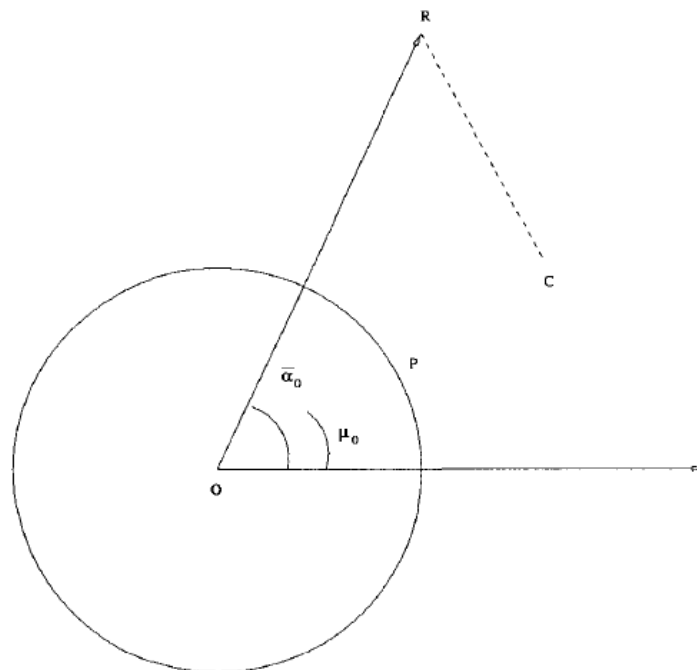




ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Εφαρμοσμένων Μαθηματικών & Φυσικών Επιστημών
Τομέας Μαθηματικών

Στατιστική Ανάλυση Κυκλικών Δεδομένων



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Λιαπτικός Δ. Ηλίας

Επιβλέπων Καθηγητής: Φουσκάκης Δημήτριος

Αθήνα, Οκτώβριος 2014

Πίνακας περιεχομένων

Περίληψη	4
Abstract.....	5
Εισαγωγή.....	6
Κεφάλαιο 1.....	8
1.1 Γραφική αναπαράσταση κυκλικών δεδομένων	8
1.2 Μέτρα θέσης και μεταβλητότητας κυκλικών δεδομένων	10
Κεφάλαιο 2.....	19
2.1 Εισαγωγή.....	19
2.2 Συνάρτηση πιθανότητας κυκλικών μεταβλητών	19
2.3 Χαρακτηριστική συνάρτηση.....	20
2.4 Μέθοδοι κατασκευής κυκλικών κατανομών	22
2.5 Ομοιόμορφη κατανομή	24
2.6 Καρδιοειδής κατανομή.....	24
2.7 Τριγωνική κατανομή	25
2.8 Κυκλική κανονική κατανομή (VonMises)	25
2.9 Παραλλαγές κυκλικής κανονικής κατανομής	28
2.10 Περιελιγμένες κυκλικές κατανομές.....	29
2.10.1 Περιελιγμένη κανονική κατανομή.....	29
2.10.2 Περιελιγμένη Cauchy κατανομή	29
2.10.3 Περιελιγμένες α-σταθερές κατανομές.....	30
2.11 Κυκλική βήτα κατανομή	30
Κεφάλαιο 3.....	32
3.1 Εισαγωγή.....	32
3.2 Εύρεση δειγματικών κατανομών	32
3.2.1 Δείγμα από κυκλική ομοιόμορφη κατανομή	32
3.2.2 Δείγμα από κυκλική κανονική κατανομή.....	34
3.3 Προσεγγίσεις σε Δείγματα μεγάλου μεγέθους	35
3.3.1 Προσεγγίσεις με σειρές.....	35
3.3.2 Προσεγγίσεις κεντρικού οριακού θεωρήματος	37
3.3.3 Προσεγγίσεις με την μέθοδο των ροπών	37
3.4 Πολλαπλά δείγματα	39
3.5 Προσεγγιστικές κατανομές για μεγάλο κ.....	41
Κεφάλαιο 4.....	42
4.1 Εισαγωγή.....	42
4.2 Κυκλική κανονική	42
4.2.1 Εκτιμήτριες μέγιστης πιθανοφάνειας	42
4.2.2 Εκτιμήτριες με την μέθοδο των ροπών.....	44
4.3 Μίξεις κυκλικής κατανομής	44
4.4 Εκτίμηση μέγιστης πιθανοφάνειας της περιελιγμένης Cauchy.....	45
4.5 Κυκλική βήτα.....	47
Κεφάλαιο 5.....	48
5.1 Εισαγωγή.....	48
5.2 Μονός Πληθυσμός.....	48
5.3 Δύο ή περισσότεροι πληθυσμοί.....	50
5.4 Έλεγχοι ομοιομορφίας.....	54
Κεφάλαιο 6.....	57
6.1 Εισαγωγή.....	57

