_X(x)$, $x>1$. Τότε η τυχαία μεταβλητή Z_1 με συνάρτηση πυκνότητας πιθανότητας $f_{Z_1}(r)$ που ορίζεται από την

$$f_{Z_1}(r) = \frac{w(r)f_X(r)}{E[w(X)]}, \quad r>1, \quad \text{με } w(r) = r^{\delta}, \ \delta < \alpha, \ \alpha > 0$$

ακολουθεί κατανομή Pareto $(1,\alpha-\delta)$, αν και μόνον αν η τυχαία μεταβλητή X ακολουθεί τη Pareto $(1,\alpha)$ κατανομή.

Απόδειξη

Ευθύ: Εστω ότι η τυχαία μεταβλητή ακολουθεί κατανομή Pareto$(1,\alpha)$ τότε,

$$f_{Z_1}(r) = \frac{r^{\delta} f_X(r)}{E(X^{\delta})}$$

$$= \frac{r^{\delta} \alpha\, r^{-(\alpha+1)}}{\int_1^{\infty} x^{\delta} \alpha\, x^{-(\alpha+1)} dx}$$

$$= (\alpha-\delta)\, r^{-(\alpha-\delta+1)} \quad , \; r > 1.$$

Δηλαδή, $Z_1 \sim$ Pareto $(1, \alpha-\delta)$.

Αντίστροφο: Εστω ότι

$$f_{Z_1}(r) = \frac{r^{\delta} f_X(r)}{C_1} \quad , \; r > 1.$$

και

$$f_{Z_1}(r) = (\alpha-\delta)\, r^{-(\alpha-\delta+1)}, \quad r > 1.$$

Τότε

$$f_X(r) = C_1\, r^{-\delta}\, (\alpha-\delta)\, r^{-(\alpha-\delta+1)}, \quad r > 1.$$

Αλλά,

$$\int_1^{\infty} f_X(r)\, dr = 1.$$

Κατά συνέπεια

$$C_1 \int_1^{\infty} (\alpha-\delta)\ r^{-(\alpha+1)}\ dr = 1$$

και επομένως,

$$C_1 = \alpha/(\alpha-\delta).$$

Αρα $f_X(r) = \alpha\ r^{-(\alpha+1)}$, $r > 1$ δηλαδή, $X \sim$ Pareto $(1,\alpha)$. Αρα το θεώρημα απεδείχθη.

Ειδικές περιπτώσεις.

α) Οταν $\delta = 1$ $Z_1 \sim$ Pareto $(1,\alpha-1)$ $\Leftrightarrow$ $X \sim$ Pareto $(1,\alpha)$

β) Οταν $\delta = -1$ $Z_1 \sim$ Pareto $(1,\alpha+1)$ $\Leftrightarrow$ $X \sim$ Pareto $(1,\alpha)$

Το θεώρημα που ακολουθεί αποτελεί γενίκευση του θεωρήματος 6.3.6 και οδηγεί σε χαρακτηρισμό της κατανομής Beta II (α,β).

Θεώρημα 6.3.7 Εστω X τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$, $x > 0$. Η τυχαία μεταβλητή Z_1 με συνάρτηση πυκνότητας πιθανότητας $f_{Z_1}(r)$ που ορίζεται από τη

$$f_{Z_1}(r) = \frac{w(r)f_X(r)}{E[w(X)]}, \quad \text{με } w(r) = r^{\delta},\ r > 0, \quad -\alpha < \delta < \beta,\ \alpha,\beta > 0$$

ακολουθεί την κατανομή Beta II $(\alpha+\delta,\beta-\delta)$, αν και μόνον αν η τυχαία μεταβλητή X ακολουθεί την κατανομή Beta II (α,β).

Απόδειξη Ομοια με εκείνην του θεωρήματος 7.3.7.

Θεώρημα 6.3.8 Εστω X τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$, $x>0$. Τότε η τυχαία μεταβλητή Z_1 με συνάρτηση πυκνότητας πιθανότητας $f_{Z_1}(r)$ που ορίζεται από την

$$f_{Z_1}(r) = \frac{w(r)f_X(r)}{E[w(X)]}, \qquad \text{με } w(r) = r^{\delta} \text{ , } r > 0, \quad \rho > -\delta, \ \rho > 0$$

ακολουθεί την κατανομή power function $(1,\rho+\delta)$, αν και μόνον αν η τυχαία μεταβλητη X ακολουθεί την κατανομή power $(1,\rho)$.

Απόδειξη Ομοια με εκείνην του θεωρήματος 6.3.6.

Ας έλθουμε τώρα στον ορισμό 6.3.6. Παρατηρούμε ότι είναι δυνατόν να οδηγηθούμε σε χαρακτηρισμό της κατανομής Pareto $(1,\alpha)$ όπως φαίνεται από το παρακάτω θεώρημα.

Θεώρημα 6.3.9 Εστω η τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$, $x > 1$ και $E(X) = \mu < \infty$. Η τυχαία μεταβλητή Z_2 με συνάρτηση πυκνότητας πιθανότητας $f_{Z_2}(r)$ που ορίζεται από την

$$f_{Z_2}(r) = C_1 \frac{\bar{F}_X(r)}{E(X)} \text{ , } r > 1.$$

ακολουθεί την κατανομή Pareto $(1,\alpha-1)$ με $C_1 = \alpha$ αν και μόνον αν η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $(1,\alpha)$.

Απόδειξη

Ευθύ. Εστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $(1,\alpha)$. Τότε,

$$f_{Z_2}(r) = C_1 \frac{\int_r^\infty \alpha\, x^{-(\alpha+1)} dx}{\alpha/(\alpha-1)}$$

$$= C_1 \frac{1}{\alpha} (\alpha-1)\, r^{-\alpha} \ , \ r > 1.$$

Αλλά, επειδή

$$\int_1^\infty f_{Z_2}(r)dr = 1$$

έπεται ότι $C_1 = \alpha$.

Επομένως,

$$f_{Z_2}(r) = (\alpha-1)\, r^{-\alpha} \ , \ r > 1$$

δηλαδή, η τυχαία μεταβλητή Z_2 ακολουθεί την κατανομή Pareto $(1,\alpha-1)$.

Αντίστροφο: Εστω ότι

$$f_{Z_2}(r) = \alpha \frac{\overline{F}_X(r)}{E(X)} \ , \ r > 1$$

και

$$f_{Z_2}(r) = (\alpha-1)\, r^{-\alpha} \ , \ r > 1.$$

Τότε,

$$\overline{F}_X(r) = \mu \frac{1}{\alpha} (\alpha-1)\, r^{-\alpha} \ , \quad -\infty < r < +\infty.$$

Επειδή όμως $\overline{F}_X(1) = 1$, συνεπάγεται ότι $\mu = \alpha/(\alpha-1)$.

Αρα, $\overline{F}_X(r) = r^{-\alpha}$, $r = 1$ δηλαδή η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $(1,\alpha)$, ο.ε.δ.

Το θεώρημα 6.3.10 που ακολουθεί οδηγεί σε χαρακτηρισμό της κατανομής Pareto ο οποίος βασίζεται στην ισότητα των κατανομών που προκύπτουν από τους ορισμούς 6.3.5 και 6.3.6.

<u>Θεώρημα 6.3.10</u> Εστω X τυχαία μεταβλητή με συνεχή συνάρτηση πυκνότητας πιθανότητας $f_X(x)$, $x \geq 1$ και $E(X) < \infty$. Ορίζουμε τις τυχαίες μεταβλητές Z_1 και Z_2 με τις σχέσεις:

$$f_{Z_1}(r) = \frac{r f_X(r)}{E(X)} \ , r > 1 \quad \text{και} \quad f_{Z_2}(r) = C_1 \frac{\overline{F}_X(r)}{E(X)} \ , r > 1 \quad \text{αντίστοιχα.}$$

Τότε, για $C_1 = \alpha$ οι κατανομές των τυχαίων μεταβλητών Z_1 και Z_2

συμπίπτουν αν και μόνον αν η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $(1,\alpha)$.

Απόδειξη

Ευθύ. Εστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $(1,\alpha)$. Τότε,

$$f_{Z_1}(r) = (\alpha-1)\, r^{-\alpha} \quad , \; r > 1$$

και

$$f_{Z_2}(r) = C_1 \frac{1}{\alpha} (\alpha-1)\, r^{-\alpha} \quad , \; r > 0.$$

Επομένως, για $C_1 = \alpha$ συνεπάγεται ότι $f_{Z_1}(r) = f_{Z_2}(r)$.

Αντίστροφο. Εστω ότι $f_{Z_1}(r) = f_{Z_2}(r)$ $r > 1$. Από τη σχέση αυτή συνεπάγεται ότι,

$$\frac{1}{\alpha}\, r\, f_X(r) = \int_r^{\infty} dF_X(x) \quad , \; r > 1 \qquad \Rightarrow$$

$$-\frac{1}{\alpha}\, r\, \bar{F}_X'(r) = \bar{F}_X(r), \; r > 1.$$

Επομένως,

$$\frac{\bar{F}_X'(r)}{\bar{F}_X(r)} = -\frac{\alpha}{r} .$$

Αρα

$$\int d \ln \bar{F}_X(r) = -\alpha \ln r + \ln C'.$$

Κατά συνέπεια,

$$\bar{F}_X(r) = C' r^{-\alpha} .$$

Αλλά, $\bar{F}_X(1) = 1$. Τότε $C' = 1$.

Επομένως, $\bar{F}_X(r) = r^{-\alpha}$.

Δηλαδή, η τυχαία μεταβλητή X ακολουθεί την κατανομή Pareto $(1,\alpha)$, ο.ε.δ.

Το συμπέρασμα που ακολουθεί αποτελεί γενίκευση του θεωρήματος 6.3.10 και αποδεικνύει ότι ισότητα μεταξύ των κατανομών που προκύπτουν από τους ορισμούς 6.3.1 και 6.3.2 οδηγεί σε μονοσήμαντο καθορισμό της κατανομής της τυχαίας μεταβλητής X

<u>Θεώρημα 6.3.11</u> Εστω X τυχαία μεταβλητή με συνεχή συνάρτηση πυκνότητας πιθανότητας $f_X(x)$, $x \geq \theta$, $\theta \geq 0$ και $E(X) < \infty$. Ορίζουμε τις τυχαίες μεταβλητές Z_1 και Z_2 με τις σχέσεις:

$$f_{Z_1}(r) = \frac{w(r) f_X(r)}{E[w(X)]}, \; r>\theta \text{ και } f_{Z_2}(r) = C_1 \frac{\bar{F}_X(r)}{E(X)}, \; r>\theta \qquad (6.3.3)$$

όπου η συνάρτηση w(.) είναι θετική και τέτοια ώστε $E[w(X)]<\infty$. Τότε η κατανομή της τυχαίας μεταβλητής X ορίζεται μονοσήμαντα αν και μόνον αν οι κατανομές των τυχαίων μεταβλητών Z_1 και Z_2 συμπίπτουν.

Απόδειξη Θέτοντας $\nu = E[w(X)]$ και $\mu = E(X)$, έχουμε από την (7.3.3) για $r>\theta$ ότι

$$\frac{w(r)f_X(r)}{\nu} = C_1 \frac{\bar{F}_X(r)}{\mu} \Rightarrow$$

$$\frac{-w(r)\bar{F}'_X(r)}{\bar{F}_X(r)} = \frac{\nu}{\mu} C_1 .$$

Θέτοντας $\kappa = \frac{\nu}{\mu} C_1$, έχουμε

$$\frac{\bar{F}'_X(r)}{\bar{F}_X(r)} = - \frac{\kappa}{w(r)} , \quad r > \theta .$$

Δηλαδή,

$$\int d \ln \bar{F}_X(r) = -\kappa \int \frac{1}{w(r)} dr + C' , \quad r > \theta .$$

Κατά συνέπεια,

$$\bar{F}_X(r) = \exp \left\{ -\kappa \int \frac{1}{w(r)} dr + C' \right\} , \quad r > \theta \qquad (6.3.4)$$

Επειδή δε $\bar{F}_X(\infty) = 0$ και $\bar{F}_X(\theta) = 1$ έπεται ότι το C' είναι προσδιορίσιμο. Αρα η $\bar{F}_X(r)$ προσδιορίζεται μονοσήμαντα από την συνάρτηση βάρους $w(r)$ και τη σταθερά $\kappa = (\nu/\mu)C_1 = \left\{E[w(X)]/E(X)\right\} C_1$.

Πόρισμα 6.3.1 Εστω ότι οι τυχαίες μεταβλητές X, Z_1 και Z_2 ορίζονται όπως στο θεώρημα 6.3.11. Εστω επίσης ότι η κατανομή της Z_1 συμπίπτει με εκείνην της Z_2. Τότε αν w(r)=r η κατανομή της X ορίζεται μονοσήμαντα ως Pareto (θ,κ).

Απόδειξη Θέτοντας όπου $w(r) = r$ η σχέση (6.3.4) γίνεται

$$\bar{F}_X(r) = \exp\left\{-\kappa \int \frac{1}{r}\, dr + C'\right\}$$

$$= \exp\left\{ \ln r^{-\kappa} C^* \right\} \quad , \quad r > \theta.$$

Αρα,

$\bar{F}_X(r) = C^* r^{-\kappa}$. Επειδή δε $\bar{F}_X(\theta) = 1 \Rightarrow C^* = \theta^{\kappa}$.

Επομένως,

$\bar{F}_X(r) = \theta^{\kappa}\, r^{-\kappa}$, $r > \theta$ δηλαδή $X \sim$ Pareto (θ,κ).

Πόρισμα 6.3.2 Εστω ότι οι τυχαίες μεταβλητές X, Z_1 και Z_2 ορίζονται όπως στο θεώρημα 6.3.11 για $\theta = 0$. Εστω επίσης ότι η κατανομή της τυχαίας μεταβλητής Z_1 συμπίπτει με εκείνην της Z_2. Τότε αν $w(r) = \beta$ η κατανομή της X ορίζεται μονοσήμαντα ως εκθετική$(\kappa\beta)$.

Απόδειξη Θέτοντας όπου $w(r) = \beta$ η σχέση (6.3.4) γίνεται για $\theta = 0$.

$$\bar{F}_X(r) = \exp\left\{-\kappa\beta r + C'\right\} \quad , \quad r > 0.$$

Αλλά $\bar{F}_X(0) = 1$. Κατα συνέπεια, $\bar{F}_X(r) = e^{-\kappa\beta r}$, $r > 0$.

Δηλαδή η τυχαία μεταβλητή X ακολουθεί εκθετική $(\kappa\beta)$ κατανομή.

Το τελευταίο θεώρημα της ενότητας διατυπώθηκε και αποδείχθηκε από τον Krishnaji (1970). Αποτελεί χαρακτηρισμό της κατανομής Pareto και βασίζεται στον ορισμό 6.3.7.

Θεώρημα 6.3.12 (Krishnaji 1970). Εστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_X(r)$, $r > \theta > 0$. Τότε η σχέση

$$\int \frac{f_X(x)}{x}\, dx = C_2 f_X(r), \quad r > \theta,\ C_2 < 1$$

είναι ικανή και αναγκαία για να ακολουθεί η X την κατανομή Pareto $\left(\theta,\ (1-C_2)/C_2\right)$.

Κ Ε Φ Α Λ Α Ι Ο 7

ΔΙΔΙΑΣΤΑΤΑ

ΣΥΣΤΗΜΑΤΑ ΚΑΤΑΝΟΜΩΝ

7.1 ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό επεκτείνονται στο χώρο των δυο διαστάσεων οι ορισμοί και τα συμπεράσματα του κεφαλαίου 7. Ετσι μελετώνται οι διδιάστατες σταθμισμένες κατανομές και οι διδιάστατες κατανομές που βασίζονται στα μερικά αθροίσματα. Αντικειμενικός σκοπός της μελέτης αυτής είναι η διατύπωση και απόδειξη ιδιοτήτων, χαρακτηριστικών ή μη, γνωστών διδιάστατων κατανομών μέσω ιδιοτήτων των σταθμισμένων κατανομών αφενός, και των κατανομών των βασισμένων στα μερικά αθροίσματα αφετέρου.

7.2 ΔΙΔΙΑΣΤΑΤΕΣ ΣΤΑΘΜΙΣΜΕΝΕΣ ΚΑΤΑΝΟΜΕΣ

Οπως αναφέρθηκε στην ενότητα 6.2, οι μονοδιάστατες σταθμισμένες κατανομές αναπτύχθηκαν τα τελευταία 25 χρόνια και αποτελούν ένα χρήσιμο εργαλείο επιλογής του κατάλληλου μοντέλου για τα δεδομένα της παρατήρησης. Οι διδιάστατες όμως σταθμισμένες κατανομές εμφανίστηκαν στη βιβλιογραφία σχετικά πρόσφατα. Πρώτη αναφορά στο θέμα γίνεται από τους Mahfoud and Patil (1982) και ακολουθεί η εργασία των Patil, Rao and Ratnaparkhi (1986).

Εστω το τυχαίο διάνυσμα (X,Y) του οποίου οι συνιστώσες είναι μη αρνητικές τυχαίες μεταβλητές με από κοινού συνάρτηση (πυκνότητας) πιθανότητας $f_{X,Y}(x,y)$ και έστω $w(x,y)$ μη αρνητική συνάρτηση βάρους. Εστω επίσης ότι η $E\left[w(X,Y)\right]$ υπάρχει. Η σταθμισμένη μορφή της $f_{X,Y}(x,y)$ δίνεται από τη σχέση

$$f^{w}_{X,Y}(x,y) = \frac{w(x,y)\ f_{X,Y}(x,y)}{E\left[w(X,Y)\right]} \ . \qquad (7.2.1)$$

Εστω $(X,Y)^w$ το διάνυσμα που έχει κατανομή την (7.2.1). Τότε

$$f^{w}_{X}(x) = \frac{E\left[w(X,Y)\,|\,X=x\right]\ f_{X}(x)}{E\left[w(X,Y)\right]} \qquad (7.2.2)$$

και

$$f^{w}_{Y|(X=x)}(y) = \frac{w(x,y)\ f_{Y|(X=x)}(y)}{E\left[w(X,Y)\right]} \qquad (7.2.3)$$

Προφανώς και οι δύο είναι αντίστοιχα σταθμισμένες μορφές της περιθωρίου κατανομής και της δεσμευμένης κατανομής του τυχαίου διανύσματος (X,Y).

Ορισμένες ειδικές μορφές συναρτήσεων βάρους που παρουσιάζουν ενδιαφέρον σε σχέση με πρακτικές εφαρμογές είναι οι

1. $w(x,y) = x^{\alpha}$
2. $w(x,y) = x$
3. $w(x,y) = \max(x,y)$
4. $w(x,y) = x + y$
5. $w(x,y) = xy$
6. $w(x,y) = x^{\alpha}\ y^{\beta}$

Αξίζει να σημειωθεί ότι στην περίπτωση που οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες, η συνάρτηση βάρους w(x,y) οφηγεί στο διάνυσμα $(X,Y)^w$ του οποίου οι συνιστώσες είναι εξαρτημένες.Το συμπέρασμα αυτό παρέχει τη δυνατότητα ορισμού διδιάστατων κατανομών.

Ας εξετάσουμε λίγο πιο αναλυτικά την περίπτωση διδιάστατης σταθμισμένης κατανομής με συνάρτηση βάρους $w(x,y) = x^{\alpha}$.

Εστω το τυχαίο διάνυσμα (X,Y) του οποίου οι συνιστώσες είναι μη αρνητικές τυχαίες μεταβλητές. Εστω επίσης ότι η από κοινού συνάρτηση (πυκνότητας) πιθανότητας των X και Y είναι η $f_{X,Y}(x,y)$ και έστω επίσης ότι η $E(X^{\alpha})$ υπάρχει. Με συνάρτηση βάρους την $w(x,y) = x^{\alpha}$ η από κοινού συνάρτηση (πυκνότητας) πιθανότητας του $(X,Y)^w$ ειναι η

$$f^{w}_{X,Y}(x,y) = \frac{x^{\alpha}\, f_{X,Y}(x,y)}{E(X^{\alpha})} \qquad (7.2.4)$$

Με βάση τη σχέση (7.2.4) προκύπτουν οι παρακάτω περιθώριες και δεσμευμένες κατανομές

$$f^{w}_{X}(x) = \frac{x^{\alpha}\, f_{X}(x)}{E(X^{\alpha})} \qquad (7.2.5)$$

$$f^{w}_{Y}(y) = \frac{E\left[X^{\alpha} \mid Y=y\right]\, f_{Y}(y)}{E\left[E\left[X^{\alpha}\right]\right]} \qquad (7.2.6)$$

$$f^{w}_{X|Y=y}(x|y) = \frac{x^{\alpha} f_{X|Y=y}(x)}{E\left[X^{\alpha}\right]} \qquad (7.2.7)$$

$$f^{w}_{Y|X=x}(y|x) = f_{Y|X=x}(y|x) \ . \qquad (7.2.8)$$

Οταν η συνάρτηση βάρους είναι $w(x,y) = x$, οι περιθώριες κατανομές των X^w και Y^w δίνονται από τις (7.2.5) και (7.2.6) με $\alpha=1$. Ετσι, η περιθώρια της Y^w είναι η σταθμισμένη περιθώρια κατανομή της Y με συνάρτηση βάρους τη συνάρτηση παλινδρόμησης της X στην Y, δηλαδή $w(y) = E(X|Y=y)$. Επιπλέον, αν η παλινδρόμηση της X στην Y είναι γραμμική, αν δηλαδή $E(X|Y=y) = \alpha y + \beta$, τότε η περιθώρια κατανομή της Y^w μπορεί να γραφτεί,

$$f^{w}_{Y}(y) = \alpha \frac{E(Y)}{E(X)} f^{*}_{Y}(y) + \left[1 - \alpha \frac{E(Y)}{E(X)}\right] f_{Y}(y).$$

7.3 ΔΙΔΙΑΣΤΑΤΑ ΣΥΣΤΗΜΑΤΑ ΚΑΤΑΝΟΜΩΝ

Στην ενότητα αυτή εξετάζεται για πολύ διδιάστατη γεννήτορα κατανομή το σύστημα των διδιάστατων σταθμισμένων κατανομών και τα διδιάστατα συστήματα των διδιάστατων κατανομών των βασισμένων σε μερικά αθροίσματα οδηγούν στην ίδια κατανομή. Ειδικότερα η παράγραφος 7.3.1 αντιμετωπίζει τη διακριτή περίπτωση ενώ η 7.3.2 τη συνεχή. Ετσι, θεμελιώνονται χαρακτηριστικές ιδιότητες για ορισμένες από τις διδιάστατες κατανομές που ορίσθηκαν στο κεφάλαιο 2 όπως η

διδιάστατη Waring, η διδιάστατη Yule, η διδιάστατη Pareto τύπου Mardia I που θα προκύπτουν από την ισότητα των συστημάτων κατανομών.