6.2 Συντελεστής συσχέτισης κυκλικών μεταβλητών	57
6.2.1 Συντελεστής ρ_c σε μοντέλο με ομοιόμορφες περιθώριες.....	59
6.3 Συντελεστής συσχέτισης κυκλικών-γραμμικών μεταβλητών	60
6.4 Παλινδρόμηση κυκλικών-γραμμικών μεταβλητών	61
6.5 Παλινδρόμηση γραμμικών-κυκλικών μεταβλητών	63
6.6 Παλινδρόμηση κυκλικών-κυκλικών μεταβλητών	63
Κεφάλαιο 7.....	65
7.1 Εισαγωγή.....	65
7.2 «CircStats» βιβλιοθήκη.....	65
Κεφάλαιο 8.....	76
8.1 Εισαγωγή.....	76
8.2 Μορφή δεδομένων.....	76
8.2 Ανάλυση του δείγματος και εξαγωγή συμπερασμάτων	77
Παράρτημα Α.....	84
Α.1 Ορισμός συναρτήσεων Bessel.....	84
Α.2 Τροποποιημένη συνάρτηση Bessel πρώτου είδους	85
Α.3 Βασικά θεωρήματα συναρτήσεων Bessel.....	87
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	90

Διάγραμμα 1-1: Διάγραμμα διασποράς μη-ομαδοποιημένων παρατηρήσεων	9
Διάγραμμα 1-2: Ροδόγραμμα παρατηρήσεων με ομαδοποιημένες παρατηρήσεις.....	10
Διάγραμμα 1-3: Συσχέτιση ορθοκανονικών και πολικών συντεταγμένων	10
Διάγραμμα 1-4: Αριθμητικός μέσος και συνιστάμενο διάνυσμα δείγματος με αντιδιαμετρικές παρατηρήσεις στον άξονα των X	11
Διάγραμμα 1-5: Κυκλική απόσταση του διανύσματος v και τυχούσας παρατήρησης.....	16
Διάγραμμα 1-6: Προβολή συνιστάμενου διανύσματος στην πολική κατεύθυνση.....	18
Διάγραμμα 2-1(α): Κυκλική αναπαράσταση συνεχούς κυκλικής κατανομής	20
Διάγραμμα 2-1(β): Γραμμική αναπαράσταση συνεχούς κυκλικής κατανομής.....	20
Διάγραμμα 7-1:Γραφική απεικόνιση συνάρτησης πυκνότητας πιθανότητας της περιελιγμένης Cauchy κατανομής για διάνυσμα παρατηρήσεων	69
Διάγραμμα 7-2:Εμπειρική κατανομή δεδομένων σε σύγκριση με την πραγματική Von Mises Κατανομή.....	72
Διάγραμμα 8-1(α): Κατευθύνσεις ανέμου στον τριγωνομετρικό κύκλο.....	76
Διάγραμμα 8-1(β): Επεξηγηματικός πίνακας κατευθύνσεων ανέμων	77
Διάγραμμα 8-2:Κυκλικά διαγράμματα κατευθύνσεων του ανέμου	78
Διάγραμμα 8-3: Τριγωνομετρικός κύκλος με παρατηρήσεις στο «κλασικό» σύστημα συντεταγμένων	79
Διάγραμμα 8-4: Εμπειρική κατανομή δεδομένων σε σύγκριση με την πραγματική Von Mises κατανομή.....	81
Διάγραμμα Α.1: Συναρτήση Bessel για διαφορετικές τιμές του v	84
Διάγραμμα Α.2: Τροποποιημένη συνάρτηση Bessel πρώτου είδους για διαφορετικές τιμές του v	85

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Στατιστική Ανάλυση Κυκλικών Δεδομένων

Επιβλέπων Καθηγητής : Φουσκάκης Δημήτριος

Περίληψη

Με τον όρο κυκλικά δεδομένα αναφερόμαστε σε δεδομένα που περιγράφουν είτε διεύθυνση είτε κατεύθυνση. Η παρατήρηση τέτοιου είδους δεδομένων και η προσπάθεια στατιστικής ανάλυσης τους, αναπτύσσεται παράλληλα με τις εφαρμοσμένες επιστήμες τα τελευταία χρόνια και ως εκ τούτου η βιβλιογραφία αλλά και η λεπτομερής περιγραφή τους είναι αρκετά περιορισμένη. Λόγω της ιδιαίτερης μορφής τους δε, τα κυκλικά δεδομένα δεν επιδέχονται της ίδιας αντιμετώπισης με τις συνήθεις (γραμμικές) μεταβλητές που παίρνουν τιμές στο \mathbb{R} , παρότι η στατιστική τους ανάλυση στηρίζεται σε σχεδόν πανομοιότυπα στάδια, όπως της εκτίμησης παραμέτρων, της δημιουργίας διαστημάτων εμπιστοσύνης, των ελέγχων υποθέσεων αλλά και της δημιουργίας μοντέλων παλινδρόμησης μεταξύ κυκλικών μεταβλητών.

Στην παρούσα εργασία γίνεται μια προσπάθεια καταγραφής βασικών ιδιοτήτων και διαφορών μεταξύ των κυκλικών και των μεταβλητών που αναπαρίστανται στην ευθεία των πραγματικών αριθμών (τις οποίες θα καλούμε «γραμμικές» στην παρούσα διπλωματική). Παρουσιάζονται οι εγγενείς δυσκολίες της στατιστικής ανάλυσης των «κατευθυνόμενων» δεδομένων και οι μετατροπές που απαιτούνται για την ορθή παρουσίασή τους και την εξαγωγή δόκιμων συμπερασμάτων. Έπειτα από την ανάλυση των βασικών κυκλικών κατανομών και των σταδίων της επεξεργασίας των δεδομένων, παρατίθεται και η εφαρμογή τους με την χρήση του στατιστικού πακέτου R , καθώς και η επεξήγηση των αντίστοιχων εντολών που χρησιμοποιήθηκαν. Τέλος, πραγματοποιείται μια συνολική ανάλυση ενός δείγματος πραγματικών κυκλικών δεδομένων και παρουσιάζονται τα αποτελέσματά της.

NATIONAL TECHNICAL UNIVERSITY OF ATHENS
FACULTY OF APPLIED MATHEMATICS
AND PHYSICAL SCIENCE
DEPARTMENT OF MATHEMATICS

Final Year Project

Statistical Analysis of Circular Data

Supervising professor: Fouskakis Dimitrios

Abstract

The term «circular data» is used in the area of Statistics to describe either directional or axial data. Observation of such kind of data and the need for appropriate statistical analysis is increasing simultaneously with the expansion of applied science. Due to that, the relevant bibliography and detailed analysis is very rare and usually referring only to specific cases. As a result of the special form of circular data we cannot perform the same statistical analysis we apply to linear data, despite the fact that the steps are practically identical, such us estimate of parameters, creation of confidence intervals, test of cases, and fitting regression models among circular variables.

At this final year project we make an effort to describe the main properties and differences of circular data comparing to linear case. We present the inherent difficulties of the appropriate analysis and the alternations that are required in order to derive the right conclusions. After the presentation of the main circular distributions and the steps of the statistical analysis we apply the theoretical approach to some circular data using the statistical package *R*, analyzing the code that we used. In the end we apply a full statistical analysis on some real circular data and we present the results.