7.3.1 Α' Περίπτωση: Διακριτή Διδιάστατη Γεννήτωρ Κατανομή

Με βάση δεδομένη διακριτή διδιάστατη γεννήτορα κατανομή που έχει συνάρτηση πυκνότητας πιθανότητας $p_{x,y} \equiv P(X=x,Y=y)$ μπορούμε να κατασκευάσουμε διδιάστατες κατανομές με έναν από τους παρακάτω τρείς τρόπους:

<u>Ορισμός 7.3.1</u> $$q'_{r,l} \equiv P(Z_1=r,Z_2=l) = \frac{w(r,l)p_{r,l}}{E[w(X,Y)]}$$

<u>Ορισμός 7.3.2</u> $$q^{+}_{r,l} \equiv P(W_1=r,W_2=l) = \frac{\sum_{x=r+1}^{\infty}\sum_{y=l+1}^{\infty} p_{xy}}{E(XY)}$$

<u>Ορισμός 7.3.3</u> $$q^{*}_{r,l} \equiv P(Y_1=r,Y_2=l) = C_2 \sum_{x=r+1}^{\infty}\sum_{y=l+1}^{\infty} \frac{p_{x,y}}{x+y}$$

Με βάση τους τρεις αυτούς ορισμούς θα διατυπώσουμε χαρακτηριστικές ιδιότητες για γνωστές διδιάστατες διακριτές κατανομές. Παρατηρώντας τον ορισμό 7.3.1 διαπιστώνουμε ότι η συνάρτηση πιθανότητας του τυχαίου διανύσματος (Z_1,Z_2) συμπίπτει με εκείνην της σχέσης (7.2.1), κατά συνέπεια $(Z_1,Z_2) \equiv (X,Y)^w$. Χαρακτηριστικές ιδιότητες για γνωστές διδιάστατες διακριτές κατανομές μπορούν να αποδειχθούν με βάση τον ορισμό της διδιάστατης

σταθμισμένης κατανομής. Το θεώρημα 7.3.1 που ακολουθεί αποτελεί χαρακτηριστική ιδιότητα της διδιάστατης κατανομής Yule.

Θεώρημα 7.3.1 Εστω (X,Y), (Z_1,Z_2) τυχαία διανύσματα με μη αρνητικές ακέραιες συνιστώσες. Εστω επίσης ότι οι συναρτήσεις πιθανότητας των (X,Y) και (Z_1,Z_2) ικανοποιούν τη συνθήκη:

$$q'_{r,m} \equiv P(Z_1=r, Z_2=m) = \frac{w(r,m)p_{r,m}}{E[w(X,Y)]} \qquad (7.3.1)$$

$$\text{με} \quad w(X,Y) = X+Y+1$$

Τότε η κατανομή του (Z_1,Z_2) είναι η διδιάστατη Waring$(2,\rho)$, $\rho>0$, αν και μόνον αν η κατανομή του (X,Y) είναι η διδιάστατη Yule$(\rho+1)$, $\rho>0$.

Απόδειξη

Ευθύ: Εστω ότι η συνθήκη (7.3.1) ικανοποιείται και έστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί διδιάστατη Yule$(\rho+1)$. Τότε,

$$q'_{r,m} = \frac{(r+m+1)\,\dfrac{(\rho+1)_{(2)}\,(r+m)!}{(\rho+2)_{(r+m+2)}}}{E(X)+E(Y)+1}, \quad r=0,1,2,\ldots \quad m=0,1,2,\ldots$$

Αλλά $X \sim$ Yule$(\rho+1)$ και $Y \sim$ Yule$(\rho+1)$.

Κατά συνέπεια,

$$E(X) = 1/\rho \quad \text{και} \quad E(Y) = 1/\rho \ .$$

Επομένως,

$$q'_{r,m} = \frac{\dfrac{(\rho+1)_{(2)}\ (r+m+1)!}{(\rho+2)_{(r+m+2)}}}{\dfrac{\rho+2}{\rho}}$$

$$= \frac{\rho(\rho+1)(\rho+2)\ 2_{(r+m)}\,1_{(1)}}{(\rho+2)(\rho+2)_{(2)}(\rho+4)_{(r+m)}}$$

$$= \frac{\rho_{(2)}}{(\rho+2)_{(2)}}\ \frac{2_{(r+m)}}{(\rho+4)_{(r+m)}} \ .$$

Δηλαδή, το τυχαίο διάνυσμα (Z_1, Z_2) ακολουθεί διδιάστατη κατανομή Waring$(2;\rho)$.

Αντίστροφο: Εστω ότι το τυχαίο διάνυσμα (Z_1, Z_2) ακολουθεί διδιάστατη κατανομή Waring$(2;\rho)$, δηλαδή,

$$q'_{r,m} = \frac{\rho_{(2)}}{(\rho+2)_{(2)}}\ \frac{2_{(r+m)}}{(\rho+4)_{(r+m)}} \quad , \quad \rho>0 \ ; \quad r=0,1,2,\ldots \quad m=0,1,2,\ldots$$

Επίσης,

$$q'_{r,m} = C\ (r+m+1)\ p_{r,m}$$

οπότε,

$$p_{r,m} = C^{-1}(r+m+1)^{-1} q'_{r,m} \ .$$

Με αντικατάσταση προκύπτει ότι,

$$p_{r,m} = C^{-1}(r+m+1)^{-1} \frac{\rho_{(2)}}{(\rho+2)_{(2)}} \frac{1_{(r+m+1)}}{(\rho+4)_{(r+m)}}$$

$$= C^{-1} \frac{\rho_{(2)}}{(\rho+2)_{(2)}} \frac{(r+m)!}{(\rho+4)_{(r+1)}}$$

$$= C^{-1} \rho_{(2)} \frac{(r+m)!}{(\rho+2)_{(r+m+2)}} \ .$$

Αλλά, $\sum_{r=0}^{\infty} \sum_{m=0}^{\infty} p_{r,m} = 1$.

Δηλαδή,

$$1 = C^{-1} \rho_{(2)} \sum_{r=0}^{\infty} \sum_{m=0}^{\infty} \frac{(r+m)!}{(\rho+2)_{(r+m+2)}}$$

ή

$$C = \frac{\rho_{(2)}}{(\rho+1)_{(2)}} \sum_{r=0}^{\infty} \sum_{m=0}^{\infty} (\rho+1)_{(2)} \frac{(r+m)!}{(\rho+2)_{(r+m+2)}} \ .$$

Αρα,

$$C = \rho/(\rho+2) \ .$$

Επομένως,

$$p_{r,m} = \frac{\rho+2}{\rho}\,\rho_{(2)}\,\frac{(r+m)!}{(\rho+2)_{(r+m+2)}}$$

$$= (\rho+1)_{(2)}\,\frac{(r+m)!}{(\rho+2)_{(r+m+2)}}\ .$$

Δηλαδή, το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη Yule $(\rho+1)$.

Με βάση τον ορισμό 7.3.2 μπορούμε να οδηγηθούμε και πάλι σε χαρακτηρισμό της διδιάστατης κατανομής Yule. Συγκεκριμένα, αποδεικνύεται ότι ισχύει το παρακάτω θεώρημα .

Θεώρημα 7.3.2 Εστω τα τυχαία διανύσματα (X,Y) και (W_1,W_2) που έχουν μη αρνητικές ακέραιες συνιστώσες. Εστω επίσης ότι οι συναρτήσεις πιθανότητας των τυχαίων αυτών διανυσμάτων ικανοποιούν τη συνθήκη:

$$q^{+}_{r,m} \equiv P(W_1=r, W_2=m) = \frac{\sum_{x=r+1}^{\infty}\sum_{y=m+1}^{\infty} p_{xy}}{E(XY)} \qquad (7.3.2)$$

$$r=0,1,2,\ldots \quad m=0,1,2,\ldots$$

Τότε η κατανομή του (W_1,W_2) είναι διδιάστατη Waring $(3,\rho)$, $\rho>0$ αν και μόνον αν η κατανομή του (X,Y) είναι διδιάστατη Yule$(\rho+2)$, $\rho>0$.

Απόδειξη

Ευθύ: Εστω ότι η συνθήκη (7.3.2) ικανοποιείται και έστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί διδιάστατη Yule$(\rho+2)$. Τότε,

$$q^{+}_{r,m} = \frac{\sum_{x=r+1}^{\infty} \sum_{y=m+1}^{\infty} \frac{(\rho+2)_{(2)}\ (x+y)!}{(\rho+3)_{(x+y+2)}}}{E(XY)} \quad , \quad r=0,1,2,\ldots \quad m=0,1,2,\ldots$$

Αλλά, $E(X) = 1/(\rho+1)$ και $E(Y) = 1/(\rho+1)$ αφού $X \sim Yule(\rho+2)$ και $Y \sim Yule(\rho+2)$.

Επίσης, $Cov(X,Y) = \frac{\rho+2}{(\rho+1)^2\rho}$.

Κατά συνέπεια,

$$E(XY) = Cov(X,Y) + E(X) + E(Y)$$

$$= \frac{2}{(\rho+1)\rho} .$$

Αρα,

$$q^{+}_{r,m} = \frac{\sum_{x=r+1}^{\infty} \sum_{y=m+1}^{\infty} \frac{(\rho+2)_{(2)}\ (x+y)!}{(\rho+3)_{(x+y+2)}}}{\frac{2}{(\rho+1)\rho}}$$

$$= \frac{\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \frac{(\rho+2)_{(2)}\ (x+y+r+m+2)!}{(\rho+3)_{(x+y+r+m+4)}}}{\frac{2}{\rho_{(2)}}}$$

$$= \frac{\rho_{(2)}\ (\rho+2)_{(2)}}{2} \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \frac{1_{(r+m+x)}\ (r+m+3)_{(x+y)}}{(\rho+3)_{(r+m+4)}\ (\rho+r+m+7)_{(x+y)}}$$

$$= \frac{\rho_{(2)}\ (\rho+2)_{(2)}\ (r+m+2)!}{2\ (\rho+3)_{(r+m+4)}} \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \frac{(r+m+3)_{(x+y)}}{(\rho+r+m+7)_{(x+y)}}$$

$$= \frac{\rho_{(2)}\ (r+m+2)!}{2(\rho+3)_{(r+m+2)}}$$

$$= \frac{\rho_{(2)}\ 1_{(2)}\ 3_{(r+m)}}{2\ (\rho+3)_{(2)}\ (\rho+5)_{(r+m)}}$$

$$= \frac{\rho_{(2)}}{(\rho+3)_{(2)}} \frac{3_{(r+m)}}{(\rho+5)_{(r+m)}} .$$

Δηλαδή, το τυχαίο διάνυσμα (W_1, W_2) ακολουθεί τη διδιάστατη κατανομή Waring $(3;\rho)$.

Αντίστροφο: Εστω ότι το τυχαίο διάνυσμα (W_1, W_2) ακολουθεί διδιάστατη κατανομή Waring$(3,\rho)$, τότε

$$q^{+}_{r,m} = \frac{\rho_{(2)}}{(\rho+3)_{(2)}} \frac{3_{(r+m)}}{(\rho+5)_{(r+m)}} \quad , \quad \rho>0.$$

Επίσης,

$$q^{+}_{r,m} = C \sum_{x=r+1}^{\infty} \sum_{y=m+1}^{\infty} p_{x,y} \ . \tag{7.3.3}$$

Από τη σχέση (7.3.3) για $r,m = 0,1,\ldots$ προκύπτουν οι παρακάτω τρεις σχέσεις:

$$q^{+}_{r,m} - q^{+}_{r+1,m} = C \sum_{x=m+1}^{\infty} p_{r+1,x} \tag{7.3.4}$$

$$q^{+}_{r,m} - q^{+}_{r,m+1} = C \sum_{x=m+1}^{\infty} p_{x,m+1} \tag{7.3.5}$$

$$q^{+}_{r,m} - q^{+}_{r+1,m+1} = C \left[\sum_{x=m+1}^{\infty} p_{r+1,x} + \sum_{x=r+1}^{\infty} p_{x,m+1} - p_{r+1,m+1} \right]. \tag{7.3.6}$$

Με βάση τις σχέσεις (7.3.4), (7.3.5) και (7.3.6) προκύπτει ότι:

$$q^{+}_{r+1,m+1} + q^{+}_{r,m} - q^{+}_{r+1,m} - q^{+}_{r,m+1} = C\, p_{r+1,m+1} \ .$$

Δηλαδή,

$$p_{r,m} = C^{-1} \left(q^{+}_{r,m} + q^{+}_{r-1,m-1} - q^{+}_{r,m-1} - q^{+}_{r-1,m} \right) \ .$$

Αλλά,

$$q^{+}_{r,m} = \frac{\rho_{(2)}}{(\rho+3)_{(2)}} \frac{3_{(r+m)}}{(\rho+5)_{(r+m)}} .$$

Επομένως,

$$p_{r,m} = C^{-1} \frac{\rho_{(2)}}{(\rho+3)_{(2)}} \left[\frac{3_{(r+m)}}{(\rho+5)_{(r+m)}} + \frac{3_{(r+m-2)}}{(\rho+5)_{(r+m-2)}} - 2 \frac{3_{(r+m-1)}}{(\rho+5)_{(r+m-1)}} \right]$$

$$= C^{-1} \frac{\rho_{(2)}}{(\rho+3)_{(2)}} \frac{(r+m)!}{2} \left[\frac{(r+m+2)(r+m+1)}{(\rho+5)_{(r+m)}} + \frac{1}{(\rho+5)_{(r+m-2)}} - 2 \frac{(r+m+1)}{(\rho+5)_{(r+m-1)}} \right]$$

$$= C^{-1} \frac{\rho_{(2)}}{(\rho+3)_{(2)}} \frac{(r+m)!}{2(\rho+5)_{(r+m)}} \left[(r+m+2)(r+m+1) + (\rho+r+m+4)(\rho+r+m+3) - 2\,(r+m+1)(\rho+r+m+4) \right]$$

$$= C^{-1} \frac{\rho_{(2)}}{(\rho+3)_{(2)}} \frac{(r+m)!}{2(\rho+5)_{(r+m)}} (\rho+2)(\rho+3)$$

$$= C^{-1} \frac{\rho_{(2)}}{2} \frac{(\rho+2)_{(2)}}{(\rho+3)_{(2)}} \frac{(r+m)!}{(\rho+5)_{(r+m)}} .$$

Αλλά, $\sum_{r=0}^{\infty} \sum_{m=0}^{\infty} p_{r,m} = 1.$

Δηλαδή,

$$1 = C^{-1} \frac{\rho_{(2)}}{2} \sum_{r=0}^{\infty} \sum_{m=0}^{\infty} \frac{(\rho+2)_{(2)}}{(\rho+3)_{(2)}} \frac{(r+m)!}{(\rho+5)_{(r+m)}} .$$

Επομένως, $C = \frac{\rho(\rho+1)}{2}$.

Αρα,

$$p_{r,m} = \frac{(\rho+2)_{(2)}}{(\rho+3)_{(2)}} \frac{(r+m)!}{(\rho+5)_{(r+m)}} .$$

Δηλαδή το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Yume $(\rho+2)$.

Ο ορισμός 7.3.3 εισήχθη από την Xekalaki (1986) και αποτελεί επέκταση στο χώρο των δύο διαστάσεων του γνωστού υποδείγματος STER [Bissinger (1965)]. Η Xekalaki (1986) ονομάζει την κατανομή του τυχαίου διανύσματος (Y_1,Y_2) διδιάστατη κατανομή STER και μέσω αυτής διατυπώνει και αποδεικνύει χαρακτηριστική ιδότητα για τη διδιάστατη κατανομή Yule. Για την πλήρότητα της παρουσίασης παραθέτουμε το συμπέρασμα που προαναφέραμε.

<u>Θεώρημα 7.3.3</u> (Xekalaki,1986) Εστω (X,Y), (Y_1,Y_2) τυχαία διανύσματα με μη αρνητικές ακέραιες συνιστώσες. Εστω επίσης ότι οι συναρτήσεις πιθανότητας των (X,Y) και (Y_1,Y_2) ικανοποιούν τη συνθήκη:

$$,_m = P(Y_1=r, Y_2=m) = \sum_{x=r+1}^{\infty} \sum_{y=m+1}^{\infty} \frac{p_{x,y}}{x+y}$$

:ε η κατανομή του τυχαίου διανύσματος (Y_1,Y_2) είναι η διδιάστατη
:ing $(2,\rho)$, $\rho>0$ αν και μόνον αν η κατανομή του (X,Y) είναι η
;ιάστατη Yule $(\rho+1)$, $\rho>0$.

Το συμπέρασμα του θεωρήματος που ακολουθεί αποτελεί ιδιότητα
; διδιάστατης κατανομής Yule και βασίζεται στην ισότητα των
:ανομών που προκύπτουν από τους ορισμούς 7.3.1 και 7.3.2.

ρημα 7.3.4 Εστω τα τυχαία διανύσματα (X,Y), (Z_1,Z_2) και
$,Y_2)$ των οποίων οι συναρτήσεις πιθανότητας ικανοποιούν τις
θήκες:

$$m = P(Z_1=r, Z_2=m) = \frac{w(r,m)p_{r,m}}{E[w(X,Y)]} \quad , \; r=0,1,2,\ldots \quad m=0,1,2,\ldots$$

με $w(X,Y) = X+Y+1$

$$m = P(Y_1=r, Y_2=m) = C_2 \sum_{x=r+1}^{\infty} \sum_{y=m+1}^{\infty} \frac{p_{x,y}}{x+y} \quad , \; r=0,1,2,\ldots \quad m=0,1,2,\ldots \qquad (7.3.7)$$

ε τα τυχαία διανύσματα (Z_1,Z_2), (Y_1,Y_2) ακολουθούν την ίδια

$$q^{*}_{r,m} = C_2 \sum_{x=r+1}^{\infty} \sum_{y=m+1}^{\infty} \frac{p_{x,y}}{x+y} .$$

Οπότε, $$q'_{r,m} = C_2 \sum_{x=r+1}^{\infty} \sum_{y=m+1}^{\infty} \frac{p_{x,y}}{x+y} . \qquad (7.3.8)$$

Από τη σχέση (7.3.8) για r,m = 0,1,..προκύπουν οι παρακάτω σχέσεις

$$q'_{r,m} - q'_{r+1,m} = C_2 \sum_{x=m+1}^{\infty} \frac{p_{r+1,x}}{x+r+1} \qquad (7.3.9)$$

$$q'_{r,m} - q'_{r+1,m} = C_2 \sum_{x=r+1}^{\infty} \frac{p_{x,m+1}}{x+m+1} \qquad (7.3.10)$$

$$q'_{r,m} - q'_{r+1,m+1} = C_2 \left[\sum_{x=m+1}^{\infty} \frac{p_{r+1,x}}{x+r+1} + \sum_{x=r+1}^{\infty} \frac{p_{x,m+1}}{x+m+1} - \frac{p_{r+1,m+1}}{r+m+2} \right] . \qquad (7.3.11)$$

Με βάση τις σχέσεις (7.3.9), (7.3.10) και (7.3.11) προκύπτει ότι:

$$q'_{r+1,m+1} + q'_{r,m} - q'_{r+1,m} - q'_{r,m+1} = C_2 \frac{p_{r+1,m+1}}{r+m+2} .$$

Αλλά,

$$q'_{r,m} = \frac{(r+m+1)p_{r,m}}{C_1} .$$

Αρα,

$$(r+m+3)p_{r+1,m+1} + (r+m+1)p_{r,m} - (r+m+2)p_{r+1,m} - (r+m+2)p_{r,m+1} =$$

$$= C \frac{p_{r+1,m+1}}{r+m+2}$$

όπου $C = C_1 C_2$.

Επομένως,

$$\left[(r+m+3)(r+m+2) - C\right]p_{r+1,m+1} + (r+m+1)(r+m+2)p_{r,m} - (r+m+2)^2 p_{r+1,m} -$$

$$-(r+m+2)^2 p_{r,m+1} = 0 .$$

Η εξίσωση αυτή διαφορών ικανοποιείται από την

$$p_{r,m} = \frac{(\rho+1)_{(2)}}{(\rho+2)_{(2)}} \frac{(r+m)!}{(\rho+4)_{(r+m)}} , \quad r=0,1,2,\ldots \quad m=0,1,2,\ldots$$

για $C = (\rho+2)(\rho+3)$.