Εισαγωγή

Σε διάφορα επιστημονικά πεδία, υπάρχει ανάγκη συλλογής μετρήσεων, που το υπό εξέταση χαρακτηριστικό τους είναι η κατεύθυνση ή η διεύθυνση. Ένα σύνολο τέτοιου είδους καλείται «κατευθυνόμενα δεδομένα». Προκύπτουν συνηθέστερα δε, σε επιστημονικά πεδία όπως αυτά της ζωολογίας, της γεωγραφίας αλλά και στις οικονομικές και ιατρικές επιστήμες. Μία από τις πρώτες επιστημονικές μελέτες στις οποίες έγινε αναγκαία η χρήση τέτοιων δεδομένων, προέκυψε στον τομέα της βιολογίας. Συγκεκριμένα υπήρχε η ανάγκη παρατήρησης και καταγραφής της κατεύθυνσης πτήσης στον ορίζοντα, μεταναστευτικών πτηνών που είχαν μόλις απελευθερωθεί.

Τα δεδομένα αυτού του είδους μπορούν να αναπαρασταθούν ως γωνίες μετρούμενες σε σχέση με κατάλληλα επιλεγμένη «μηδενική» διεύθυνση και φορά θετικής περιστροφής. Το σημαντικό στην συγκεκριμένη ανάλυση είναι ότι δεν μας απασχολεί το μέγεθος της παρατήρησης αλλά η κατεύθυνσή του. Έτσι οι μετρήσεις μπορούν να μετατραπούν σε σημεία πάνω στην περιφέρεια κύκλου οποιασδήποτε διαμέτρου (βολικότερα και συνηθέστερα μοναδιαίας ακτίνας) με επίκεντρο το κέντρο του κύκλου, ή ως μοναδιαία διανύσματα με αρχή το κέντρο του κύκλου και πέρασ το σημείο της μέτρησης.

Λόγω αυτής της αναπαράστασης, τα δεδομένα αυτής της μορφής (δύο διαστάσεων) αναφέρονται συχνά στην βιβλιογραφία ως *κυκλικά δεδομένα*. Αντίστοιχα τα δεδομένα τριών διαστάσεων αναπαρίστανται με δύο γωνίες, ενώ οι μετρήσεις καταγράφονται στην επιφάνεια μοναδιαίας σφαίρας και για αυτό αναφέρονται ως σφαιρικά δεδομένα.

Η ανάγκη στατιστικής ανάλυσης δεδομένων τέτοιου είδους καθίσταται αναγκαία όσο περισσότερο αναπτύσσονται οι εφαρμοσμένες επιστήμες. Υπάρχουν πολλά παραδείγματα που καταδεικνύουν την σημασία των κυκλικών μεταβλητών στην σύγχρονη εποχή αλλά αυτά που θα παραθέσουμε αμέσως δείχνουν τον εξαιρετικά χρήσιμο χαρακτήρα τους. Καταρχάς ανάλυση κυκλικών μεταβλητών χρησιμοποιείται από το σύνολο του οικονομικού κόσμου παγκοσμίως για την εύρεση του κατάλληλου ισοζυγίου προσωπικού εργασίας, παρατηρώντας την ώρα μέγιστης ανάγκης εξυπηρέτησης πελατών. Επίσης σημαντικές εφαρμογές προκύπτουν και στον κλάδο της ιατρικής επιστήμης, εντοπίζοντας τον κατάλληλο χρόνο χορήγησης φαρμακευτικών σκευασμάτων με σκοπό την βέλτιστη απορρόφηση και επίδραση στον ασθενή. Η συγκεκριμένη ανάλυση βρίσκει εφαρμογή σε κάθε περιοδικό φαινόμενο με γνωστή περίοδο όπως μια ημέρα, ή ένας μήνας ή ένας χρόνος, καθώς οι παρατηρήσεις της τυχαίας μεταβλητής που περιγράφει το φαινόμενο μπορούν να αναπαρασταθούν σε κύκλο του οποίου η περιφέρεια αντιστοιχεί στην περίοδο. Ακριβώς σε αυτό το σημείο βασίζεται και η διαφορά των «γραμμικών μεταβλητών» με τις αντίστοιχες κυκλικές, καθώς στις τελευταίες δεν μας απασχολεί ο χρόνος του συμβάντος αλλά η στιγμή στην οποία παρατηρήθηκε μέσα στην περίοδο.