Να σημειωθεί ότι εξετάζεται η μοναδικότητα της λύσης αυτής ώστε να μπορέσει η σχέση (7.3.7) να αποτελέσει χαρακτηριστική ιδιότητα της διδιάστατης κατανομής Yule (ρ).

7.3.2 Β' Περίπτωση: Συνεχής Διδιάστατη Γεννήτωρ Κατανομή

Με βάση δεδομένη συνεχή διδιάστατη γεννήτορα κατανομή που έχει συνάρτηση πυκνότητας πιθανότητας $f_{X,Y}(x,y)$ $x>0$, $y>0$ μπορούμε να δημιουργήσουμε διδιάστατες κατανομές με έναν από τους παρακάτω τρεις τρόπους:

<u>Ορισμός 7.3.4</u> $$f_{Z_1,Z_2}(r,m) = \frac{w(r,m)\, f_{X,Y}(r,m)}{E[w(X,Y)]} \quad , r>0,\ m>0$$

<u>Ορισμός 7.3.5</u> $$f_{W_1,W_2}(r,m) = C_1 \frac{\bar{F}_{X,Y}(r,m)}{E(XY)} \quad , r>0,\ m>0$$

<u>Ορισμός 7.3.6</u> $$f_{Y_1,Y_2}(r,m) = C_2 \int_m^{\infty}\int_r^{\infty} \frac{f_{X,Y}(x,y)}{x+y+1}\, dx\, dy \ , \ r>0, m>0$$

Με βάση τους τρείς αυτούς ορισμούς θα διατυπώσουμε χαρακτηριστικές ιδιότητες για γνωστές συνεχείς διδιάστατες κατανομές. Ο πρώτος χαρακτηρισμός αφορά τη διδιάστατη κατανομή Pareto τύπου Mardia I και βασίζεται στον ορισμό 7.3.4. Ειδικότερα, αποδεικνύεται ότι η ιδιότητα του αμετάβλητου της μορφής *(form invariance)* που χαρακτηρίζει τη μονοδιάστατη κατανομή Pareto διατηρείται και στο χώρο των δύο διαστάσεων.

Θεώρημα 7.3.5 Εστω (X,Y), (Z_1,Z_2) τυχαία διανύσματα των οποίων οι συνιστώσες είναι μεγαλύτερες από τη μονάδα. Εστω επίσης ότι οι συναρτήσεις πυκνότητας πιθανότητας των (X,Y) και (Z_1,Z_2) ικανοποιούν τη συνθήκη

$$f_{Z_1,Z_2}(r,m) = \frac{w(r,m)\, f_{X,Y}(r,m)}{E[w(X,Y)]} \quad , \; r>1, \; m>1 \tag{7.3.11}$$

με $w(X,Y) = X+Y-1$.

Τότε η κατανομή του (Z_1,Z_2) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho>0$ αν και μόνον αν η κατανομή του (X,Y) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho+1$.

Απόδειξη

Ευθύ: Εστω ότι η συνθήκη (7.3.11) ικανοποιείται και έστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho+1$. Τότε,

$$f_{Z_1,Z_2}(r,m) = \frac{(r+m-1)(\rho+1)(\rho+2)(r+m-1)^{-(\rho+3)}}{E(X) + E(Y) - 1} \quad , \; r>1, \; m>1 \; .$$

Αλλά $X \sim \text{Pareto}(1,\rho+1)$ και $Y \sim \text{Pareto}(1,\rho+1)$.

Κατά συνέπεια,

$$E(X) = E(Y) = (\rho+1)/\rho \ .$$

Επομένως,

$$f_{Z_1,Z_2}(r,m) = \frac{(\rho+1)\rho+2)(r+m-1)^{-(\rho+2)}}{2\left[(\rho+1)/\rho\right] - 1}$$

$$= \frac{(\rho+1)(\rho+2)(r+m-1)^{-(\rho+2)}}{(\rho+2)/\rho}$$

$$= \rho(\rho+1)(r+m-1)^{-(\rho+2)} .$$

Δηλαδή, διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho>0$.

Αντίστροφο: Εστω ότι το τυχαίο διάνυσμα (Z_1,Z_2) ακολουθει τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο ρ. Τότε,

$$f_{Z_1,Z_2}(r,m) = \rho(\rho+1)(r+m-1)^{-(\rho+2)} \ , \ \rho>0.$$

Επίσης,

$$f_{Z_1,Z_2}(r,m) = C\ (r+m-1)f_{X,Y}(r,m).$$

Τότε,

Θεώρημα 7.3.5 Εστω (X,Y), (Z_1,Z_2) τυχαία διανύσματα των οποίων οι συνιστώσες είναι μεγαλύτερες από τη μονάδα. Εστω επίσης ότι οι συναρτήσεις πυκνότητας πιθανότητας των (X,Y) και (Z_1,Z_2) ικανοποιούν τη συνθήκη

$$f_{Z_1,Z_2}(r,m) = \frac{w(r,m)\, f_{X,Y}(r,m)}{E[w(X,Y)]} \quad , \; r>1, \; m>1 \tag{7.3.11}$$

με $w(X,Y) = X+Y-1$.

Τότε η κατανομή του (Z_1,Z_2) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho>0$ αν και μόνον αν η κατανομή του (X,Y) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho+1$.

Απόδειξη

Ευθύ: Εστω ότι η συνθήκη (7.3.11) ικανοποιείται και έστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho+1$. Τότε,

$$f_{Z_1,Z_2}(r,m) = \frac{(r+m-1)(\rho+1)(\rho+2)(r+m-1)^{-(\rho+3)}}{E(X) + E(Y) - 1} \quad , \; r>1, \; m>1 \; .$$

Αλλά $X \sim \text{Pareto}(1,\rho+1)$ και $Y \sim \text{Pareto}(1,\rho+1)$.

Κατά συνέπεια,

$$E(X) = E(Y) = (\rho+1)/\rho \ .$$

Επομένως,

$$f_{Z_1,Z_2}(r,m) = \frac{(\rho+1)\rho+2)(r+m-1)^{-(\rho+2)}}{2\left\lfloor(\rho+1)/\rho\right\rfloor - 1}$$

$$= \frac{(\rho+1)(\rho+2)(r+m-1)^{-(\rho+2)}}{(\rho+2)/\rho}$$

$$= \rho(\rho+1)(r+m-1)^{-(\rho+2)}.$$

Δηλαδή, διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho>0$.

Αντίστροφο: Εστω ότι το τυχαίο διάνυσμα (Z_1,Z_2) ακολουθει τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο ρ. Τότε,

$$f_{Z_1,Z_2}(r,m) = \rho(\rho+1)(r+m-1)^{-(\rho+2)} \ , \ \rho>0.$$

Επίσης,

$$f_{Z_1,Z_2}(r,m) = C\ (r+m-1)f_{X,Y}(r,m).$$

Τότε,

Δηλαδή, το τυχαίο διάνυσμα (Z_1,Z_2) ακολουθεί διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho+1$.

Με βάση τον ορισμό 7.3.5 μπορούμε να οδηγηθούμε σε χαρακτηρισμό της διδιάστατης κατανομής Pareto τύπου Mardia I. Συγκεκριμένα μπορούμε να διατυπώσουμε και να αποδείξουμε το παρακάτω θεώρημα.

Θεώρημα 7.3.6 Εστω (X,Y), (W_1,W_2) τυχαία διανύσματα των οποίων οι συνιστώσες είναι μεγαλύτερες από τη μονάδα. Εστω επίσης ότι οι συναρτήσεις πυκνότητας πιθανότητας των (X,Y) και (W_1,W_2) ικανοποιούν τη συνθήκη

$$f_{W_1,W_2}(r,m) = C_1 \frac{\bar{F}_{X,Y}(r,m)}{E(XY)}, \quad r>1,\ m>1. \tag{7.3.12}$$

Τότε η κατανομή του (W_1,W_2) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο ρ αν και μόνον αν η κατανομή του (X,Y) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho+2$.

Απόδειξη.

Ευθύ: Εστω ότι η συνθήκη (7.3.12) ικανοποιείται και έστω ότι το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho+2$. Τότε, για $r>1$, $m>1$

$$f_{W_1,W_2}(r,m) = C_1 \frac{(r+m-1)^{-(\rho+2)}}{\frac{\rho^2+\rho+1}{\rho(\rho+1)}}$$

$$f_{X,Y}(r,m) = C^{-1}(r+m-1)^{-1} f_{Z_1,Z_2}(r,m).$$

Με αντικατάσταση προκύπτει ότι,

$$f_{X,Y}(r,m) = C^{-1}\rho(\rho+1)(r+m-1)^{-(\rho+3)}.$$

Αλλά,

$$\int_1^\infty \int_1^\infty f_{X,Y}(r,m)\, dr\, dm = 1,$$

δηλαδή,

$$C^{-1}\rho(\rho+1) \int_1^\infty \int_1^\infty (r+m-1)^{-(\rho+3)}\, dr\, dm = 1$$

ή

$$C^{-1}\rho(\rho+1)\, \frac{1}{(\rho+2)(\rho+1)} = 1 .$$

Αρα, $C = \rho/(\rho+2)$

και επομένως,

$$f_{Z_1,Z_2}(r,m) = (\rho+1)(\rho+2)(r+m-1)^{-(\rho+3)}.$$

$$= C_1 \frac{\rho(\rho+1)}{\rho^2+\rho+1} (r+m-1)^{-(\rho+2)}.$$

Αλλά,

$$\int_1^\infty \int_1^\infty f_{W_1,W_2}(r,m)\, dr\, dm = 1.$$

Επομένως,

$$C_1 \frac{\rho(\rho+1)}{\rho^2+\rho+1} \int_1^\infty \int_1^\infty (r+m-1)^{-(\rho+2)}\, dr\, dm = 1.$$

Δηλαδή,

$$C_1 = \rho^2+\rho+1.$$

Επομένως,

$$f_{W_1,W_2}(r,m) = \rho(\rho+1)(r+m-1)^{-(\rho+2)}.$$

Δηλαδή, το τυχαίο διάνυσμα (W_1,W_2) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο ρ.

Αντίστροφο: Εστω ότι το τυχαίο διάνυσμα (W_1,W_2) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο ρ. Τότε,

για $r>1$, $m>1$

$$f_{W_1,W_2}(r,m) = \rho(\rho+1)(r+m-1)^{-(\rho+2)} .$$

Επίσης,

$$f_{W_1,W_2}(r,m) = C\ \bar{F}_{X,Y}(r,m)$$

όπου $C = C_1/E(XY)$.

Αρα,

$$\rho(\rho+1)(r+m-1)^{-(\rho+2)} = C\ \bar{F}_{X,Y}(r,m).$$

Επομένως,

$$\bar{F}_{X,Y}(r,m) = C^{-1}\rho(\rho+1)(r+m-1)^{-(\rho+2)}.$$

Αλλά από τον ορισμό των X, Y προκύπτει ότι

$C = \rho(\rho+1)$.

Δηλαδή,

$$\bar{F}_{X,Y}(r,m) = (r+m-1)^{-(\rho+2)}.$$

Το τυχαίο διάνυσμα (X,Y) ακολουθεί δηλαδή τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho+2$.

Με βάση τον ορισμό 7.3.6 είναι δυνατόν να οδηγηθούμε σε χαρακτηρισμό της διδιάστατης Pareto τύπου Mardia I. Συγκεκριμένα μπορεί να αποδεχθεί το παρακάτω θέωρημα .

Θεώρημα 7.3.7 Εστω τα τυχαία διανύσματα (X,Y), (Y_1,Y_2) των οποίων οι συνιστώσες είναι μεγαλύτερες από τη μονάδα. Εστω επίσης ότι οι συναρτήσεις πυκνότητας πιθανότητας των (X,Y) και (Y_1,Y_2) ικανοποιούν τη συνθήκη

$$f_{Y_1,Y_2}(r,m) = C_2 \int_m^{\infty} \int_r^{\infty} \frac{f_{X,Y}(x,y)}{x+y-1}\, dx\, dy \ , \ r>1, \ m>1. \qquad (7.3.13)$$

Τότε η κατανομή του (Y_1,Y_2) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho>0$, αν και μόνον αν η κατανομή του (X,Y) είναι η διδιάστατη Pareto τύπου Mardia I με παράμετρο $\rho+1$.

Απόδειξη.

Ευθύ: Εστω ότι η συνθήκη (7.3.13) ικανοποιείται και έστω ότι το (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I . Τότε,

$$f_{Y_1,Y_2}(r,m) = C_2 \int_m^{\infty} \int_r^{\infty} \frac{(\rho+1)(\rho+2)(x+y-1)^{-(\rho+3)}}{x+y-1}\, dx\, dy$$

$$= C_2(\rho+1)(\rho+2)\int_m^{\infty} \int_r^{\infty} (x+y-1)^{-(\rho+4)}\, dx\, dy$$

$$= C_2 \frac{(\rho+1)(\rho+2)}{\rho+3} \int_m^\infty (r+y-1)^{-(\rho+3)}\, dy$$

$$= C_2 \frac{\rho+1}{(\rho+3)(\rho+2)} (r+m-1)^{-(\rho+2)}.$$

Αλλά,

$$\int_1^\infty \int_1^\infty f_{Y_1,Y_2}(r,m)\, dr\, dm = 1.$$

Δηλαδή,

$$C_2 \frac{\rho+1}{\rho+3} \int_1^\infty \int_1^\infty (m+r-1)^{-(\rho+2)}\, dr\, dm = 1,$$

οπότε,

$$C_2 = (\rho+3)\rho .$$

Επομένως,

$$f_{Y_1,Y_2}(r,m) = \rho(\rho+1)(r+m-1)^{-(\rho+2)} .$$

Δηλαδή, το (Y_1,Y_2) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho>0$.

Αντίστροφο: Εστω ότι το τυχαίο διάνυσμα (Y_1, Y_2) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I . Τοτε, για $r>1$, $m>1$.

$$f_{Y_1,Y_2}(r,m) = \rho(\rho+1)(r+m-1)^{-(\rho+2)} ,$$

Επίσης,

$$f_{Y_1,Y_2}(r,m) = C_2 \int_m^\infty \int_r^\infty \frac{f_{X,Y}(x,y)}{x+y-1}\, dx\, dy .$$

Αρα,

$$\rho(\rho+1)(r+m-1)^{-(\rho+2)} = C_2 \int_m^\infty \int_r^\infty \frac{f_{X,Y}(x,y)}{x+y-1}\, dx\, dy .$$

Παραγωγίζοντας ως προς r και m, προκύπτει ότι

$$\rho(\rho+1)(\rho+2)(\rho+3)(r+m-1)^{-(\rho+4)} = C_2 \frac{f_{X,Y}(r,m)}{r+m-1} .$$

Από τη σχέση αυτή έπεται ότι

$$f_{X,Y}(r,m) = C_2^{-1}\rho(\rho+1)(\rho+2)(\rho+3)(r+m-1)^{-(\rho+3)} .$$

Αλλά,

$$\int_1^\infty \int_1^\infty f_{X,Y}(r,m)\, dr\, dm = 1$$

και επομένως,

$$C_2 = \rho(\rho+3).$$

Κατά συνέπεια,

$$f_{X,Y}(r,m) = (\rho+1)(\rho+2)(r+m-1)^{-(\rho+3)}.$$

Δηλαδή, το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho+1$, $\rho>0$.

Το θεώρημα που ακολουθεί αποτελεί ιδιότητα της διδιάστατης κατανομής Pareto τύπου Mardia I, που βασίζεται στην ισότητα των κατανομών που προκύπτουν από τους ορισμούς 7.3.4 και 7.3.6.

<u>Θεώρημα 7.3.8</u> Εστω τα τυχαία διανύσματα (X,Y), (Z_1,Z_2) και (Y_1,Y_2) των οποίων οι συνιστώσες είναι μεγαλύτερες από τη μονάδα. Εστω επίσης ότι οι συναρτήσεις πυκνότητας πιθανότητας αυτών ικανοποιούν τις παρακάτω δυο συνθήκες:

$$f_{Z_1,Z_2}(r,m) = \frac{w(r,m)\, f_{X,Y}(r,m)}{E[w(X,Y)]}, \quad r>1,\ m>1 \qquad (7.3.14)$$

με $w(X,Y) = X+Y-1$

και

$$f_{Y_1,Y_2}(r,m) = C_2 \int_m^{\infty} \int_r^{\infty} \frac{f_{X,Y}(x,y)}{x+y-1}\, dx\, dy \ , \ r>1, \ m>1 \qquad (7.3.15)$$

Τότε τα τυχαία διανύσματα (Z_1,Z_2) και (Y_1,Y_2) ακολουθούν την ίδια κατανομή αν το τυχαίο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho+1$, $\rho>0$.

Απόδειξη.

Εστω ότι το τυχαιο διάνυσμα (X,Y) ακολουθεί τη διδιάστατη κατανομή Pareto τύπου Mardia I με παράμετρο $\rho+1$ τότε, για $r>1, m>1$

$$f_{Z_1,Z_2}(r,m) = \frac{(r+m-1)(\rho+1)(\rho+2)(r+m-1)^{-(\rho+3)}}{(\rho+2)/\rho}$$

$$= \rho(\rho+1)(r+m-1)^{-(\rho+2)}$$

και

$$f_{Y_1,Y_2}(r,m) = C_2 \int_m^{\infty} \int_r^{\infty} \frac{(\rho+1)(\rho+2)(x+y-1)^{-(\rho+3)}}{x+y-1}$$

$$= C_2 \frac{\rho+1}{\rho+3} (r+m-1)^{-(\rho+2)}.$$

Μετά τον υπολογισμό της σταθεράς C_2 από την αρχική συνθήκη

$$\int_1^{\infty}\int_1^{\infty} f_{Y_1,Y_2}(r,m)\, dr\, dm = 1$$

ροκύπτει ότι:

$$Y_1,Y_2(r,m) = \rho(\rho+1)(r+m-1)^{-(\rho+2)}, \quad r>1,\ m>1.$$

ηλαδή,

$$Z_1,Z_2(r,m) = f_{Y_1,Y_2}(r,m), \quad r>1,\ m>1.$$

ικάζεται ότι για το συγκεκριμένο θεώρημα ισχύει και το ντίστροφο. Εστω ότι

$$Z_1,Z_2(r,m) = f_{Y_1,Y_2}(r,m), \quad r>1,\ m>1,$$

ηλαδή,

$$r+m-1)f_{X,Y}(r,m) = C\int_m^{\infty}\int_r^{\infty} \frac{f_{X,Y}(x,y)}{x+y-1}\, dx\, dy .$$

αραγωγίζοντας ως προς r και m προκύπτει ότι

$$\frac{}{r} f_{X,Y}(r,m) + \frac{\partial}{\partial m} f_{X,Y}(r,m) + (m+r-1)\frac{\partial^2}{\partial r\,\partial m} f_{X,Y}(r,m) = C\frac{f_{X,Y}(r,m)}{r+m-1} .$$

διαφορική αυτή εξίσωση επαληθεύεται από την

$$f_{X,Y}(r,m) = (\rho+1)(\rho+2)(r+m-1)^{-(\rho+3)}$$

με $C = (\rho+2)(\rho+3)$.

Να σημειωθεί ότι εξετάζεται η μοναδικότητα της λύσης αυτής ώστε να μπορέσει η σχέση (7.3.15) να αποτελέσει χαρακτηριστική ιδιότητα της διδιάστατης κατανομής Pareto τύπου Mardia I.

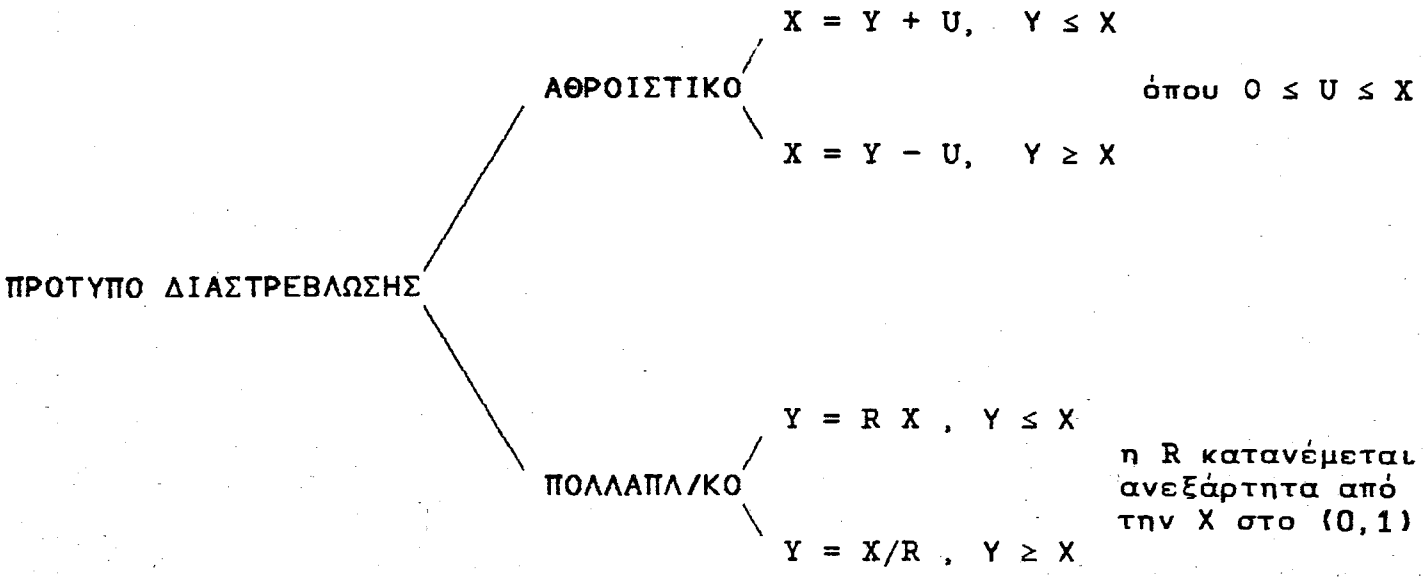
ΚΕΦΑΛΑΙΟ 8

ΠΡΟΤΥΠΟ ΔΙΑΣΤΡΕΒΛΩΣΗΣ

8.1 ΕΙΣΑΓΩΓΗ

Στη πράξη συχνά παρουσιάζονται καταστάσεις όπου η αρχική παρατήρηση Χ υπόκειται σε μια διαδικασία φθοράς (καταστροφής). Με άλλα λόγια η τελική παρατηρούμενη τιμή είναι το μη καταστραμένο τμήμα της Χ, έστω δε αυτό Υ. Ετσι για παράδειγμα, το δηλούμενο εισόδημα ή ο αριθμός των τροχαίων ατυχημάτων που αναφέρονται είναι γενικά μικρότερα ή ίσα με το πραγματικό εισόδημα ή τον πραγματικό αριθμό ατυχημάτων γιατί τα άτομα τείνουν να δηλώνουν μικρότερο εισόδημα ή λιγώτερα ατυχήματα για λόγους φορολογικούς ή ασφαλιστικούς αντίστοιχα *(underrorting)*. Επίσης η καταγραφή σημάτων ραντάρ, μεγεθών σμηνών πουλιών ή κοπαδιών ζώων μπορεί να καταλήξει σε μια παρατήρηση το πολύ ίση με τον πραγματικό αριθμό σημάτων πουλιών ή ζώων λόγω κακής ορατότητας ή άλλων μορφών μεροληψίας. Από την άλλη πλευρά είναι πολύ συνηθισμένο το φαινόμενο, όπως για παράδειγμα στις απαιτήσεις από ασφαλιστικές εταιρείες, η δηλούμενη τιμή μιας παρατήρησης να είναι μεγαλύτερη από την πραγματική τιμή της (overreported ή overrecorded). Αυτές οι περιπτώσεις μπορούν να αντιμετωπισθούν με ενιαίο τρόπο αν θεωρηθεί ότι εμπίπτουν στο πλαίσιο ενός μοντέλου όπου η αρχική παρατηρηση Χ υποβάλλεται σε μια διαδικασία φθοράς που οδηγεί στην παρατηρούμενη τιμή Υ η οποία εν γένει διαφέρει από την Χ. Θα ονομάσουμε το μοντέλο αυτό **Πρότυπο Διαστρέβλωσης** και θα διακρίνουμε δύο περιπτώσεις: το **Προσθετικό Πρότυπο Διαστρέβλωσης** και το **Πολλαπλασιαστικό Πρότυπο Διαστρέβλωσης**. Οι παραπάνω ορισμοί μπορούν να αποδοθούν σχηματικά με το διάγραμμα 8.1.1

Δ Ι Α Γ Ρ Α Μ Μ Α 8.1.1



Θα ξεκινήσουμε την μελέτη του προτύπου διαστρέβλωσης παραθέτοντας σύντομα τη βιβλιογραφία που ήδη υπάρχει στο θέμα αυτό κάνοντας παράλληλα προσπάθεια ενοποίησης των συπερασμάτων. Ο κύριος όγκος της υπάρχουσας βιβλιογραφίας καλύπτει το αθροιστικό πρότυπο διαστρέβλωσης στο πλαίσιο της γνωστής **συνθήκης Rao-Rubin.** Η περίπτωση αυτή, που αφορά διακριτές κυρίως κατανομές, δεν αποτελεί αντικείμενο της παρούσας μελέτης. Θα αναφερθούμε σύντομα σε αυτήν για την πληρότητα της παρουσίασης.

Εστω X μη αρνητική τυχαία μεταβλητή που παίρνει ακέραιες τιμές και έστω $P_n = P(X = n)$ με $n = 0,1,2,\ldots$. Εστω $S(r|n) = P(Y=r|X=n)$ η πιθανότητα με την οποία η αρχική τιμή n της τυχαίας

μεταβλητής X παρατηρείται ως r ($\rho \leq n$), και έστω Y η παρατηρούμενη μεταβλητή. Τότε ,

$$P(Y=r) = \sum_{n=r}^{\infty} P_n \, S(r|n) \qquad r = 0,1,2,... \qquad (8.1.1.)$$

Πρώτος ο C.R. Rao (1963) στην πρωτοπόρα εργασία του *"On Discrete Distributions Arising out of Methods of Ascertainement"* έθεσε τα θεμέλια της μελέτης του προτύπου διαστρέβλωσης. Έτσι υποθέτοντας ότι

$$S(r|n) = \binom{n}{r} p^r (1-p)^{n-r}, \qquad 0 < p < 1 \qquad (8.1.2)$$

παρατήρησε ότι όποτε η τυχαία μεταβλητή X ακολουθούσε την κατανομή Poisson ίσχυε η σχέση

$$P(Y=r) = P(Y=r|X-Y=0). \qquad (8.1.3)$$

Αργότερα οι Rao-Rubin (1964) χαρακτήρισαν την κατανομή Poisson με βάση τη σχέση (8.1.3.) η οποία έμεινε στη βιβλιογραφία με το όνομα **συνθήκη των Rao-Rubin**. Η συνθήκη αυτή αντικατέστησε την υπόθεση της πλήρους ανεξαρτησίας που έθεσε ο Moran (1952) κατά τον χαρακτηρισμό της κατανομής Poisson με εκείνην της μερικής ανεξαρτησίας των Y και X-Y. Για να αποδείξουν τον χαρακτηρισμό της κατανομής Poisson οι Rao-Rubin χρησιμοποιούν το θεώρημα του Bernstein για τις απόλυτα μονότονες συναρτήσεις. Ο Shanbhag (1974) κατέληξε στο ίδιο συμπέρασμα χρησιμοποιώντας απλούστερη μεθοδολογία, καθώς και οι Srivastava and Singh (1975). Παραλλαγές

(variants) της συνθήκης Rao-Rubin χρησιμοποίησαν οι Patil and Ratnaparkhi (1975) και ο Consul (1975) για να χαρακτηρίσουν την αρνητική διωνυμική κατανομή καθώς και την κατανομή Lagrangian-Poisson. Παραλλαγές της συνθήκης Rao-Rubin χρησιμοποιούν επίσης και οι Patil and Ratnaparkhi (1977) επεκτείνοντας τα συμπεράσματα του προηγούμενου άρθρου τους, ο Gupta (1981) ο οποίος γενικεύει το συμπέρασμα των Patil and Ratnaparkhi (1975) καθώς και ο Kourouklis (1986) ο οποίος αφενός μεν γενικεύει το συμπέρασμα του Gupta (1981) αφετέρου διατυπώνει και αποδεικνύει χαρακτηρισμό της γενικευμένης κατανομής Polya-Eggenberger. Ο Shanbhag (1977) έδωσε μια επέκταση *(extension)* του χαρακτηρισμού της κατανομής Poisson με βάση τη συνθήκη Rao-Rubin. Ο Panaretos (1982a) επεκτείνει αυτά τα συμπεράσματα και στις περικομένες *(truncated)* κατανομές. Γενικεύσεις της συνθήκης Rao-Rubin διατυπώθηκαν και χρησιμοποιήθηκαν για την απόδειξη χαρακτηριστικών ιδιοτήτων κατανομών από την Talwalker (1975) και τους C.R. Rao et al (1980). Ο Panaretos (1982b) μελετά την περίπτωση κατά την οποία η παράμετρος p της διωνυμικής κατανομής της σχέσης (8.1.2) είναι τυχαία μεταβλητή. Εύστοχες παρατηρήσεις για περιπτώσεις όπου το συμπέρασμα των Rao and Rubin (1964) δεν εφαρμόζεται γίνεται από τον Shanbhag (1974) και τους Shanbhag and Panaretos (1979). Προσπάθεια μεταφοράς της συνθήκης των Rao-Rubin στις δυο διαστάσεις γίνεται απο τους Talwalker (1970), Shanbhag (1974), Panaretos (1977), Kourouklis (1986). Τέλος, οι Taillie and Patil (1981) διατυπώνουν ένα συνεχές ανάλογο της συνθήκης των Rao-Rubin και έτσι χαρακτηρίζουν μια μορφή εκθετικής οικογένειας της οποίας μέλος είναι και η κατανομή Gamma.

Από τα αποτελέσματα τα οποία έχουν εμφανισθεί στη βιβλιογραφία στο πλαίσιο του προτύπου διαστρέβλωσης στενότερα συνδεδεμένα με το κεφάλαιο αυτό είναι εκείνα για τα οποία η διαστρέβλωση συνδέεται με underreporting εισοδήματος ή πλούτου. Χαρακτηρισμοί αυτού του τύπου χρησιμοποιούν συχνά κατάλληλα επιλεγμένες συναρτήσεις παλινδρόμησης. Οι συναρτήσεις αυτές βελτιώνουν τις δυνατότητες εφαρμογής των χαρακτηριστικών ιδιοτήτων αφού η πρόσβαση σε πληροφορίες σχετικές με τους regressors και τη regressing τυχαία μεταβλητή είναι ευκολώτερη. Έτσι, θεωρώντας το πολλαπλασιαστικό πρότυπο διαστρέβλωσης $Y = RX$ με R και X ανεξάρτητες τυχαίες μεταβλητές και υποθέτοντας ότι η R ακολουθεί συγκεκριμένη κατανομή πιθανοτήτων, ο Krishnaji (1970) απέδειξε ότι αν η συνάρτηση παλινδρόμησης μιας τυχαίας μεταβλητής Z στην X και στο παρατηρούμενο μέρος αυτής Y είναι γραμμική τότε η κατανομή της αρχικής τυχαίας μεταβλητής X ορίζεται μονοσήμαντα ως Pareto. Η Xekalaki (1984b) απέδειξε ότι ιδιότητα ανάλογη με αυτήν που μελετήθηκε από τον Krishnaji για τη κατανομή Pareto στη συνεχή περίπτωση ισχύει για την κατανομή Yule στη διακριτή περίπτωση. Δηλαδή, μεταξύ των διακριτών κατανομών εισοδήματος η Yule είναι η μόνη κατανομή που διατηρεί τη γραμμικότητα της συνάρτησης παλινδρόμησης. Στο πλαίσιο του προσθετικού προτύπου διαστρέβλωσης $X = Y + U$ και χωρίς να επιβάλλουν περαιτέρω υποθέσεις για το U οι Revankar et al (1974) απέδειξαν ότι η γραμμικότητα της παλινδρόμησης του καταστραμένου μέρους της X επί της αρχικής τιμής, χαρακτηρίζει την κατανομή της X ως Pareto. Η Xekalaki (1989) μελετά το διακριτό ανάλογο της χαρακτηριζουσας συνθήκης των Revankar et al και αποδεικνύει ότι αν το μη παρατηρούμενο μέρος του εισοδήματος παλινδρομεί γραμμικά στη

παρατηρούμενη τιμή τότε η πραγματική κατανομή εισοδήματος ορίζεται μονοσήμαντα ως Yule.

Στο κεφάλαιο αυτό αποδεικνύονται, στο πλαίσιο του προτύπου διαστρέβλωσης, ορισμένα θεωρήματα τα οποία επιτρέπουν τον χαρακτηρισμό γενικών οικογενειών κατανομών. Κατά συνέπεια, συμπεράσματα που ήδη υπάρχουν στη βιβλιογραφία και τα οποία προαναφέρθηκαν, αποδεικνύεται ότι αποτελούν ειδικές περιπτώσεις των γενικών αυτών θεωρημάτων. Τα θεωρήματα καλύπτουν τόσο τη συνεχή όσο και τη διακριτή περίπτωση.

9.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ UNDERREPORTING: ΣΥΝΕΧΗΣ ΠΕΡΙΠΤΩΣΗ.

Στην ενότητα αυτή αποδεικνύουμε δυο θεωρήματα τα οποία στο πλαίσιο του προσθετικού ή πολλαπλασιαστικού προτύπου διαστρέβλωσης. Τα θεωρήματα αυτά αποδεικνύουν ότι η συναρτησιακή μορφή της παλινδρόμησης επι στη X ή/και στη Y μιας κατάλληλα επιλεγμένης τυχαίας μεταβλητής ορίζει μονοσήμαντα την κατανομή της X. Στα δυο θεωρήματα της ενότητας αυτής η X υποτίθεται ότι είναι απόλυτα συνεχής τυχαία μεταβλητή.

Εστω X η πραγματική παρατήρηση η οποία είναι τυχαία μεταβλητή με τιμές στο $(0,+\infty)$. Εστω Y, $Y \leq X$ το παρατηρούμενο μέρος της X και έστω ότι $Y = RX$, όπου R είναι τυχαία μεταβλητή ανεξάρτητη από την X και η οποία κατανέμεται σύμφωνα με την κατανομή power function στο (0,1). Το πρόβλημα το οποίο θα μελετήσουμε είναι η επίδραση στην κατανομή της X της συνάρτησης παλινδρόμησης $E(Z|Y=y)$

μιας τυχαίας μεταβλητής Z (η οποία είναι ανεξάρτητη από την R) επί της Y όταν η συνάρτηση παλινδρόμησης $E(Z|X=x)$ έχει δεδομένη μορφή. Θα εξετασθεί επίσης και το αντίστροφο, δηλαδή η επίδραση της μορφής της $E(Z|Y=y)$ επί της κατανομής της X. Θα αποδειχθεί ότι η κατανομή της X ορίζεται μονοσήμαντα από τη συνάρτηση $E(Z|Y=y)$.

Θεώρημα 8.2.1 Εστω X μια απόλυτα συνεχής τυχαία μεταβλητή με μη εκφυλισμενη κατανομή και έστω

$$h(x) = E(Z|X=x) , \qquad x>0 \tag{8.2.1}$$

όπου Z είναι μια άλλη τυχαία μεταβλητή με αυθαίρετη συνάρτηση κατανομής.
Επιπλέον, έστω R τυχαια μεταβλητή ανεξάρτητη των Z και X της οποίας η συνάρτηση πυκνότητας πιθανότητας δίνεται από την

$$f_R(r) = \begin{cases} \rho r^{\rho-1} & \text{για } 0\le r<1, \quad \rho>0 \\ 0 & \text{οπουδήποτε αλλού} \end{cases} \tag{8.2.2}$$

Τότε οι συναρτήσεις $h(y)$, $y>0$ και

$$\lambda(y) = E(Z|Y=RX=y) \qquad y>0 \tag{8.2.3}$$

προσδιορίζουν μονοσήμαντα την κατανομή της X.

Απόδειξη Εστω $F_W(w)$ και $f_W(w)$ η από κοινού συνάρτηση κατανομής και η από κοινού συνάρτηση πυκνότητας πιθανότητας του τυχαίου διανύσματος $W = (w_1,\ldots,w_n)$, $n\ge1$. Τότε , αν η Z παίρνει τιμές

σέ ένα διάστημα έστω R_Z

$$E(Z|Y=y) = \int_{R_Z} z\, f_{Z|(Y=y)}(z)\, dz$$

δηλαδή,

$$E(Z|Y=y) = \int_{R_Z} z\, \frac{f_{Z,Y}(z,y)}{f_Y(y)}\, dz \ . \qquad (8.2.4)$$

Αλλά,

$$F_{Z,Y}(z,y) = P\ (Z\le z,\ Y\le y)$$

$$= P\ (Z\le z,\ RX\le y)$$

$$= \int_0^\infty P\ (Z\le z,\ RX\le y \mid R=r)\ f_R(r)\, dr.$$

Από την (8.2.2) και επειδή η R ειναι ανεξαρτητη και της X και της Z, η τελευταία εξίσωση γίνεται:

$$F_{Z,Y}(z,y) = \int_0^1 F_{Z,Y}\left(z, \frac{y}{r}\right) \rho r^{\rho-1}\, dr \ .$$

Αρα,

$$f_{Z,Y}(z,y) = \int_0^1 f_{Z,Y}\left(z, \frac{y}{r}\right) \frac{1}{r}\, \rho r^{\rho-1}\, dr$$

$$= \int_y^\infty f_{Z,Y}(z,x)\ \rho \left(\frac{y}{x}\right)^{\rho-2} \frac{y}{x^2}\, dx$$

δηλαδή,

$$f_{Z,Y}(z,y) = \rho y^{\rho-1} \int_y^{\infty} f_{Z,Y}(z,x)\, x^{-\rho}\, dx. \qquad (8.2.5)$$

Κατά συνέπεια συνδυάζοντας τις (8.2.4) και (8.2.5)

$$E(Z|Y=y) = \left(f_Y(y)\right)^{-1} \rho y^{\rho-1} \int_{R_Z} z \left(\int_y^{\infty} f_{Z,X}(z,x)\, x^{-\rho}\, dx \right) dz$$

$$= \left(f_Y(y)\right)^{-1} \rho y^{\rho-1} \int_{R_Z} z \left(\int_y^{\infty} f_{Z|(X=x)}(z,x) f_X(x)\, x^{-\rho} dx \right) dz$$

$$= \left(f_Y(y)\right)^{-1} \rho y^{\rho-1} \int_y^{\infty} \left(\int_{R_Z} z\, f_{Z|(X=x)}(z)\, dz \right) f_X(x)\, x^{-\rho}\, dx\,.$$

Χρησιμοποιώντας την (8.2.1) η τελευταία αυτή εξίσωση γράφεται ισοδύναμα ως εξής:

$$E(Z|Y=y) = \left(f_Y(y)\right)^{-1} \rho y^{\rho-1} \int_y^{\infty} h(x)\, f_X(x)\, x^{-\rho}\, dx. \qquad (8.2.6)$$

Αλλά,

$$F_Y(y) = P(Y \le y)$$

$$= P(RX \le y)$$

$$= \int_0^1 P(RX \le y \mid R=r)\ f_R(r)\ dr$$

$$= \int_0^1 P\left(X \le \frac{y}{r}\right)\ f_R(r)\ dr$$

$$= \int_0^1 F_X\left(\frac{y}{r}\right)\ f_R(r)\ dr\ .$$

Επομένως,

$$f_Y(y) = \int_0^1 f_X\left(\frac{y}{r}\right)\frac{1}{r}\ f_R(r)\ dr$$

$$= \int_y^\infty f_X(x)\ \rho\ \left(\frac{y}{x}\right)^{\rho-2}\ \frac{y}{x^2}\ dx\ .$$

Δηλαδή,

$$f_Y(y) = \rho y^{\rho-1} \int_y^\infty f_X(x)\ x^{-\rho}\ dx\ . \qquad (8.2.7)$$

Επομένως, συνδυάζοντας τις σχέσεις (8.2.6) και (8.2.7) προκύπτει ότι

$$E(Z \mid Y=y) = \frac{\displaystyle\int_y^\infty h(x)\ f_X(x)\ x^{-\rho}\ dx}{\displaystyle\int_y^\infty f_X(x)\ x^{-\rho}\ dx}\ .$$

Αντικαθιστώντας την (8.2.3) στην τελευταία αυτή εξίσωση έχουμε ότι

$$\int_y^\infty h(x)\, f_X(x)\, x^{-\rho}\, dx = \lambda(y) \int_y^\infty f_X(x)\, x^{-\rho}\, dx \ .$$

Παραγωγίζοντας ως προς y προκύπτει ότι

$$-h(y) f_X(y) y^{-\rho} = \lambda'(y) \int_y^\infty f_X(x) x^{-\rho}\, dx - \lambda(y)\, f_X(y)\, y^{-\rho} \ . \qquad (8.2.8)$$

Εστω

$$K(y) = \int_y^\infty f_X(x) x^{-\rho}\, dx \ .$$

Επομένως,

$$K'(y) = -f_X(y)\, y^{-\rho} \ .$$

Αντικαθιστώντας τα K(y) και K'(y) στην (8.2.8) προκύπτει ότι

$$h(y)\, K'(y) = \lambda'(y) K(y) + \lambda(y) K'(y),$$

δηλαδή

$$[h(y) - \lambda(y)] K'(y) = \lambda'(y) K(y).$$

Προφανώς, $h(y)-\lambda(y)\neq 0$ γιατί διαφορετικά θα είχαμε ότι $f_X(x)=0$ το οποίο αντιφάσκει την υπόθεση ότι δηλαδή η κατανομή της X είναι

μη εκφυλισμένη.

Επομένως,

$$\frac{K'(y)}{K(y)} = \frac{\lambda'(y)}{h(y)-\lambda(y)}$$

Η λύση της διαφορικής αυτής εξίσωσης είναι

$$K(y) = C \exp\left\{ \int \frac{\lambda'(y)}{h(y)-\lambda(y)}\, dy \right\}$$

η οποία ισοδύναμα γράφεται

$$\int_y^{\infty} f_X(x)\, x^{-\rho}\, dx = C \exp\left\{ \int \frac{\lambda'(y)}{h(y)-\lambda(y)}\, dy \right\} .$$

Παραγωγίζοντας οδηγούμαστε στη σχέση

$$f_X(y)\, y^{-\rho} = -C \exp\left\{ \int \frac{\lambda'(y)}{h(y)-\lambda(y)}\, dy \right\} \frac{\partial}{\partial y} \left\{ \int \frac{\lambda'(y)}{h(y)-\lambda(y)}\, dy \right\} ,$$

δηλαδή, στην.

$$f_X(y) = C\, y^{\rho}\, \frac{\lambda'(y)}{\lambda(y)-h(y)} \exp\left\{ \int \frac{\lambda'(y)}{h(y)-\lambda(y)}\, dy \right\} . \qquad (8.2.9)$$

Από τη σχέση (8.2.9) είναι προφανές ότι η $f_X(y)$ ορίζεται μονασήμαντα από τις συναρτήσεις παλινδρόμησης $h(y) = E(Z|Y=y)$.