Οι εν λόγω μετρήσεις εμφανίζουν μία σειρά από *ιδιαιτερότητες*:

- Η αναπαράσταση τους δεν είναι μοναδική καθώς εξαρτάται από την επιλογή της «μηδενικής» κατεύθυνσης και της θετικής φοράς περιστροφής.
- Δεν υπάρχει φυσική διάταξη των παρατηρήσεων καθώς το εάν μια μέτρηση είναι μεγαλύτερη από μία άλλη εξαρτάται από τις προαναφερθείσες επιλογές.
- Η αρχή με το τέλος συμπίπτουν με τον τρόπο αναπαράστασης μας, άρα και οι μετρήσεις μας εμφανίζουν μία περιοδικότητα.
- Η ανάγκη μοντελοποίησης ασύμμετρων πακέτων δεδομένων που εμφανίζονται συνηθέστερα εμπεριέχει εγγενείς δυσκολίες.

Δομή Εργασίας

- Στο πρώτο κεφάλαιο παρουσιάζονται οι τρόποι γραφικής αναπαράστασης και τα κύρια μέτρα θέσης και μεταβλητότητας των κυκλικών μεταβλητών.
- Στο δεύτερο κεφάλαιο αναλύονται οι βασικές θεωρητικές κυκλικές κατανομές και κάποιες βασικές ιδιότητες τους. Ιδιαίτερη σημασία δίνεται στην κυκλική κανονική κατανομή VonMises που είναι η πιο διαδεδομένη και εμφανίζει σημαντικά πλεονεκτήματα.
- Στο τρίτο κεφάλαιο παρουσιάζονται οι εκτιμήσεις των κατανομών βασικών στατιστικών ποσοτήτων για την κυκλική ομοιόμορφη και κυκλική κανονική κατανομή. Επίσης δίνονται κάποιες χρήσιμες προσεγγίσεις ανάλογα με την μορφή και το μέγεθος του δείγματος.
- Στο τέταρτο κεφάλαιο αναλύεται η σημειακή εκτίμηση των παραμέτρων της κυκλικής κανονικής κατανομής αλλά και κάποιον μεικτών κατανομών.
- Στο πέμπτο κεφάλαιο προχωρούμε στους ελέγχους υποθέσεων για την μέση κατεύθυνση και την παράμετρο συγκέντρωσης της Von Mises κατανομής καθώς και σε κάποιους ελέγχους ομοιομορφίας.
- Στο έκτο κεφάλαιο αναλύονται οι συντελεστές συσχέτισης μεταξύ κυκλικών μεταβλητών αλλά και μεταξύ κυκλικών και γραμμικών μεταβλητών. Επίσης, παρουσιάζονται τα μοντέλα παλινδρόμησης μεταξύ κυκλικών και γραμμικών μεταβλητών.
- Στο έβδομο κεφάλαιο παρουσιάζονται και αναλύονται οι βασικές εντολές της R που χρησιμοποιούνται για την πρακτική εφαρμογή της ανάλυσης που προηγήθηκε στην παρούσα εργασία. Τέλος, παρουσιάζονται κάποια παραδείγματα της εφαρμογής τους σε κυκλικά δεδομένα.
- Στο όγδοο κεφάλαιο γίνεται μια βασική στατιστική ανάλυση μέσω της R , ενός δείγματος μετεωρολογικών κυκλικών δεδομένων.
- Στο παράρτημα αναλύεται ένα μαθηματικό βοήθημα για τις συναρτήσεις Bessel που χρησιμοποιούνται στην εργασία. Παρουσιάζονται κάποιες χρήσιμες ιδιότητες και εφαρμογές τους στην ανάλυση των κυκλικών μεταβλητών.