και $\lambda(y) = E(Z|Y=y)$.

Στη συνέχεια θα θεωρήσουμε ειδικές μορφές για τις συναρτήσεις παλινδρόμησης οι οποίες εμπλέκονται στο θεώρημα 8.2.1. Έτσι με τα δύο πορίσματα τα οποία θα διατυπώσουμε θα οδηγηθούμε σε χαρακτηρισμό της κατανομής Pareto αφενός και της κατανομής F του Fisher αφετέρου. Συμπεράσματα ήδη γνωστά στη βιβλιογραφία θα αποδειχθεί ότι αποτελούν ειδικές περιπτώσεις των πορισμάτων αυτών.

<u>Πόρισμα 8.2.1</u> Εστω οι τυχαίες μεταβλητές X,Y και Z που ορίζονται όπως στο θεώρημα 9.2.1. Εστω επίσης $h(x) = E(Z|X=x) = \delta+\beta x^{\alpha}$. Τότε, η συνθήκη $\lambda(y) = E(Z|Y=y) = \delta+\gamma y^{\alpha}$ ορίζει μονοσήμαντα την κατανομή της X ως Pareto, για όλα τα α, με τη προυπόθεση ότι $\alpha\gamma/(\gamma-\beta) > \rho$.

<u>Απόδειξη</u> Θέτοντας όπου $h(y)=\delta+\beta y^{\alpha}$ και $\lambda(y)=\delta+\gamma y^{\alpha}$ η σχέση (8.2.9) γίνεται

$$f_X(y) = Cy^{\rho}\,\frac{\gamma\alpha\, y^{\alpha-1}}{(\gamma-\beta)y^{\alpha}}\,\exp\left\{\int \frac{\gamma\alpha\, y^{\alpha-1}}{(\beta-\gamma)y^{\alpha}}\,dy\right\}$$

$$= C\,\frac{\alpha\gamma\, y^{\rho-1}}{\gamma-\beta}\,\exp\left\{\ln y^{\{\alpha\gamma/(\beta-\gamma)\}}\right\}$$

$$= C\,\frac{\alpha\gamma}{\gamma-\beta}\; y^{-\{\alpha\gamma/(\gamma-\beta)\}+\rho-1}\,.$$

Επομένως, $f_X(y) = \alpha_0 \theta_0^{\alpha_0} y^{-(\alpha_0+1)}$; $y \geq \theta$, $\alpha_0 > 0, \theta > 0$

όπου $C = \alpha_0 = \{\alpha\gamma/(\gamma-\beta)\} - \rho$

και $\theta_0^{\alpha} = \alpha\gamma/(\gamma-\beta)$.

Παρατήρηση 1 Η περίπτωση $\alpha=1$ έχει μελετηθεί από τον Krishnaji (1970) το συμπέρασμα του οποίου προκύπτει σαν ειδική περίπτωση του πορίσματος 8.2.1.

Πόρισμα 8.2.2 Εστω οι τυχαίες μεταβλητές X,Y και Z που ορίζονται όπως στο θεώρημα 8.2.1. Εστω επίσης $h(x) = E(Z|X=x) = \delta_1+\beta x$. Τότε, η συνθήκη $\lambda(y) = E(Z|Y=y) = \delta_0+\gamma y$ ορίζει μονοσήμαντα την κατανομή της X ως F του Fisher με ν_1 και ν_2 βαθμούς ελευθερίας.

Απόδειξη Θέτοντας όπου $h(y)=\delta_1+\beta y$ και $\lambda(y)=\delta_0+\gamma y$ η σχεση (8.2.9) γίνεται

$$f_X(y) = Cy^{\rho} \frac{\gamma}{(\delta_0-\delta_1)+(\gamma-\beta)y} \left[(\delta_1-\delta_0)+(\beta-\gamma)y\right]^{\{\gamma/(\beta-\gamma)\}}$$

$$= C_1 \frac{y^{\rho}}{\left[(\delta_0-\delta_1)+(\beta-\gamma)y\right]^{\{\gamma/(\beta-\gamma)\}}}$$

όπου $C_1 = -C\gamma > 0$.

Η σχέση στην οποία καταλήξαμε έχει τη μορφή της F του Fisher με ν_1 και ν_2 βαθμούς ελευθερίας , όπου

$\nu_1 = \beta-\gamma = 2\rho+2 \in Z^+$

$$\nu_2 = \delta_1 - \delta_0 \in Z^+$$

επίσης,

$$\delta_1 - \delta_0 = \{2\gamma/(\beta-\gamma)\} - (\beta-\gamma)$$

και

$$C_1 = \left\{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right) \Big/ \Gamma(\nu_1/2)\Gamma(\nu_2/2)\right\} \nu_1^{(\nu_1/2)} \quad \nu_2^{(\nu_2/2)} .$$

Κατά συνέπεια η τυχαία μεταβλητή X ακολουθεί την κατανομή F του Fisher στη περίπτωση κατά την οποία οι δοσμένες συναρτήσεις παλινδρόμησης είναι γραμμικές και επίσης $\rho = (\beta - \gamma - 2)/2 > 0$ και $\delta_1 - \delta_0 = \left\{2\gamma/(\beta-\gamma)\right\} - (\beta-\gamma)$

<u>Παρατήρηση 2</u> Εστω $f_X(x;\theta) = x^{-\theta} \exp\{-A(x)+B(\theta)\}$ η συνάρτηση πυκνότητας πιθανότητας της λογαριθμοεκθετικής οικογένειας. Αντικαθιστώντας την $f_X(x;\theta)$ στην σχέση (8.2.9) προκύπτει ότι

$$y^{-\theta} \exp\{-A(x)+B(\theta)\} = C\, y^{\rho}\, D(y) \exp\left\{-\int D(y)\, dy\right\}$$

όπου $D(y) = \dfrac{\lambda'(y)}{h(y)-\lambda(y)}$

Κατά συνέπεια,

$$\ln y^{-\theta} - A(y) + B(\theta) = \ln C + \ln y^{\rho} + \ln D(y) - \int D(y)\, dy$$

Παραγωγίζοντας την εξίσωση αυτή ως προς y προκυπτει οτι

$$-\theta\frac{\theta}{y} - A'(y) = \rho\,\frac{1}{y} + \frac{D'(y)}{D(y)} - D(y)$$

ή

$$(\theta+\rho)\,\frac{1}{y} = D(y) - A'(y) - \frac{D'(y)}{D(y)}$$

Επομένως, είναι προφανές ότι για κάθε συγκεκριμένη μορφή της A(y) έχουμε μια διαφορική εξίσωση ως προς A(y) η λύση της οποίας οδηγεί σε μια μονοσήμαντη σχέση μεταξύ των h(y) και λ(y).

Στη συνέχεια θα μελετήσουμε τη περίπτωση του προσθετικού προτύπου διαστρέβλωσης. Εστω X τυχαία μεταβλητή με μη εκφυλισμένη κατανομή πιθανότητας στο (m,∞) όπου m>0. Θεωρούμε ότι η X παριστάνει τη πραγματική τιμή ενώ η Y ,Y ≤ X το παρατηρούμενό της μέρος. Υποθέτουμε επίσης ότι

$$Y = X - U \quad \text{και} \quad 0 < U < \max(0, X-m). \tag{8.2.10}$$

Τότε, αποδεικνύεται ότι ισχύει το παρακάτω θεώρημα.

<u>Θεώρημα 8.2.2</u> Εστω X,Y και U τυχαίες μεταβλητές που ορίζονται όπως παραπάνω και έστω ότι ισχύει η (8.2.10). Τότε οι συναρτήσεις

$$h(x) = E(U|X=x) \qquad x > m \tag{8.2.11}$$

και

$$g(y) = E(U|X>y) \tag{8.2.12}$$

ορίζουν μονοσήμαντα την κατανομή της X.

<u>Απόδειξη</u> Εστω $F_W(w)$ και $f_W(w)$ η συνάρτηση κατανομής και η συνάρτηση πυκνότητας πιθανότητας αντίστοιχα, ενός τυχαίου διανύσματος $W = (w_1,\ldots,w_n)$, $n \geq 1$. Τότε αν η U παίρνει τιμές σε ένα διάστημα έστω R_U έχουμε

$$g(y) = E(U|X>y) = \int_{R_U} u\, f_{U|(X>y)}(u)\, du \ .$$

Επομένως,

$$\int_y^\infty f_X(x)\, g(y)\, dx = \int_{R_U} u \left(\int_y^\infty f_{U,X}(u,x)\, dx \right) du$$

ή

$$\left(\int_y^\infty f_X(x)\, dx \right) g(y) = \int_{R_U} u \left(\int_y^\infty f_{U|(X=x)}(u)\, f_X(x)\, dx \right) du \ .$$

Η σχέση αυτή ειναι ισοδύναμη με την

$$\left[1-F_X(y)\right] g(y) = \int_y^\infty f_X(x) \left(\int_{R_U} u\, f_{U|(X=x)}(u)\, du \right) dx \ .$$

Αντικαθιστώντας την (8.2.11) η τελευταία αυτή εξίσωση γράφεται

$$\left[1-F_X(y)\right]\ g(y) = \int_y^{\infty} h(x)\ dF_X(x)\ .$$

Επομένως,

$$\left[1-F_X(y)\right]\ g(y) = \int_m^{\infty} h(x)\ dF_X(x) - \int_m^{\infty} h(x)\ dF_X(x)$$

$$= E\left[h(x)\right] + \int_m^{\infty} h(x)\ d\left[1-F_X(x)\right]\ .$$

Ολοκληρώνοντας κατά μέρη και επειδή $F_X(m)=0$

$$\left[1-F_X(y)\right]\ g(y) = E\left[h(x)\right] + \left[h(x)\left[1-F_X(x)\right]\right]_m^y - \int_m^y \left[1-F_X(x)\right]\ h'(x)\ dx$$

$$= E\left[h(x)\right] + h(y)\left[1-F_X(y)\right] - h(m) - \int_m^y \left[1-F_X(x)\right]h'(x)dx\ .$$

Παραγωγίζοντας ως προς y προκύπτει ότι

$$\bar{F}_X'(y)\ g(y) + \bar{F}_X(y)\ g'(y) = \bar{F}_X'(y)\ h(y) + \bar{F}_X(y)\ h'(y) - \bar{F}_X'(y)\ h'(y)$$

όπου $\bar{F}_X(y) = 1-F_X(y)$.

Επομένως,

$$\bar{F}_X'(y)\left[g(y)-h(y)\right] = -\bar{F}_X(y)\ g(y).$$

προφανώς $g(y)-h(y) \neq 0$, διαφορετικά θα ήταν $\bar{F}_X(y)=0$ για κάθε $y>0$ το οποίο θα μπορούσε να οδηγήσει σε εκφυλισμένη κατανομή για τη X. Επομένως,

$$\frac{\bar{F}'_X(y)}{\bar{F}_X(y)} = \frac{g'(y)}{g(y)-h(y)}$$

Η λύση της διαφορικής αυτής εξίσωσης είναι

$$\bar{F}_X(y) = C \exp\left\{ \int \frac{g'(y)}{g(y)-h(y)}\, dy \right\} \qquad (8.2.13)$$

ή ισοδύναμα

$$f_X(y) = C \frac{g'(y)}{g(y)-h(y)} \exp\left\{ \int \frac{g'(y)}{h(y)-g(y)}\, dy \right\} \qquad (8.2.14)$$

Από τη σχέση (8.2.14) είναι προφανές ότι η $f_X(x)$ ορίζεται μονοσήμαντα από τις συναρτήσεις $g(y) = E(U|X=y)$ και $h(y) = E(U|X>y)$.

Το πόρισμα το οποίο ακολουθεί αποτελεί ειδική περίπτωση του θεωρήματος 8.2.2 και οδηγεί σε χαρακτηρισμό της κατανομής Pareto. Αποδεικνύεται δε ότι θεώρημα ήδη γνωστό στη βιβλιογραφία αποτελεί ειδική περίπτωση του πορίσματος 8.2.3.

Πόρισμα 8.2.3 Εστω οι τυχαίες μεταβλητές X,Y και U που ορίζονται όπως στο θεώρημα 8.2.2. Εστω επίσης $h(x) = E(U|X=x) = \delta+\beta x^{\alpha}$. Τότε, η συνθήκη $g(y) = E(U|X>y) = \delta+\gamma y^{\alpha}$ ορίζει μονοσήμαντα τη κατανομή της X σαν Pareto για όλα τα α.

Απόδειξη Θέτοντας όπου $g(y)=\delta+\beta y^{\alpha}$ και $g(y)=\delta+\gamma y^{\alpha}$ η σχέση (8.2.14) γίνεται:

$$f_X(y) = C\,\frac{\gamma\alpha y^{\alpha-1}}{(\gamma-\beta)y^{\alpha}}\exp\left\{\int\frac{\gamma\alpha y^{\alpha-1}}{(\gamma-\beta)y^{\alpha}}\,dy\right\}$$

$$= C\,\frac{\alpha\gamma y^{-1}}{\gamma-\beta}\,y^{-\{\alpha\gamma/(\gamma-\beta)\}}$$

$$= C\,\frac{\alpha\gamma}{\gamma-\beta}\,y^{-\{\alpha\gamma/(\gamma-\beta)\}-1}.$$

Δηλαδή,

$$f_X(y) = \alpha_0\, y^{-(\alpha_0+1)} \qquad y \geq 1,\ \alpha_0 > 0$$

όπου $\alpha_0^{*} = \alpha\gamma/(\gamma-\beta)$

και $C = 1.$

Παρατήρηση 3 Η περίπτωση $\alpha=1$ έχει μελετηθεί από τους Revankar et al (1974) το συμπέρασμα των οποίων μπορεί να θεωρηθεί σαν ειδική περίπτωση του πορίσματος 8.2.3.

Παρατήρηση 4 Εστω $F_X(y)$ η συνάρτηση επιβίωσης της κατανομής Pareto τρίτου είδους, δηλαδή

$$\bar{F}_X(y) = \theta(y+\delta)^{-\alpha}\, e^{-\beta\gamma} \qquad \alpha,\theta > 0,\ \beta \geq 0 \qquad \delta \in R,\ y \geq 1 \ . \qquad (8.2.15)$$

Αντικαθιστώντας την (8.2.15) στην (8.2.13) προκύπτει ότι

$$\theta(y+\delta)^{-\alpha}\, e^{-\beta y} = C \exp\left\{ \int \frac{g'(y)}{h(y)-g(y)}\, dy \right\}$$

$$\ln\theta - \alpha\ln(y+\delta) - \beta y = \ln C + \int \frac{g'(y)}{h(y)-g(y)}\, dy \ .$$

Παραγωγίζοντας ως προς y έχουμε ότι

$$-\alpha \frac{1}{\gamma+\delta} - \beta = \frac{g'(y)}{h(y)-g(y)}$$

ή ισοδύναμα

$$(y+\delta)g'(y) + \alpha h(y) - \alpha g(y) + \beta(y+\delta)h(y) - \beta(y+\delta)g(y) = 0 \ .$$

Θέτοντας $Q(y) = \left[\frac{\alpha}{y+\delta} + \beta\right]$, η τελευταία εξίσωση γίνεται

$$\frac{dg(y)}{g(y)} - Q(y)g(y) = -Q(y)h(y)$$

η οποία είναι γραμμική διαφορική εξίσωση πρώτης τάξης ως προς g(y). Επομένως,

$$g(y)\, e^{\int Q(y)dy} = -\int Q(y)h(y)\, e^{\int Q(y)dy}\, dy + C\ .$$

Επειδή δε,

$$\int Q(y)dy = \ln(y+\delta)^{-\alpha} - \beta y$$

η εξίσωση μετασχηματίζεται στην

$$g(y)(y+\delta)^{-\alpha}\, e^{-\beta y} = -\alpha\int h(y)(y+\delta)^{-\alpha-1}\, e^{-\beta y}\, dy$$

$$-\beta\int h(y)(y+\delta)^{-\alpha}\, e^{-\beta y}\, dy \qquad (8.2.16)$$

Είναι επομένως προφανές ότι υπάρχει μια σχετικά απλή σχέση μεταξύ g(y) και h(y) συναρτήσει των παραμέτρων της κατανομής Pareto , και είναι ενδιαφέρον να παρατηρήσουμε ότι για $\beta=0$ (η περίπτωση της κατανομής Pareto δευτέρου είδους) η σχέση (8.2.16) γίνεται

$$g(y)(y+\delta)^{-\alpha} = -\alpha\int h(y)\,(y+\delta)^{-(\alpha+1)}\, dy + C$$

ενώ, για $\beta=0$, $\delta=0$ (περίπτωση της κατανομής Pareto πρώτου είδους) η σχέση (8.2.16) απλοποιείται ακόμη περισσότερο και γίνεται

$$g(y)\ y^{-\alpha} = -\alpha \int h(y)\ y^{-(a+1)}\ dy + C\ .$$

8.3 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ OVERREPORTING: ΣΥΝΕΧΗΣ ΠΕΡΙΠΤΩΣΗ.

Εστω ότι η X δηλώνει την πραγματική παρατήρηση και η Y το παρατηρούμενό της μέρος. Η ενότητα αυτή αντιμετωπίζει την περίπτωση προβλημάτων, στο πλαίσιο πάντα του προτύπου διαστρέβλωσης όπου η παρατηρούμενη τιμή Y είναι μια υπερεκτίμηση της πραγματικής παρατήρησης X, δηλαδή $Y \geq X$. Προσθετικό πρότυπο αυτού του τύπου μελετήθηκε από τον Panareto (1983) ο οποίος το ονομάζει generating model. Στο πλαίσιο του μοντέλου αυτού. αλλά θεωρώντας τόσο τη προσθετική όσο και τη πολλαπλασιαστική μορφή του προτύπου αποδεικνύουμε ότι η συναρτησιακή μορφή της παλινδρόμησης στη X ή στη Y μιας κατάλληλα επιλεγμένης τυχαίας μεταβλητής ορίζει μονοσήμαντα την κατανομή της X η οποία και πάλι εδώ θεωρείται ότι είναι απόλυτα συνεχής. Πριν προχωρήσουμε στην απόδειξη των κυρίως συμπερασμάτων της ενότητας θα αποδείξουμε το παρακάτω λήμμα.

<u>Λήμμα 8.3.1</u> Εστω X τυχαία μεταβλητή με απόλυτα συνεχή συνάρτηση κατανομής και έστω R τυχαία μεταβλητή ανεξάρτητη από την X με συνάρτηση πυκνότητας πιθανότητας

$$f_R(r) = \begin{cases} \rho r^{\rho-1} & \text{για} \quad 0 < r \leq 1, \ \rho > 0 \\ 0 & \text{οπουδήποτε αλλού} \end{cases} \qquad (8.3.1)$$

Υποθέτουμε ότι $P\left[Y = \frac{X}{R} < x_0\right] > 0$ για κάποιο $x_0 > 0$. Τότε η κατανομή του $\frac{X}{R}$ περικομένη από τα δεξιά στο x_0 συμπίπτει με την κατανομή της X αν και μόνο αν η X ακολουθεί κατανομή power στο $(0, x_0)$.

<u>Απόδειξη</u> Χρησιμοποιώντας τον συμβολισμό της ενότητας 8.2 έχουμε ότι

$$F_Y(y) = P(Y \leq y) = P(\frac{X}{R} \leq y)$$

$$= \int_0^1 P\left(\frac{X}{R} \leq y \,|\, R=r\right) \, dF_R(r)$$

$$= \int_0^1 P(X \leq yr) \, dF_R(r)$$

δηλαδή,

$$F_Y(y) = \int_0^1 F_X(yr) \, dF_R(r) \, . \qquad (8.3.2)$$

Παρατηρούμε ότι η σχέση (8.3.2) ισχύει για οποιαδήποτε κατανομή της R με μόνη προυπόθεση το πεδίο τιμών της να περιέχεται στο $(0,1)$.

<u>Ικανό</u> Εστω ότι η X κατανέμεται σύμφωνα με την κατανομή power

στο (0,x). Τότε,

$$\bar{F}_X(x) = \begin{cases} \left(\frac{x}{x_0}\right)^{\alpha} & \text{για } x<x_0,\ x_0>0,\ \alpha>0 \\ 1 & \text{οπουδήποτε αλλού} \end{cases} \qquad (8.3.3)$$

και κατά συνέπεια,

$$\bar{F}_X(xyx_0) = \left(\frac{xyx_0}{x_0}\right)^{\alpha} = \left(\frac{xx_0}{x_0}\right)^{\alpha}\left(\frac{yx_0}{x_0}\right)^{\alpha}$$

δηλαδή,

$$\bar{F}_X(xyx_0) = \bar{F}_X(xx_0)\ \bar{F}_X(yx_0)\ . \qquad (8.3.4)$$

Από τη σχέση (8.3.2) έχουμε,

$$F_Y(x_0) = \int_0^1 F_X(x_0 r)\ dF_R(r)\ . \qquad (8.3.5)$$

Επίσης για κάθε $y<x_0$ και επειδή ισχύουν οι σχέσεις (8.3.2) και (8.3.5) προκύπτει ότι,

$$F_Y(y)-F_Y(x_0)F_X(y) = \int_0^1 F_X(yr)\ dF_R(r) - \int_0^1 F_X(x_0 r)F_X(y)\ dF_R(r)$$

ή ισοδύναμα,

$$F_Y(y)-F_Y(x_0)F_X(y) = \int_0^1 \left[F_X(yr)-F_X(x_0r)F_X(y)\right] dF_R(r) \ . \qquad (8.3.6)$$

Από την (8.3.4) προκύπτει ότι

$$F_X(yr) = F_X\left[\frac{y}{x_0} x_0\right] F_X(rx_0)$$

δηλαδή,

$$F_X(yr) = F_X(y) \ F_X(rx_0).$$

Κατά συνέπεια το δεύτερο μέλος της σχέσης (8.3.6) είναι ίσο με το μηδέν και επομένως

$$\frac{F_X(y)}{F_Y(x_0)} = F_X(y) \ . \qquad (8.3.7)$$

Επομένως,

$$P\left(Y = \frac{X}{R} \le y \mid Y<x_0\right) = \frac{P(Y \le y, \ Y <x_0)}{P(Y<x_0)}$$

$$= \frac{F_Y(y)}{F_Y(x_0)}$$

$$= F_X(y) \ .$$

Αναγκαίο Συνδυάζοντας τις σχέσεις (8.3.1) και (8.3.2) προκύπτει ότι

$$F_Y(y) = \int_0^1 F_X(yr)\ f_R(r)\ dr$$

$$= \rho\int_0^1 F_X(yr)\ r^{\rho-1}\ dr$$

$$= \rho\int_0^y F_X(x)\ \left(\frac{x}{y}\right)^{\rho-1}\ \frac{1}{y}\ dx$$

δηλαδή,

$$F_Y(y) = \rho y^{-\rho}\int_0^y F_X(x)\ x^{\rho-1}\ dx$$

ή ισοδύναμα

$$(1/\rho)\ F_Y(y)\ y^{\rho} = \int_0^y F_X(x)\ x^{\rho-1}\ dx.$$

Παραγωγίζοντας ως προς y προκύπτει,

$$\frac{\partial}{\partial y}\left[(1/\rho)\ F_Y(y)\ y^{\rho}\right] = F_X(y)\ y^{\rho-1}\ .$$

Χρησιμοποιώντας τη σχέση (8.3.7) η τελευταία αυτή εξίσωση γράφεται

$$\frac{\partial}{\partial y}\left[(1/\rho)\ F_Y(x_0)\ F_X(y)\ y^{\rho}\right] = F_X(y)\ y^{\rho-1}\ . \qquad (8.3.8)$$

Θέτοντας στη σχέση (8.3.8) όπου $F_X(y)\ y^{\rho} = K(y)$ προκύπτει ότι

$$(1/\rho)F_X(x_0)\ \frac{\partial}{\partial y}\ K(y) = K(y)\ \frac{1}{y}\ .$$

Δηλαδή,

$$\frac{K'(y)}{K\ (y)} = \frac{\rho}{yF_X(x_0)}\ .$$

Λύνοντας τη διαφορική αυτή εξίσωση προκύπτει ότι:

$$F_X(y) = C\ y^{\rho\left(\frac{1}{F_Y(x_o)}\ -1\right)}\ .$$

Επομένως, το λήμμα αποδείχτηκε.

Ας θεωρήσουμε τώρα το πολλαπλασιαστικό generating μοντέλο όπου X αντιπροσωπεύει την πραγματική παρατήρηση και Y το παρατηρούμενο μέρος αυτής το οποίο είναι τουλάχιστον ίσο με το X. Ειδικώτερα ας υποθέσουμε ότι η σχέση μεταξύ των X και Y καθορίζεται απο την $Y = X/R$ όπου R είναι τυχαία μεταβλητή ανεξάρτητη από την X και ακολουθεί την κατανομή power στο διάστημα (0,1). Τότε αποδεικνύεται ότι ισχύει το παρακάτω θεώρημα.

Θεώρημα 8.3.1 Εστω Z τυχαία μεταβλητή με αυθαίρετη κατανομή και έστω X τυχαία μεταβλητή με απόλυτα συνεχή, μη εκφυλισμένη κατανομή πιθάνοτητας. Εστω R τυχαία μεταβλητή, ανεξάρτητη των Z και X με συνάρτηση πυκνότητας πιθανότητας που δίνεται από την

$$f_R(r) = \begin{cases} \rho r^{\rho-1} & \text{όταν } 0<\rho<1,\ \rho>0 \\ 0 & \text{οπουδήποτε αλλού.} \end{cases}$$

Τότε οι συναρτήσεις,

$$h(y) = E\left(Z \mid X = x\right)$$

και

$$\lambda(y) = E\left(Z \mid Y = \frac{X}{R} = y\right)$$

προσδιορίζουν μονοσήμαντα την κατανομή της X.

Απόδειξη Με επιχειρήματα ανάλογα με εκείνα που χρησιμοποιήθηκαν για την απόδειξη του θεωρήματος 8.2.1 καταλήγουμε ότι

$$E\left(Z \mid Y=y\right) = \left[f_Y(y)\right]^{-1} \rho\, y^{-(\rho+1)} \cdot \int_0^y h(x)\, f_X(x)\, x^{\rho}\, dx.$$

Επιπλέον,

$$f_Y(y) = \rho\, y^{-(\rho+1)} \int_0^y f_Y(x)\, x^{\rho}\, dx.$$

Επομένως,

$$\int_0^y h(x)\, f_X(x)\, x^{\rho}\, dx = \lambda(y) \int_0^y f_X(x)\, x^{\rho}\, dx.$$

Παραγωγίζοντας ως προς y και λύνοντας τη διαφορική εξίσωση που προκύπτει, δεδομένου ότι η X έχει μη εκφυλισμένη κατανομή, προκύπτει ότι

$$f_X(x) = C\, y^{-\rho}\, \frac{\lambda'(y)}{h(y)-\lambda(y)}\, \exp\left\{\int \frac{\lambda'(y)}{h(y)-\lambda(y)}\, dy\right\}. \qquad (8.3.9)$$

Επομένως, η $f_X(x)$ καθορίζεται μονοσήμαντα από τις $\lambda(y)$ και $h(y)$.

Θα αποδείξουμε τώρα υπό μορφή πορίσματος μια χρήσιμη χαρακτηριστική ιδιότητα της κατανομής Pareto. Το πόρισμα προκύπτει από το θεώρημα 8.3.1 στην ειδική περίπτωση που οι συναρτήσεις παλινδρόμησης είναι γραμμικές.

Πόρισμα 8.3.1 Εστω οι τυχαίες μεταβλητές X,Y και Z οι οποίες ορίζονται όπως στο θεώρημα 8.3.1. Εστω επίσης ότι $E(Z|X=x) = \alpha+\delta x$, $\alpha,\delta \in \mathbb{R}$. Τότε $E(Z|Y=y) = \alpha+\beta y$, $\beta \in \mathbb{R}$ αν και μόνον αν η X ακολουθεί κατανομή Pareto με την προυπόθεση ότι $\beta/(\delta-\beta) < \rho$.

Απόδειξη Θέτοντας όπου $h(y)=\alpha+\delta y$ και $\lambda(y)=\alpha+\beta y$ η σχέση (8.3.9) γίνεται:

$$f_X(x) = C\, y^{-\rho}\, \frac{\beta}{(\delta-\beta)\, y}\, \exp\left\{\int \frac{\beta}{(\delta-\beta)\, y}\, dy\right\}$$

$$= C\, \frac{\beta}{\delta-\beta}\, y^{\left[\beta/(\delta-\beta)\right]-\rho-1}.$$

Δηλαδή, η X ακολουθεί την κατανομή Pareto με παραμέτρους $(\rho+a)^{-1/a}$ και a όπου $a = \left[\beta/(\delta-\beta)\right] - \rho$ και $C = a$ με την προυπόθεση ότι $\beta/(\delta-\beta) < \rho$.

Ας θεωρήσουμε τέλος την περίπτωση κατά την οποία έχουμε υπερεκτίμηση της πραγματικής τιμής στο πλαίσιο του προσθετικού όμως generating model. Αποδεικνύεται τότε, ότι ισχύει το παρακάτω θεώρημα.

Θεώρημα 8.3.2 Εστω Y,X και U τυχαίες μεταβλητές που ορίζονται όπως στο θεώρημα 8.2.2 και έστω ότι $Y = X+U$. Τότε οι συναρτήσεις

$h(x)=E(U|X=x)$

και

$g(y)=E(U|X<y)$

ορίζουν μονοσήμαντα την κατανομή της X.

Απόδειξη Ακολουθώντας τα βήματα της απόδειξης του θεωρήματος 8.2.2 καταλήγουμε ότι:

$$F_X(y) = C \exp\left\{ \int \frac{g'(y)}{h(y)-g(y)} dy \right\}.$$

Επομένως, η κατανομή της X προσδιορίζεται μονοσήμαντα από τις συναρτήσεις h(y) και g(y).

8.4 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ UNDERREPORTING: ΔΙΑΚΡΙΤΗ ΠΕΡΙΠΤΩΣΗ.

Στην ενότητα αυτή θα διατυπώσουμε και θα αποδείξουμε θεωρήματα τα οποία, στο πλαίσιο είτε του προσθετικού είτε του πολλαπλασιαστικού προτύπου διαστρέβλωσης, αποδεικνύουν ότι και στην περίπτωση που η τυχαία μεταβλητή X είναι διακριτή, η μορφή της συνάρτησης παλινδρόμησης στη X ή/και στη Y μιας κατάλληλα επιλεγμένης τυχαίας μεταβλητής ορίζει μονοσήμαντα την κατανομή της X.

Εστω X η πραγματική παρατήρηση η οποία υποθέτουμε ότι είναι μη αρνητική τυχαία μεταβλητή που παίρνει ακέραιες τιμές. Εστω επίσης Y, $Y \leq X$ το παρατηρούμενο μέρος της X και έστω ότι $Y = [RX]$ όπου R τυχαία μεταβλητή ανεξάρτητη από την X και η οποία κατανέμεται ομοιόμορφα στο διάστημα (0,1). Με $[\alpha]$ συμβολίζουμε το ακέραιο μέρος του α. Με το θεώρημα 8.4.1 αποδεικνύεται ότι η κατανομή της X ορίζεται μονοσήμαντα από τη συνάρτηση παλινδρόμησης $E(Z|Y=y)$ μιας τυχαίας μεταβλητής Z (η οποία κατανέμεται ανεξάρτητα από την R) επί της Y όταν η συνάρτηση παλινδρόμησης $E(Z|X=x)$ έχει δοσμένη μορφή.

Θεώρημα 8.4.1 Εστω X τυχαία μεταβλητή με ακέραιες και θετικές τιμές της οποίας η κατανομή είναι μη εκφυλισμένη. Ορίζουμε τη συνάρτηση

$$h(x) = E(Z|X=x) \quad x=1,2,... \tag{8.4.1}$$

όπου Z είναι τυχαία μεταβλητή με αυθαίρετη συνάρτηση κατανομής.

Εστω Y τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές τέτοια ώστε Y=[RX] όπου R είναι τυχαία μεταβλητή ανεξάρτητη από τις X και Z και η οποία κατανέμεται ομοιόμορφα στο διάστημα (0,1). Τότε η συνάρτηση

$$\lambda(y) = E(Z|Y=y) \quad y=0,1,2,... \tag{8.4.2}$$

προσδιορίζει μονοσήμαντα την κατανομή της X, αν $\lambda(y_1)\neq\lambda(y_2)$ όταν $y_1\neq y_2$.

<u>Απόδειξη</u> Εστω ότι p_x, q_y είναι οι συναρτήσεις πιθανότητας των X και Y αντίστοιχα. Χωρίς να βλάπτουμε τη γενικότητα ας υποθέσουμε ότι η τυχαία μεταβλητή Z παίρνει ακέραιες τιμές. Τότε με βάση τον ορισμό της τυχαίας μεταβλητής Y προκύπτει ότι:

$$E(Z|Y=y) = \sum_z z\ P(Z=z|Y=y)$$

$$= \frac{1}{q_y} \sum_z z\ P(Z=z, [RX]=y)$$

$$= \frac{1}{q_y} \sum_z z\ P(Z = z, y \le RX < y+1)$$

$$= \frac{1}{q_y} \sum_z z \sum_{x=y+1}^{\infty} P(Z = z,\ \frac{y}{x} \le R < \frac{y+1}{x} \mid X=x)\ p_x$$

$$= \frac{1}{q_y} \sum_z z \sum_{x=y+1}^{\infty} P(Z = z \mid X = x)\, P\left(\frac{y}{x} \le R < \frac{y+1}{x}\right) p_x$$

$$= \frac{1}{q_y} \sum_{x=y+1}^{\infty} \frac{p_x}{x} \sum_z z\, P(Z=z \mid X=x).$$

Δηλαδή,

$$E(Z \mid Y=y) = \frac{1}{q_y} \sum_{x=y+1}^{\infty} h(x)\, \frac{p_x}{x}.$$

Χρησιμοποιώντας τη σχέση (8.4.2) προκύπτει ότι

$$\lambda(y) = \frac{1}{q_y} \sum_{x=y+1}^{\infty} h(x)\, \frac{p_x}{x}. \qquad (8.4.3)$$

Αλλά,

$$I_y = P(Y=y)$$

$$= P\Big([Rx] = y\Big)$$

$$= P(y \le RX < y+1)$$

$$= \sum_{x=y+1}^{\infty} P(y \le RX < y+1 | X=x)\ p_x$$

$$= \sum_{x=y+1}^{\infty} P\left(\frac{y}{x} \le R < \frac{y+1}{x} \right) p_x \ .$$

Δηλαδή,

$$q_y = \sum_{x=y+1}^{\infty} \frac{p_x}{x} \ . \qquad (8.4.4)$$

Συνδυάζοντας τις σχέσεις (8.4.3) και (8.4.4) προκύπτει,

$$\lambda(y) \sum_{x=y+1}^{\infty} \frac{p_x}{x} = \sum_{x=y+1}^{\infty} h(x)\ \frac{p_x}{x}$$

ή ισοδύναμα,

$$\sum_{x=y+1}^{\infty} \frac{p_x}{x} = \frac{1}{\lambda(y)} \sum_{x=y+1}^{\infty} h(x)\ \frac{p_x}{x} \ . \qquad (8.4.5)$$

Εξειδικεύοντας τη σχέση (8.4.5) για y=r και y=r+1 και αφαιρώντας τις εξισώσεις που προκύπτουν έχουμε ότι,

$$\sum_{x=y+1}^{\infty} \frac{p_x}{x} = \frac{\lambda(r+1)-h(r+1)}{\lambda(r)-\lambda(r+1)} \frac{p_{r+1}}{r+1}, \qquad r=0,1,2,\ldots .$$

Εφαρμόζοντας την ίδια τεχνική φθάνουμε σε μια εξίσωση της μορφής

$$A_{r+1}Q_{r+1} - (A_r+1)Q_r = 0 , \qquad r=0,1,2,\ldots \qquad (8.4.6)$$

όπου

$$A_r = \frac{\lambda(r+1)-h(r+1)}{\lambda(r)-\lambda(r+1)}$$

$$Q_r = \frac{p_{r+1}}{r+1} \qquad r=0,1,2,\ldots .$$

Από τη σχέση (8.4.6) προκύπτει ότι $\lambda(r+1)-h(r+1)\neq 0$, γιατί αλλιώς θα είχαμε ότι $p_r=0$, $r=0,1,\ldots$ το οποίο έρχεται σε αντίφαση με την υπόθεση που έχουμε θέσει, ότι δηλαδή η κατανομή της X είναι μη εκφυλισμένη. Επομένως, η σχέση (8.4.6) παίρνει τη μορφή

$$Q_{r+1} - \frac{A_{r+1}}{r+1} Q_r = 0 , \qquad r=0,1,2,\ldots$$

ή ισοδύναμα,

$$p_{r+1} - \frac{[\lambda(r-1)-h(r)][\lambda(r+1)-\lambda(r)](r+1)}{[\lambda(r+1)-h(r+1)][\lambda(r)-\lambda(r-1)]r} p_r = 0.$$

Η μοναδική λύση της εξίσωσης αυτής διαφορών δίνεται από τη σχέση

$$P(X=r)=P(X=1)\ r\ \frac{\lambda(r)-\lambda(r-1)}{\lambda(1)-\lambda(0)}\ \prod_{i=1}^{r-1}\ \frac{\lambda(i-1)-h(i)}{\lambda(i+1)-h(i+1)}\quad r=2,3,\ldots \qquad (8.4.7)$$

Είναι επομένως προφανές ότι η p_x ορίζεται μονοσήμαντα από τις συναρτήσεις $h(x) = E(Z|X=x)$ και $\lambda(y) = E(Z|Y=y)$.

Το πόρισμα που ακολουθεί βασίζεται στο θεώρημα 8.4.1 και οδηγεί σε ένα πολύ χρήσιμο χαρακτηρισμό της κατανομής Yule. Εδώ τίθεται σαν υπόθεση η γραμμικότητα των εμπλεκομένων συναρτήσεων παλινδρόμησης.

<u>Πόρισμα 8.4.1</u> Εστω οι τυχαίες μεταβλητές X, Y και Z που ορίζονται όπως στο θεώρημα 8.4.1. Εστω επίσης ότι

$$h(x) = E(Z|X=x) = \alpha+\beta x, \qquad x=1,2,\ldots$$

Τότε η συνθήκη,

$$\lambda(y) = E(Z|Y=y) = \gamma+(\gamma-\alpha)y, \quad y=0,1,2,..$$

ορίζει μονοσήμαντα την κατανομή της X σαν Yule(ρ), όπου $\rho = \beta/(\gamma-\alpha-\beta)$, με την προυπόθεση ότι $\beta/(\gamma-\alpha-\beta) > 0$.

Απόδειξη Θέτοντας όπου $h(x)=\alpha+\beta x$ και $\lambda(y)=\gamma+(\gamma-\alpha)y$ η σχέση (8.4.7) γίνεται

$$P(X=r) = P(X=1)\, r \prod_{i=1}^{r-1} \frac{(\gamma-\alpha-\beta)i}{[2\gamma-2\alpha-\beta+(\gamma-\alpha-\beta)i]}$$

ή ισοδύναμα,

$$P(X=r) = P(X=1) \prod_{i=1}^{r-1} \frac{i+1}{\rho+2+i}$$

όπου $\rho = \dfrac{\beta}{\gamma-\alpha-\beta}$.

Δηλαδή,

$$P(X=r) = P(X=1) \frac{r!}{(\rho+3)_{(r-1)}} .$$

Επειδή δε,

$$\sum_{r=1}^{\infty} P(X=r) = 1$$

η παραπάνω εξίσωση γίνεται

$$1 = P(X=1) \sum_{r=0}^{\infty} \frac{(r+1)!}{(\rho+3)_{(r)}}$$

ή ισοδύναμα,

$$1 = P(X=1) \sum_{r=0}^{\infty} \frac{2_{(r)} 1_{(r)}}{(\rho+3)_{(r)}} \frac{1}{r!} .$$

Από τη τελευταία αυτή εξίσωση προκύπτει ότι

$$P(X = 1) = \frac{\rho}{\rho+2} .$$

Η τυχαία μεταβλητή λοιπόν X κατανέμεται σαν Yule (ρ) όπου $\rho = \beta/(\gamma-\alpha-\beta)$.

<u>Παρατήρηση 1</u> Την περίπτωση την οποία διαπραγματεύεται το πόρισμα 8.4.1 έχει μελετήσει η Xekalaki (1984). Κατά συνέπεια το ήδη δημοσιευμένο συμπέρασμα της μπορεί να θεωρηθεί σαν ειδική περίπτωση του θεωρήματος 8.4.1.

Στη συνέχεια θα μελετήσουμε την προσθετική περίπτωση του προτύπου διαστρέβλωσης. Εστω X τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές της οποίας η κατανομή είναι μη εκφυλισμένη στο $\{m, m+1, \ldots\}$, $m \in I^+ \cup \{0\}$. Θεωρούμε ότι η τυχαία αυτή μεταβλητή αντιπροσωπεύει την πραγματική παρατήρηση ενώ η τυχαία μεταβλητή Y, $Y \leq X$ αντιστοιχεί στο παρατηρούμενο μέρος αυτής. Επιπλέον υποθέτουμε ότι

$$Y = X - U \quad \text{και} \quad 0 \le U \le \max(0, X-m). \tag{8.4.8}$$

Τότε αποδεικνύεται ότι ισχύει το παρακάτω θεώρημα:

Θεώρημα 8.4.2 Εστω οι τυχαίες μεταβλητές X,Y και U που ορίζονται όπως πιο πάνω και έστω ότι ισχύει η σχέση (8.4.8). Εστω επίσης ότι

$$h(x) = E(U|X=x) \qquad x = m, m+1, \ldots \tag{8.4.9}$$

Τότε η συνάρτηση

$$g(y) = E(U|X \ge y) \qquad y = m, m+1, \ldots \tag{8.4.10}$$

ορίζει μονοσήμαντα την κατανομή της X, με την προυπόθεση ότι $g(y_1) \ne g(y_2)$ για $y_1 \ne y_2$.