Κεφάλαιο 1

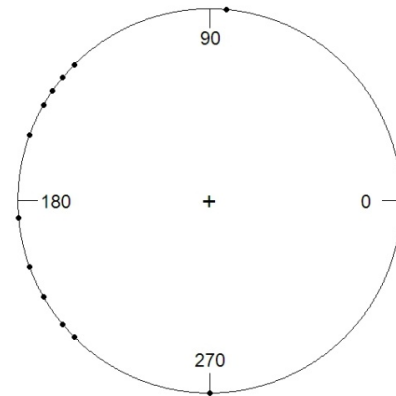
Τρόποι γραφικής αναπαράστασης και μέτρα θέσης και μεταβλητότητας κυκλικών δεδομένων

1.1 Γραφική αναπαράσταση κυκλικών δεδομένων

Όπως και στην κλασική στατιστική ανάλυση γραμμικών μεταβλητών, σημαντικότερο ρόλο παίζει πρώτα από όλα η μείωση του όγκου της πληροφορίας και μια πρώτη εκτίμηση των βασικών μέτρων θέσης και μεταβλητότητας μέσω κατάλληλης γραφικής αναπαράστασης των δεδομένων μας. Όπως έχουμε ήδη επισημάνει στην εισαγωγή, τα κυκλικά δεδομένα περιγράφονται από την γωνία που σχηματίζουν με την μηδενική κατεύθυνση που έχουμε επιλέξει και ανάλογα με την μορφή που έχει το πακέτο των δεδομένων μας και την ανάλυση που θέλουμε να κάνουμε, ενδείκνυται και διαφορετική αναπαράστασή τους. Οι βασικότερες μέθοδοι αναπαράστασης για κυκλικά δεδομένα είναι οι ακόλουθες:

- **Κυκλικό διάγραμμα διασποράς**

Η μέθοδος αυτή ενδείκνυται για μη-ομαδοποιημένες παρατηρήσεις και αναπαριστά τις παρατηρήσεις μας με σημεία πάνω στην περιφέρεια μοναδιαίου κύκλου στην προέκταση από το κέντρο του κύκλου της κατάλληλης νοητής κατεύθυνσης. Είναι ιδιαίτερα χρήσιμη για δεδομένα που δεν εμφανίζουν κάποια ομαδοποίηση γιατί τότε θα συνέπιπτε η αναπαράσταση στο ίδιο σημείο της περιφέρειας του κύκλου πολλών παρατηρήσεων. Από το κυκλικό διάγραμμα διασποράς μπορούμε εύκολα να συμπεράνουμε το μέγεθος καθώς και την διασπορά των παρατηρήσεων στο δείγμα μας. Για παράδειγμα στο Διάγραμμα 1-1 έχουμε την απεικόνιση δώδεκα κατευθύνσεων αποχώρησης πτηνών και μπορούμε να παρατηρήσουμε τόσο την επιλεγμένη μηδενική κατεύθυνση (σημειώνεται ως αυτή που αντιστοιχεί στις 0°), όσο και την υψηλότερη συγκέντρωση των παρατηρήσεων στο δεύτερο και τρίτο τεταρτημόριο.