Απόδειξη Εστω p_x η συνάρτηση πιθανότητας της X. Τότε,

$$g(y) = E(U|X \ge y)$$

$$= \sum_u u \, P(U|X \ge y)$$

$$= \sum_u u \, \frac{P(U=u, \; X \ge y)}{P(X \ge y)}$$

$$= \sum_{u} u \sum_{x=y}^{\infty} \frac{P(U=u,\ X=x)}{P(X \geq y)}$$

$$= \sum_{x=y}^{\infty} \sum_{u} u \frac{P(U=u|X=x)\ p_x}{P(X \geq y)} .$$

Δηλαδή,

$$g(y)\ P(X \geq y) = \sum_{x=y}^{\infty} p_x \sum_{u} u\ P(U=u|X=x)$$

ή ισοδύναμα,

$$g(y)\ P(X \geq y) = \sum_{x=y}^{\infty} p_x\ h(x) \quad . \qquad (8.4.11)$$

Εξειδικεύοντας τη σχέση (8.4.11) για y=r και y=r+1 και αφαιρώντας τις προκύπτουσες εξισώσεις προκύπτει ότι

$$g(r+1)\ P(X \geq r+1) - g(r)\ P(X \geq r) = -\ P(X=r)\ h(r) .$$

Κατά συνέπεια,

$$[g(r+1) - h(r)]\ P(X \geq r+1) - [g(r) - h(r)]\ P(X \geq r) = 0.$$

Είναι προφανές ότι $g(r+1)\neq g(r)$ γιατί αλλιώς η κατανομή της X θα ήταν εκφυλισμένη και αυτό αντιβαίνει τις υποθέσεις που έχουμε κάνει αναφορικά με την τυχαία μεταβλητή X.

Η μοναδική λύση της παραπάνω εξίσωσης διαφορών υπό τον περιορισμό $P(X\geq m) = 1$ δίνεται από τη σχέση

$$P(X\geq y) = \prod_{r=m}^{y-1} \frac{g(r)-h(r)}{g(r+1)-h(r)} . \qquad (8.4.12)$$

Επομένως το θεώρημα αποδείχτηκε.

Ας εφαρμόσουμε τώρα τη χαρακτηρίζουσα συνθήκη του θεωρήματος 8.4.2 στην ειδική περίπτωση που οι συναρτήσεις $h(x) = E(U|X=x)$ και $g(y) = E(U|X\geq y)$ είναι γραμμικές. Θα οδηγηθούμε έτσι σε έναν ενδιαφέροντα χαρακτηρισμό της περικομένης κατανομής Yule.

<u>Πόρισμα 8.4.2</u> Εστω X,Y και U τυχαίες μεταβλητές που ορίζονται όπως στο θεώρημα 8.4.2. Εστω επίσης ότι

$$h(x) = E(U|X=x) = \alpha+\beta x$$

Τότε η συνάρτηση

$$g(y) = E(U|X\geq y) = \gamma+\delta y$$

ορίζει μονοσήμαντα την κατανομή της X σαν (m-1)-περικομένη Yule με συνάρτηση πιθανότητας την

$$P(X=x) = \rho \frac{(m+1)_{(x-m)}}{(1+\rho+m)_{(x-m+1)}}, \quad x = m, m+1, \ldots \qquad (8.4.13)$$

με την προυπόθεση ότι $\gamma-\alpha = \delta-\beta$.

Απόδειξη Θέτοντας όπου $h(x)=\alpha+\beta x$ και $g(y)=\gamma+\delta y$ η σχέση (8.4.12) γίνεται

$$P(X \geq r) = \prod_{i=m}^{r-1} \frac{(\gamma-\alpha)+(\delta-\beta)i}{(\gamma-\alpha)+\delta+(\delta-\beta)i}$$

ή ισοδύναμα,

$$P(X \geq r) = \left(\frac{\gamma-\alpha}{\delta-\beta} + m\right)_{(r-m)} \Big/ \left(\frac{\gamma-\alpha+\delta}{\delta-\beta} + m\right)_{(r-m)} . \qquad (8.4.14)$$

Κατά συνέπεια στην περίπτωση που $\gamma-\alpha = \delta-\beta$ από τη σχέση (8.4.14) προκύπτει ότι

$$P(X \geq r) = \frac{(m+1)_{(r-m)}}{(1+\rho+m)_{(r-m)}}$$

$$= (\rho+1+r) \frac{(m+1)_{(r-m)}}{(1+\rho+m)_{(r-m+1)}} .$$

Δηλαδή,

$$P(X \geq r) = [1 + \frac{1}{\rho}(r+1)]\ \rho\ \frac{(m+1)_{(r-m)}}{(1+\rho+m)_{(r-m+1)}} \quad ..$$

Αλλά αυτή είναι, σύμφωνα με συμπέρασμα της Xekalaki (1984), η ικανή και αναγκαία συνθήκη για να ακολουθεί η X την (m-1)-περικομένη κατανομή Yule με συνάρτηση πιθανότητας την (8.4.13).

<u>Παρατήρηση 2</u> Την περίπτωση με την οποία ασχολείται το πόρισμα 8.4.2 μελετάει η Xekalaki (1989). Κατά συνέπεια το ήδη δημοσιευμένο αποτέλεσμά της μπορεί να θεωρηθεί σαν ειδική περίπτωση του θεωρήματος 8.4.2.

8.5 ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΙΔΙΟΤΗΤΕΣ ΣΤΟ ΠΛΑΙΣΙΟ ΤΟΥ OVERREPORTING: ΔΙΑΚΡΙΤΗ ΠΕΡΙΠΤΩΣΗ.

Εστω X η πραγματική παρατήρηση και Y το παρατηρούμενό της μέρος. Η ενότητα αυτή αντιμετωπίζει την περίπτωση που η καταγραφόμενη τιμή Y είναι μια υποεκτίμηση της πραγματικής παρατήρησης Y. Ειδικότερα υποθέτουμε ότι $Y = \left[\frac{X}{R}\right]$ όπου R είναι τυχαία μεταβλητή η οποία κατανέμεται ομοιόμορφα στο διάστημα (0,1) και η οποία είναι ανεξάρτητη από την X. Τότε, το παρακάτω θεώρημα εξετάζει την επίδραση της κατανομής της X στη συνάρτηση

παλινδρόμησης μιας τυχαίας μεταβλητής Z στην Y όταν η μορφή της συνάρτησης παλινδρόμησης της Z στην X είναι δοσμένη.

Θεώρημα 8.5.1 Εστω X τυχαία μεταβλητή με ακέραιες και θετικές τιμές της οποίας η κατανομή είναι μη εκφυλισμένη.
Ας θεωρήσουμε τη συνάρτηση

$$h(x) = E(Z|X=x), \qquad x=1,2,... \tag{8.5.1}$$

όπου Z είναι τυχαία μεταβλητή με αυθαίρετη κατανομή.
Εστω Y τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές τέτοια ώστε $Y = \left[\frac{X}{R}\right]$ όπου R είναι τυχαία μεταβλητή ανεξάρτητη από τις X και Z η οποία κατανέμεται ομοιόμορφα στο (0,1). Τότε η συνάρτηση

$$\lambda(y) = E(Z|Y=y) \qquad y=0,1,2,... \tag{8.5.2}$$

προσδιορίζει μονοσήμαντα την κατανομή της X, με την προυπόθεση ότι $\lambda(y_1)\neq\lambda(y_2)$ για $y_1\neq y_2$.

Απόδειξη Εστω $p_x = P(X=x)$ η κατανομή πιθανότητας της X. Από τον ορισμό της $E(Z|Y=y)$ προκύπτει ότι

$$\lambda(y) = \frac{1}{q_y}\,\frac{1}{y(y+1)}\sum_{x=1}^{y} x\,h(x)\,p_x$$

όπου,

$$q_y = P\left(\left[\frac{X}{R}\right] = y\right)$$

$$= P\left(y \le \frac{X}{R} < y+1\right)$$

$$= \sum_{x=1}^{y} P\left(\frac{x}{y+1} < R \le \frac{x}{y}\right) p_x$$

$$= \frac{1}{y(y+1)} \sum_{x=1}^{y} x\, p_x \ .$$

Κατά συνέπεια,

$$\lambda(y) \sum_{x=1}^{y} x\, p_x = \sum_{x=1}^{y} x\, h(x)\, p_x \qquad y=0,1,2,\ldots \ . \qquad (8.5.3)$$

Εξειδικεύοντας τη σχέση (8.5.3) για y=r και y=r+1 και αφαιρώντας τις εξισώσεις που προκύπτουν έχουμε ότι

$$\sum_{x=1}^{r} x\, p_x = \frac{\lambda(r+1)-h(r+1)}{\lambda(r)-\lambda(r+1)} (r+1)\, p_{r+1} \ . \qquad (8.5.4)$$

Επαναλαμβάνοντας την ίδια διαδικασία για r και r-1, από τη σχέση (8.5.4) προκύπτει ότι

$$p_{r+1} - \frac{[\lambda(r)-\lambda(r+1)][\lambda(r-1)-h(r)]r}{[\lambda(r+1)-h(r+1)][\lambda(r-1)-\lambda(r)](r+1)} p_r = 0$$

αφού $\lambda(r+1)\neq h(r+1)$ σαν συνέπεια των υποθέσεων που έχουμε θέσει για την κατανομή της X.

Η μοναδική λύση της εξίσωσης αυτής διαφορών δίνεται από τη σχέση,

$$P(X=r)=P(X=1)\frac{\lambda(r)-\lambda(r+1)}{\lambda(0)-\lambda(1)}\frac{1}{r}\prod_{i=1}^{r-1}\frac{\lambda(r-1)-h(r)}{\lambda(r+1)-h(r+1)}, \quad r=2,3,\ldots . \quad (8.5.5)$$

Κατά συνέπεια η p_x ορίζεται μονοσήμαντα από τις συναρτήσεις παλινδρόμησης $h(x)=E(Z|X=x)$ και $\lambda(y)=E(Z|Y=y)$.

Το θεώρημα 8.5.2 αντιμετωπίζει το πρόβλημα του overreporting στο πλαίσιο του προσθετικού προτύπου διαστρέβλωσης.

Θεώρημα 8.5.2 Εστω X τυχαία μεταβλητή με ακέραιες και μη αρνητικές τιμές της οποίας η κατανομή είναι μη εκφυλισμένη στο $\{m,m+1,\ldots\}$, $m \in I^+$. Εστω επίσης οι τυχαίες μεταβλητές Y και U με ακέραιες και μη αρνητικές τιμές οι οποίες είναι τέτοιες ώστε

$Y = X + U$ και $0 \le U \le X + m$.

Τότε οι συναρτήσεις

$$g(y) = E(U|X \le y)$$

και

$$h(x) = E(U|X=x)$$

ορίζουν μονοσήμαντα την κατανομή της X.

Απόδειξη Εστω $p_x = P(X=x)$ η συνάρτηση πιθανότητας της X. Τότε,

$$g(y) = E(U|X \le y)$$

$$= \sum_u u \frac{P(U=u, X \le y)}{P(X \le y)}$$

$$= \frac{1}{P(X \le y)} \sum_{x=m}^{y} P(X=x) \sum_u u\, P(U=u|X=x)$$

$$= \frac{1}{P(X \le y)} \sum_{x=m}^{y} p_x\, E(U|X=x) \ .$$

Δηλαδή,

$$g(y)\ P(X\le y) = \sum_{x=m}^{y} P(X=x)\ h(x)\ . \qquad (8.5.6)$$

Εξειδικεύοντας τη σχέση (8.5.6) για y=r και y=r+1 και αφαιρώντας τις εξισώσεις που προκύπτουν έχουμε ότι

$$[g(r)-h(r)]\ P(X\le r) - [g(r-1)-h(r)]\ P(X\le r-1) = 0, \quad r=m,m+1,\ldots\ .$$

Είναι προφανές ότι $g(r)\neq h(r)$ γιατί αλλιώς η κατανομή της X θα ήταν εκφυλισμένη και αυτό αντιβαίνει τις υποθέσεις που έχουμε θέσει για την X. Η μοναδική λύση της παραπάνω εξίσωσης διαφορών υπό τον περιορισμό ότι $P(X\ge m) = 1$ δίνεται από την

$$P(X\le y) = P(X=m) \prod_{r=m}^{y-1} \frac{g(r-1)-h(r)}{g(r)-h(r)}\ .$$

Κατά συνέπεια η κατανομή της X ορίζεται μονοσήμαντα από τις συναρτήσεις g(r) και h(r).

Ε Π Ι Λ Ο Γ Ο Σ

Οι Engel και Zijlstra (1980), μελετώντας προβλήματα προσφοράς και ζήτησης, διετύπωσαν και απέδειξαν χαρακτηριστικές ιδιότητες της εκθετικής κατανομής από ήδη υπάρχοντες χαρακτηρισμούς της γεωμετρικής κατανομής. Οδηγήθηκαν δε στο συμπέρασμα ότι ορισμένες χαρακτηριστικές ιδιότητες των κατανομών αυτών οι οποίες σχετίζονται με τη δομή της συνάρτησης κατανομής τους μπορούν να θεωρηθούν αντίστοιχες. Επίσης οι Xekalaki και Panaretos (1988) οδηγούνται σε χαρακτηρισμούς της κατανομής Pareto οι οποίοι είναι αντίστοιχοι με ήδη υπάρχοντες χαρακτηρισμούς της κατανομής Yule. Και στην περίπτωση αυτή οι χαρακτηριστικές ιδιότητες απορρέουν από τη δομή των συναρτήσεων κατανομών. Τέλος ο Πανάρετος (1988) αποδεικνύει ότι και χαρακτηριστικές ιδιότητες οι οποίες συνδέονται με την έννοια της ανεξαρτησίας στατιστικών συναρτήσεων ή της δεσμευμένης αναμενόμενης τιμής μπορούν να θεωρηθούν αντίστοιχες για τα δύο ζευγάρια κατανομών που προαναφέραμε. Τα συμπεράσματα της διατριβής εδραιώνουν αυτή τη σχέση δυισμού που υπάρχει μεταξύ των κατανομών Pareto και Yule. Τέλος η επιτυχής μεταφορά αρκετών συμπερασμάτων στο χώρο των δυο διαστάσεων υποδηλώνει πως υπάρχει αντίστοιχη σχέση μεταξύ διδιάστατων μορφών των κατανομών αυτών και συγκεκριμένα μεταξύ της διδιάστατης Pareto τύπου Mardia I έτσι όπως αυτή ορίζεται από την (2.3.8) και της διδιάστατης Yule έτσι όπως αυτή ορίζεται από την (2.8.3).

Β Ι Β Λ Ι Ο Γ Ρ Α Φ Ι Α

Ahsanullah,M. and Kabir,A.B.M. (1973) A characterization of the Pareto distribution. *The Canadian Journal of Statistics,* 1, 109-112.

Ahsanullah,M Houchens,R.L. (1989) A note on record values from a Pareto distribution. *The Pakistan Journal of Statistics,* 5, 51-57.

Aitchison,J. and Brown,J.A.C. (1957) *The Lognormal Distribution,* Cambridge University Press, London.

Arnold.B.C. (1982) *Pareto Distributions.* International Co-operative Publ. House, Fairland, Maryland USA.

Arnold,B.C. and Press,J. (1983) Bayesian inference for Pareto Populations. *Journal of Econometrics,* 21, 287-306.

Arnold,B.C. (1987) Bivariate distributions with Pareto conditionals. *Statistics & Probability Letters,* 5, 263-266.

Arnold,B.C. and Press,J. (1989) Bayesian Estimation and Prediction for Pareto Data. *Journal of the American Statistical Association,* 84, 1079-1084.

Adnan, Awad.M. (1987) Characterization of distributions through maximization of alpha-entropies. *Journal of Mathematical Control and Information.* 1, 1-7.

Basu,A.P. (1971) Bivariate failure rate. *Journal of the American Statistical Association,* 66, 103-104.

Basu,A.P. and Block,H.W. (1975) On characterizing univariate and multivariate exponential distributions with applications. In *Statistical Distributions in Scientific Work,* Vol 3, G.P. Patil, S. Kotz and J.K. Ord (eds.), D. Reidel Publ. Co., Dordrecht, Holland, 399-421.

Beg,M.I. and Kirmani,S.N.U.A. (1974) On a characterization of exponential and related distributions. *The Australian Journal of Statistics,* 16, 163-6.

Van Belle,G. and Schneiderman,M. (1973) Some statistical aspects of environmental pollution and protection. *Int. Stat. Rev., 41,315-331.*

Berk,R.H. (1977) Characterizations via conditional distributions. *J. Appl. Prob.* 14, 806-816.

Bissinger,B.H. (1965) A type resisting distribution generated from considerations of an inventory decision model. In *Classicaland Contagious Discrete Distributions,* G.P. Patil (ed.), Pergamon Press and Statistical Publishing Society, Calcutta, 15-17.

Blitz,R.C. and Brittain,J.A. (1964) An extension of the Lorenz diagram to the correlation of two variables. *Metron,* 23, 137-143.

Block,H.W. (1977a) A family of bivariate life distributions. In *The Theory and Applications of Reliability* , Vol.1, C.P. Tsokos and I. Shimi (eds.) Academic Press, New York, 349-371.

Block,H.W. (1977b) Multivariate reliability classes. In *Applications in Statistics,* P.R. Krishnaiah, (ed.), North Holland, 79-88.

Block,H.W. and Basu,A.P. (1974) A continuous bivariate exponential extension. *Journal of the American Statistical Association,* 69, 1031-1037.

Block,H.W. and Savits,T.H. (1980a) Multivariate distributions in reliability theory and life testing. In *Statistical Distributions inScientific Work,* Vol. 5, C. Taillie, G.P. Patil and B.A. Baldessari (eds.), Holland: D. Reidel Publ. Co.

Block,H.W. and Savits,T.H. (1980b) Multivariate increasing failure rate average distributions. *The Annals of Probability,* 8 (4), 793-801.

Block,H.W. and Savits,T.H. (1981a) Multivariate classes of life distributions in reliability theory. *Mathematics and Operations Research,* 6 (3), 453-461.

Block,H.W. and Savits,T.H. (1980b) Multidimensional IFRA Processes. *The Annals of Probability,* 9 (1), 162-166.

Brindley,E.C. and Thompson,W.A. (1972) Dependence and aging aspects of multivariate survival. *Journalof the American Statistical Association,* 67, 822-830.

Bryson,M.C. and Siddiqui,M.M. (1969) Some criteria for aging. *Journal of the American Statistical Association,* 64, 1472-1483.

Brown,M. (1972) Low density traffic systems. *Advances in Appl. Probbility,* 4, 177-192.

Consul,P.C. (1975) Some new characterizations of discrete Lagrangian distributions. In *Statistical Distributions in Scientific Work, Vol 3.* G.P. Patil, S. Kotz, J.K. Ord (eds.), D. Reidel Publ. Co. 270-290.

Cook,R.D. and Martin,F.B. (1974). A model for quadrat sampling with "visibility bias", *J. Amer. Statist. Assoc.,* 69, 345-349.

Cox,D.R. (1962) *Renewal Theory.* Frome and Long: Butler & Tannner, Ltd.

Cox,D.R. (1969) Some sampling problems in technology. In *New Developments in Survey Sampling,* (Johnson and Smith, Eds). 289-316. University of North Carolina Press.

Dallas,A.C. (1976) Characterizing the Pareto and power distributions. *Annals of Institute of Statistical Mathematics,* 28,491-7.

Dimaki,C. (1977) Characterizations and estimation procedures for the continuous Pareto distributions. Unpublished M.Sc. thesis, Department of Statistics, University of Warwick, England.

Δημάκη,Κ. (1988a) Χαρακτηριστικές ιδιότητες της κατανομής Pareto. *Πρακτικά Διήμερου Κλασικής και Μπεϋζιανής Στατιστικής,* Ελληνικό Στατιστικό Ινστιτούτο, Αθήνα, Φεβρουάριος 1988.

Dimaki,C. (1988b) Some characterizations of the Pareto distribution by order statistics. *Επιστημονική Επετηρίδα,* Ανωτάτη Σχολή Οικονομικών και Εμπορικών Επιστημών, Αθήνα.

Dimaki,C. and Xekalaki,E. (1990) Identifiability of Income Distributions in the Context of Damage and Generating Models, *Communications in Statistics - Part A: Theory and Methods,* 19(8), 2757-2766.

imaki,C. and Xekalaki,E. (1991a) Properties of Probability Distributions Used in Income Analysis. (submitted).

imaki,C. and Xekalaki,E. (1991b) On a refinement of Kotlarski's theorem. (submitted).

imaki,C. and Xekalaki,E. (1991c) Characterizations of the Pareto Distributions based on Order Statistics. (submitted).

ıgel,J. and Zijlstra,M. (1980) A characterization of the gamma distribution by the negative binomial distribution. *Journal of Applied Probability,* 17, 1138-1144.

ɔstein,B. and Sobel,M. (1953) Life testing. *Journal of the American Statistical Association,* 48, 486-502.

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

ary,J.D. and Marshall,A.W. (1979) Multivariate distributions with increasing hazard rate average. *The Annalsof Probability,* 7 (2), 359-370.

ller,W. (1966) *Introduction to Probability Theory and Applications,* 2, 10-14, Wiley & Sons, New York.

erguson,T.S. (1967) On characterizing distributions by properties of order statistics. *Sankhya, Series A,* 29, 265-278.

isher,R.A. (1934) The effect of methods of ascertainement upon the estimation of frequencies, *Ann. Eugen.* 6, 13-25.

isz,M. (1958) Characterization of some probability distributions. *Skand. Aktuariedskr.,* 1-2, 65-70.

riday,D.S. and Patil,G.P. A bivariate exponential model with applications to reliability and computer generation of random variables. In *The Theory and Applications of Reliability* , Vol. 1, C.P. Tsokos and I.N. Shimi (eds.), Academic Press, New York, 527-549.

ılambos,J. (1972) Characterization of certain populations by independence of order statistics. *Journal of Applied Probability,* 9, 224-230.

ılambos,J. (1975) Characterizations of probability distributions by properties of order statistics. In *Statistical Distributions in Scientific Work,* G.P. Patil et al (editors), Dordrecht-Holland, D. Reidel.

ılambos,J. (1978) *The asymptotic Theory of Extreme Order Statistics,* J.Wiley, New York.

Galambos,J. and Kotz,S. (1978) *Characterizations of Probability Distributions.* Lecture Notes in Mathematics #675, Springer-Verlag, Berlin.

Govindarajulu,Z. (1966) Characterization of the exponential and power distribution . *Skand. Aktuarietidskr.*, 49, 132-136.

Gumbel,E.J. (1960) Bivariate exponential distributions. *J. Amer. Statist. Assoc.*,55, 698-707.

Gupta,R.C. (1980) On the mean residual life function in survival studies. In *Statistical Distributions in Scientific Work,* Vol. 5, C. Taillie, G.P. Patil and B.A. Baldessari (eds.), Holland: D. Reidel Publ. Co.

Gupta,R.C. (1981) On the Rao-Rubin characterization of the Poisson distribution. In *Statistical Distributions in Scientific Work, Vol 4,* C. Taillie et al (eds.), D. Reidel Publ. Co., 341-347.

Hagstroem,K.G. (1925) La loi de Pareto et la reassurance. *Skand. Aktuar.*, 8, 65-88.

Haldane,J.B.S. (1938) The estimation of the frequency of recessive conditions in man, *Ann. Eugen.* 7,255-262.

Hamdan,M.A. (1972) On a characterization by conditional expectations. *Technometrics,* 14, 497-499.

nis,C.L. and Chakraborty,R. (1984) Non-random sampling in human genetics: Familial correlation, *J. Math. Appl. Med. & Biol.*, 1, 193-213.

rris,R. (1970) A multivariate definition for increasing hazrd rate distribution functions. *The Annals of Mathematical Statistics*, 41 (2), 713-717.

ang,J.S. (1975) A note on order statistics from a Pareto distribution. *Scandinavisk Actuarietidskr.*, 3, 187-190.

win,J.O. (1975) The generalised Waring distribution. *J. R. Statist. Soc., Series A*, 18-31 (Part I), 204-227 (Part II), 374-384 (Part III).

nardan,K.G. and Rao,B.R. (1984) Cerrtain inter-relationships between two classes of probability models for ascertainement of families in segregation analysis. *J. Ind. Statist. Assoc.*, Vol. 22.

be,D.G. (1969) On characterizing the multivariate linear exponential distribution. *Canad. Math. Bull.*, 12 (5), 567-572.

lbfleisch,J.D. and Prentice,R.L. (1980) *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York.

ndall,M.G. (1961) Natural Law in the Social Sciences. *Journal of the Royal Statistical Society*, A, 124, 1-16.

ndall,M.G. and Moran,P.A.P. (1963) *Geometrical Probability*, Charles Griffin & Co., Ltd., London.

tlarski,I.I. (1975) On characterization of probability distributions by conditional expectations. *Statistical Distributions in Scientific Work, Vol. 3*, G.P. Patil et al (eds). D. Reidel Publ. Co., Holland.

urouklis,S. (1986) Characterizations of some discrete distributions based on a variant of the Rao-Rubin condition, *Communications in Statistics, Part A: Theory and Methods*, !5(3), 839-851.

ishnaji,N. (1970) Characterization of the Pareto distribution through a model of uderreported incomes. *Econometrica*, 38, 251-255.

ishnaji,N. (1974) Characterization of some discrete distributions based on a damage model, *Sankhya, Ser. A*, 36, 204-213.

urent,A.G. (1974) On characterization of some distributions by trunction properties. *Journal of the American Statistical Association*, 69, 823-827.

Laurent,A.G. (1975) Failure and mortality from wear and ageing. The Teissier model. In *Statistical Distributions in Scientific Work,* Vol. 2, G.P. Patil, S. Kotz and J.K. Ord (eds.), Reidel, Dordrecht and Boston, 301-320.

Lin,Gwo-Dong (1988) Characterization of distributions via relationships between two moments of order statistics. *Journal of Statistical Planning and Inference,* 19, 73-80.

Mahfoud,M. and Patil,G.P. (1981) Size biased sampling, weighted distributions, and Bayesian estimation, In *Statistics in Theory and Practice: Essays in Honor of Bertil Matern.* Bo Ranneby ,(ed.) Umea, Sweeden : Swedish Univ. Agri. Sci., 173-187.

Mahfoud, M. and Patil,G.P. (1982) On weighted distributions, In *Statistics and Probability: Essays in Honor of C.R. Rao,* G. Kallianpur, P.R. Krishnaji, J.K. Ghosh (eds), North Hollland, Amsterdam, 479-492.

Malik,H.J. (1970) A characterization of the Pareto distribution. *Skandinavisk Aktuarietidskrift,* 53, 115-117.

Mann,N.R., Schafer,R.E. and Singpurwalla,N.D. (1974) *Methods for Statistical Analysis of Reliability and Life Data.* John Wiley & Sons, New York.

Mardia,K.V. (1962) Multivariate Pareto distributions, *Ann. Math. Statist.*, 33, 1008-1015.

Mardia,K.V. (1964) Some results on the order statistics of the multivariate normal and Pareto type I populations. *Annals of Mathematical Statistics*, 35, 1815-1818.

Mardia,K.V. (1970) *Families of Bivariate Distributions.* Charles Griffin and Co. Ltd., London.

Marshall,A.W. (1975) Multivariate distributions with monotone hazard rate. In *Reliability and Fault Tree Analysis*, R. Barlow, J. Fussel and N. Singpurwalla (eds.), SIAM, Philadelphia, 259-284.

arshall,A.W. and Olkin,I. (1967) A multivariate exponential distribution, *J. Amer. Statist. Assoc.*, 62, 30-44.

athai,A.M. (1967) On the structural properties of the conditional distributions, *Can. Math. Bull.*, 10, 239-245.

cKay,A.T. (1934) Sampling from batches. *J. R. Statist. Soc., Ser. B*, 1, 207-216.

enon,M.V. (1966) Characterization theorems for some univariate probability distributions, *J. R. Statist. Soc. Ser. B*, 28, 143-145.

Moran,P.A.P. (1952) A characteristic property of the Poisson distribution, *Proc. Camb. Phil. Soc. Maths. Phys. Sci.*, 48, 206-207.

Morgenstern,D. (1956) Einfache Beispiele zweidimensionaler Verteilungen. *Mitt. für Math. Stat.*, 8,21-29.

Morrison,D.G. (1978) On linearly increasing mean residual lifetimes. *Journal of Applied Probability*, 15, 617-620.

Muth,E.G. (1977) Reliability models with positive memory derived from the mean residual life function. In *The Theory and Applications of Reliability*, Vol II, C.P. Tsokos and I.N. Shimi (eds.), Academic Press, New York, 401-435.

Neel and Schull (1966) *Human Heredity*, 211-227. University of Chicago Press, Chicago.

Ord,J.K. (1972) *Lectures notes on analysis of spatial patterns and ecological relations*. Advanced Institute on Statistical Ecology in the United States, 241.

Ord,J.K. (1975) Statistical models for personal income distributions. In *Statistical Distributions in Scientific Work*, 2, G.P. Patil, S. Kotz and J.K. Ord (eds.), Reidel, Dordrecht and Boston, 151-158.

Panaretos,J. (1977) A characterization of a general class of multivariate discrete distributions, *Colloquia Mathematica Societatis Janos Bolyai 21. Analytic Function Methods in Probability Theory,* Debrecen (Hungary), 243-252.

Panaretos,J. (1981) On the joint distribution of two discrete random variables. *Ann. Inst. Stat. Maths.*, 33, 191-198.

Panaretos,J. (1982a) On characterizing some discrete distributions using an extension of the Rao-Rubin theorem, *Sankhya, Ser. A*, 44, 415-422.

Panaretos,J. (1982b) An extension of the damage model, *Metrika*, 29, 189-194.

Panaretos,J. (1982c) On a structural property of finite distributions. *J. R. Statist. Soc. Ser. B*, 26, 209-211.

Panaretos,J. (1983) A Generating Model Involving Pascal and Logarithmic Series Distributions. *Communications in Statistics, Part A: Theory and Methods*, 12, 841-848.

Panaretos,J. (1988) Απεικονίσεις χαρακτηριστικών ιδιοτήτων κατανομών μέσω ανελίξεων Poisson. *Πρακτικά Διήμερου Κλασσικής και Μπεϋζιανής Στατιστικής*, Ελληνικό Στατιστικό Ινστιτούτο, Αθήνα, Φεβρουάριος 1988.

Patil,G.P. (1981) Studies in statistical ecology involving weighted distributions. In *Statistics Applications and New Directions: Proceedings of ISI Golden Jubille International Conference,* J.K. Ghosh and J. Roy (eds.), Calcutta, India, Statistical Publ. Soc., 478-503.

Patil,G.P., Boswell,M.T. and Friday,D.S. (1975) Chance mechanism in computer generation of random variables. In *Statistical Distributions in Scientific Work,* 2, G.P. Patil, S.Kotz and J.K. Ord (eds.) Reidel, Dordrecht and Boston, 37-50.

Patil,G.P. and Ord,J.K. (1975) On size-biased sampling and related form-invariant weighted distributions, *Sankhya, 37, 48-61.*

Patil,G.P. and Rao, C.R. (1977) Weighted distributions: a survey of their applications, In *Applications of Statistics,* P.R. Krishnaiah, (ed.), North-Holland, Amsterdam, 383-405.

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.

Patil,G.P. and Ratnaparkhi,M.V. (1975) Problems of damaged random variables and related characterizations. In *Statistical Distributions in Scientific Work, Vol. 3,* G.P. Patil, S. Kotz, J.K. Ord (eds.), D. Reidel Publ. Co., 225-270.

Patil,G.P. and Ratnaparkhi,M.V. (1977) Characterizations of certain statistical distributions based on additive damage models involving Rao-Rubin condition and some of its variants, *Sankhya, Ser. B,* 39(1), 65-75.

Patil,G.P. and Seshadri,V. (1963) A characterization of a bivariate distribution by the marginal and the conditional distributions of the same component. *Ann. Inst. Statist. Math. Tokyo,* 15, 215-221.

Patil,G.P. and Seshadri,V. (1964) Characterization theorems for some univariate probability distributions, *J.R. Statistic. Society, Ser. B,* 26, 286-292.

Pielou,E.C. (1969) *An Introduction to Mathematical Ecology,* Wiley, New York, 147.

Rao,C.R. (1965) On discrete distributions arising out of methods of ascertainement. In *Classical and Contagious Discrete Distributions,* G.P. Patil (ed.), Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.

Rao,C.R. (1985) Weighted distributions arising out of methods of ascertainement: What population does a sample , *The 151 Centenary Volume,* A.C. Atkinson and S.E. Fienberg (eds.), 543-569.

Rao,C.R. and Rubin,H. (1964) On a characterization of the Poisson distribution, *Sankhya, Ser. A.* 26, 295-298.

Rao,C.R., Srivastava,R.C., Talwalker,S. Edgar,G.A. (1980) Characterization of probability on a generalised Rao-Rubin condition, *Sankhya, Series A,* 42(3), 161-169.

Reinhardt,H.E. (1968) Characterizing the exponential distribution, *Biometrics,* 24, 437-438.

Revankar,N.S., Hartley,M.J. and Pagano,M. (1974) A Characterization of the Pareto Distribution, *The Annals of Statistics,* 2, 599-601.

Rogers,G.S. (1963) An alternative proof of the characterization of the density ax^b. *American Mathematical Monthly,* 70, 857-8.

Rosenzweig,M., Zelen,M., VonHoff,D.D., et al (1978) Waiting for a bus: Does it explain age-dependent differences in response to chemotherapy of early breast cancer, *New England J. Med.,* 299, 1363-1364.

Rossberg,H.J. (1972) Characterization of the exponential and the Pareto distributions by means of some properties of the distributions which the differences and quotients of order statistics are subject to. *Math. Operationsforschung und Statistik,* 3, 207-216.

Roy,D. and Mukherjee,S.P. (1986) A note on characterizations of the Weibull distribution. *Sankhya, Series A,* 48 (2), 250-253.

Samanta,M. (1972) Characterization of the Pareto distribution. *Skandinavisk Aktuarietidskrift,* 55, 191-192.

Scheaffer,R.L. Size biased sampling, *Technometrics,* 14, 635-644.

Schotz,W.E. and Zelen,M. (1971) Effect of lenght sampling bias on labeled mitotic index waves, *J. Theor. Biol.,* 32, 383-404.

Schultz,D.M. (1975) Mass-size distributions- A review and a proposed new model. In *Statistical Distributions in Scientific Work,* 2, G.P. Patil, S. Kotz and J.K. Ord (eds.), Reidel, Dordrecht and Boston, 275-288.

Shanbhag,D.N. (1970) The characterizations for exponential and geometric distributions. *J. Amer. Statist. Assoc.,* 65,1256-1259.

Shanbhag,D.N. (1974) An elementary proof for the Rao-Rubin characterization of Poisson distribution, *Journal of Applied Probability*, 11, 211-215.

Shanbhag,D.N. (1977) An extension of the Rao-Rubin characterization of the Poisson distribution, *Journal of Applied Probability*, 14, 640-646.

Shanbhag,D.N. (1989) Review of "Pareto Distributions" by B.C.Arnold. *Journal of the Royal Statistical Society*, A, 152, 253-254.

Shanbhag,D.N. and Panaretos,J. (1979) Some results related to the Rao-Rubin characterization of the Poisson distribution, *Austral. J. Statist.* 21(1), 78-83.

Simon,H.A. (1955) On a class of skew distribution functions. *Biometrika*, 42, 425-440.

Simon,R (1975) Length biased sampling in the estimation of antigen frequencies. National Cancer Institute. A preliminary paper.

Simon,R. (1980) Length biased sampling in etiologic studies. *Amer. J. Epidemiol.*, 111, 444-452.

Smart,R.J. (1963) Alcoholism, birth order, and family size, *J. Abnorm. Soc. Psychology*, 66, 17-23.

Srivastava,M.S. (1965) A characterization of Pareto's distribution and $(k+1)x^k/\theta^{k+1}$. *Annals of Mathematical Statistics,* 36, 361-362.

Srivastava,R.C. and Singh,J. (1975) On some characterizations of the binomial and Poisson distributions based on a damage model, In *Statistical Distributions in Scientific Work, Vol. 3,* G.P. Patil, S. Kotz, J.K. Ord (eds.) D. Reidel Publ. Co., 271-278.

Stene,J. (1981) Probability distributions arising from the ascertainment and the analysis of data on human families and other groups, In *Statistical Distributions in Scientific Work,* 6, C. Taillie, G.P. Patil and B.A. Baldessari (eds.), D. Reidel, Dordrecht, 233-264.

Taillie,C. and Patil,G.P. Rao-Rubin condition for a certain class of continuous damage models, In *Statistical Distributions in Scientific Work, Vol. 4,* C. Taillie et al (eds.) D. Reidel Publi. Co., 365-373.

Taguchi,T (1972a) On the two -dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to to the two dimensional case-I. *Annals of Institute of Statistical Mathematics,* 24, 355-382.

Taguchi,T (1972b) On the two -dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to to the two dimensional case-II. *Annals of Institute of Statistical Mathematics,* 24, 599-619.

Taguchi,T (1973) On the two -dimensional concentration surface and extensions of concentration coefficient and Pareto distribution to to the two dimensional case-III. *Annals of Institute of Statistical Mathematics,* 25, 215-237.

Takahashi,M. (1966) Theoretical basis for cell cycle analysis. I. Labeled mitosis wave method. *J. Theoret. Biol., 13, 202-211.*

Takahashi,M. (1968) Theoretical basis for cell cycle analysis. II. Further studies on labeled mitosis wave method. *J. Theoret. Biol.*, 18, 195-209.

Talwalker,S. (1970) A characterizationof the double Poisson distribution, *Sankhya, Series A,* 32, 265-270.

Talwalker,S. (1975) Models in medicine and toxicology, In *Statistical Distributions in Scientific Work, Vol 2,* G.P. Patil, S. Kotz, J.K. Ord (eds.) D. Reidel Publis. Co., Dordrecht, Holland, 263-274.

Talwalker,S. (1986) Functional equations in characterizations of discrete distributions by Rao-Rubin condition and its variants, *Communications in Statistics, Part A: Theory and Methods,* 15(3) 961-979.

Thompson,W.A. (1981) On the foundations of reliability. *Technometrics,* 23 (1), 1-13.

Tsokos,C.P. and Shimi,I.N. (1977) The Bayesian and non-parametric approach to reliability studies: A survey of recent work. In *The Theory and Applications of Reliability,* Vol II, C.P. Tsokos and I.N. Shimi (eds.), Academic Press, New York, 401-435.

Uppuluri,V.R. and Patil,S.A. (1976) Inferences about rare events. Oak Redge National Laboratory, Technical Report, 1-6.

Vardi,Y., Shepp,L.A. and Logan,B.F. (1981) Distribution functions invariant under residual-lifetime and length biased sampling, Z. Wahrscheinlichtkeit theorie verw, *Geberte,* 56, 427-441.

Wallach,R.V. and Sichel,H.S. (1963) The measurement of mechanical strength and abradability of coke by means of statistical parameters, *J. Inst. Fuel.,* 36, 421-435.

Wang,Y.H. and Srivastava,R.C. (1980) A characterization of the exponential and related distributions by linear regression. *Annals of Statistics,* 8, 217-220.

Warren,W.G. (1974) Statistics in forest-products research. In *Proc. 4th Conf. Advisory Group of Forest Statisticians* , IUFRO.

Warren,W.G. (1975) Statistical distributions in forestry and forest products research, In *Statistical Distributions in Scientific Work,* 2, G.P. Patil, S. Kotz and J.K. Ord (eds.), Reidel, Dordrecht and Boston, 369-384.

Watson,G.S. (1971) Estimating functionals of particle size distributions. *Biometrika,* 58, 483-490.

Xekalaki,E. (1977a) Bivariate and multivariate extensions of the generalized Waring distribution, Ph.D. thesis, University of Bradford, U.K.

Xekalaki,E. (1977b) On characterizing the bivariate Poisson, Binomial andnegative distributions. *Colloquia Mathematica Societatis Janos Bolyai 21. Analytic Function Methods in Probability Theory.* Debrecen, Hungary, 369-379.

Xekalaki,E. (1981) Chance mechanisms for the univariate generalised Waring distribution. *Statistical Distributions in Scientific Work, Vol. 4 (Models, Structures and Characterizations).* C. Taillie et al (eds), Holland: D. Reidel Publ. Co.

Xekalaki,E. (1983a) A property of the Yule distribution and its applications. *Communic. in Statistics Ser. A (Theory and Methods)*, 12(10), 1181-1189.

Xekalaki,E. (1983b) Hazard functions and life distributions in discrete time. *Communic. in Statistics Ser. A (Theory and Methods)*, 12(21), 2503-2509.

Xekalaki,E. (1984a) Models leading to the bivariate generalized Waring distribution. *Utilitas Mathematicae*, 25, 263-290.

Xekalaki,E. (1984b) Linear regression and the Yule distribution. *Journal of Econometrics*, 24, 397-403.

Xekalaki,E. (1986) The bivariate Yule distribution and some of its properties. *Statistics*, 17(2), 311-317.

Xekalaki,E. (1988a) Η επίδραση της κατανομής της ζήτησης σε μοντέλα ανανέωσης αποθέματος, *Πρακτικά Διήμερου Κλασσικής και Μπεϋζιανής Στατιστικής*, Ελληνικό Στατιστικό Ινστιτούτο, Αθήνα, Φεβρουάριος 1988.

Xekalaki,E. (1988b) Buying and Stocking under Uncertainty, *Επιστημονική Επετηρίδα*, Ανωτάτη Σχολή Οικονομικών και Εμπορικών Επιστημών, Αθήνα.

Xekalaki,E. and Panaretos,J. (1988) On the association of the Pareto and the Yule distributions. *Theoriya Veroyatnostei i ee Primeneniya. (Theory of Probability and its Applications)*, 33, 206-210.

Zelen,M. (1969) On the theory of screening for chronic disease, *Biometrika*, 56, 601-614.

Zelen,M. (1971) Problems in the early detection of disease and the finding of faults. *Bull. Int. Statist. Inst.*, 44, 649-659.

Zelen,M. (1974) Problems in cell kinetics and the early detection of disease, In *Reliability and Biometry*, F. Prochan and R.J. Serfling (eds.), SIAM, Philadelphia, 701-706.

A B S T R A C T

The primary objective of this thesis is to establish characterizations for distributions used in income analysis among other fields of application. The thesis consists of eight chapters which are briefly described.

The first chapter is of an introductory nature. Chapter 2 provides a classification of bivariate Pareto and Yule distributions that already exist in the literature or are introduced for the first time in this thesis by genesis scheme. Characteristic properties based on conditional expectations and/or conditional probabilities are shown in chapter 3 while the fourth chapter contains characterizations for the bivariate distributions introduced in chapter 2. In chapter 5 the distributions Pareto and Yule are studied in the context of reliability theory and several characterization results concerning both the univariate and bivariate cases are proved. Chapters 6 and 7 contain characteristic properties obtained mainly for Pareto and Yule distributions within the framework of systems of distributions. Emphasis is given to the system of weighted distributions. Results concerning both the univariate and bivariate cases are proved. Finally chapter 8 contains characteristic properties of continuous and discrete probability distributions in the context of what is defined as a distortion model.