Διάγραμμα 1-1: Διάγραμμα διασποράς μη-ομαδοποιημένων παρατηρήσεων

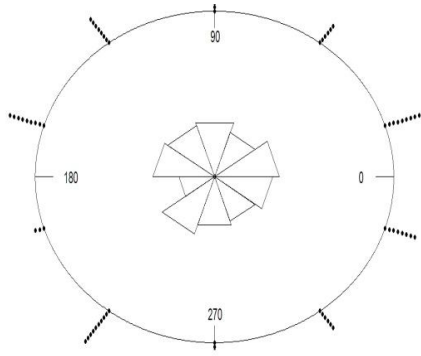
- **Κυκλικό ιστογράμμα**

Αυτή η μέθοδος απεικόνισης είναι πιο χρήσιμη σε ομαδοποιημένα δεδομένα. Κάθε τμήμα του ιστογράμματος έχει ως επίκεντρο το μέσο του κάθε διαστήματος των παρατηρήσεων και είτε το μήκος του είτε το εμβαδόν του αντιστοιχούν στην συχνότητα των παρατηρήσεων στο εν λόγω εύρος¹. Έχει μεγάλη εφαρμογή σε συλλογές δεδομένων που εμφανίζουν αποτελέσματα με ομαδοποιημένα χαρακτηριστικά, όπως για παράδειγμα ο χρόνος εκτέλεσης μιας ιατρικής επέμβασης. Συνήθως στα ομαδοποιημένα δεδομένα δεν μας απασχολούν οι ακριβείς τιμές των παρατηρήσεων αλλά το κατά πόσο διαφέρουν μεταξύ τους και εάν τυχόν είναι συγκεντρωμένες σε κάποιο διάστημα.

- **Ροδόγραμμα**

Αυτή η μέθοδος είναι παρόμοια με του κυκλικού ιστογράμματος καθώς εδώ τα τμήματα του ιστογράμματος αντικαθίστανται από κυκλικούς τομείς, στους οποίους διαιρούμε την περιφέρεια του κύκλου. Και σε αυτή την περίπτωση συνηθέστερα το εμβαδό του κάθε τομέα αντιστοιχεί στην συχνότητα των παρατηρήσεων, κάτι που επιτυγχάνεται θέτοντας την ακτίνα του τομέα ανάλογη της τετραγωνικής ρίζας της σχετικής συχνότητας και όχι της συχνότητας του δείγματος. Στο Διάγραμμα 1-2 βλέπουμε την απεικόνιση σε ροδόγραμμα 75 παρατηρήσεων χωρισμένα σε 10 τομείς. Απεικονίζεται το πλήθος των παρατηρήσεων στοιβαγμένα στην περιφέρεια του κύκλου για κάθε τομέα γεγονός που μας επιβεβαιώνει ότι όπου υπάρχει μεγαλύτερη συγκέντρωση παρατηρήσεων το τμήμα έχει εμφανώς μεγαλύτερη ακτίνα.

¹ Γραμμικό ιστογράμμα: Ως συνέχεια του παραπάνω μπορεί να επιλεγεί ένα σημείο του κυκλικού ιστογράμματος γύρω από το οποίο θα αναπτυχθεί το γραμμικό ιστογράμμα. Για την λήψη ορθών συμπερασμάτων έχει μεγάλη σημασία η επιλογή του κατάλληλου σημείου κοπής. Για παράδειγμα, εάν τα δεδομένα μας παρουσιάζουν ένα κομβικό σημείο, (σημείο με μεγαλύτερη συχνότητα στο δείγμα) το σημείο κοπής οφείλει να είναι κοντά σε αυτό, ειδάλως οδηγούμαστε στο ανακριβές συμπέρασμα ύπαρξης δύο κόμβων.

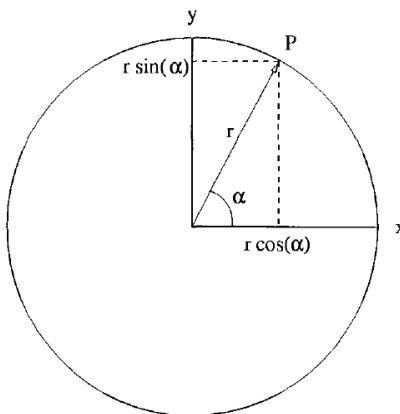


Διάγραμμα 1-2: Ροδόγραμμα παρατηρήσεων με ομαδοποιημένες παρατηρήσεις

1.2 Μέτρα θέσης και μεταβλητότητας κυκλικών δεδομένων

Όπως αναφέρθηκε στην εισαγωγή τα κατευθυνόμενα δεδομένα, μπορούν να αναπαρασταθούν μοναδικά μέσω δύο συντεταγμένων. Στο ορθοκανονικό σύστημα ως $\vec{u} = (x, y)$ και σε πολικές συντεταγμένες ως $\vec{u} = (r, \alpha)$ όπου r είναι η ακτίνα του κύκλου και α η κατεύθυνση της παρατήρησης. Οι δύο αυτές μορφές εναλλάσσονται μεταξύ τους, όπως φαίνεται και στο Διάγραμμα 1-3, μέσω των παρακάτω σχέσεων:

$$\begin{cases} x = r \cos \alpha \\ y = r \sin \alpha \end{cases} \quad (1.1)$$



Διάγραμμα 1-3: Συσχέτιση ορθοκανονικών και πολικών συντεταγμένων

Εδώ χρειάζεται ιδιαίτερη προσοχή, καθώς οι δύο συντεταγμένες φαίνονται ανεξάρτητες ενώ δεν είναι, καθώς δεσμεύονται από την σχέση:

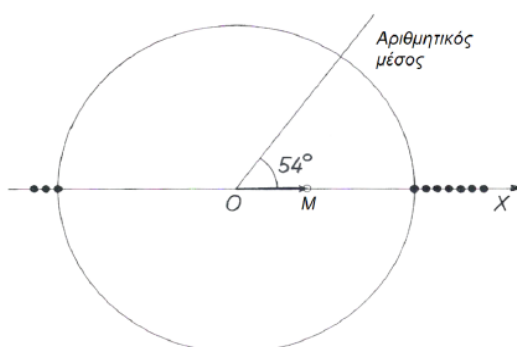
$$x^2 + y^2 = r^2. \quad (1.2)$$

Τέλος, στην περίπτωση που έχουμε σφαιρικά δεδομένα όπως για παράδειγμα δεδομένα γεωγραφικού μήκους και πλάτους, αυτά αναπαρίστανται μοναδικά με τρεις μεταβλητές. Στο καρτεσιανό σύστημα συντεταγμένων με $\vec{u} = (x, y, z)$ και σε σφαιρικές συντεταγμένες με $\vec{u} = (r, \alpha_1, \alpha_2)$ όπου r η ακτίνα της σφαίρας και (α_1, α_2) το ζεύγος των κατευθύνσεων που χαρακτηρίζουν την παρατήρηση μας. Συγκεκριμένα α_1 είναι η γωνία που σχηματίζει το διάνυσμα θέσης της παρατήρησης μας με τον άξονα των x ενώ α_2 η γωνία που σχηματίζει η προβολή του διανύσματος θέσης στο επίπεδο $y - z$ με τον άξονα των y . Επειδή όπως αναφέραμε και στην εισαγωγή της παρούσας διπλωματικής εργασίας στα κυκλικά δεδομένα δεν μας απασχολεί το μέγεθος της παρατήρησης αλλά η κατεύθυνση, μπορούμε να θεωρήσουμε τα διανύσματα θέσης που τα αναπαριστούν ως μοναδιαία ($r = 1$) και με αυτή τη σύμβαση θα συνεχίσουμε σε όλο το κείμενο της εργασίας αυτής. Κατά συνέπεια στην περίπτωση των σφαιρικών δεδομένων μπορούμε να εναλλάσσουμε την αναπαράστασή τους βάσει του ακόλουθου τύπου:

$$(x, y, z) = (\cos \alpha_1, \sin \alpha_1 \cos \alpha_2, \sin \alpha_1 \sin \alpha_2). \quad (1.3)$$

➤ Μέτρα θέσης

Σύμφωνα με την μέχρι τώρα ανάλυση, είναι προφανές ότι για να βρούμε την μέση κατεύθυνση ενός δείγματος, δεν μπορούμε να χρησιμοποιήσουμε τον συνηθισμένο αριθμητικό μέσο, καθώς μπορεί να δώσει κατεύθυνση που να μην υπάρχει καν στο δείγμα μας. Για παράδειγμα, εάν έχουμε δείγμα μετρήσεων στις διευθύνσεις 0° και 180° , ο αριθμητικός τους μέσος θα βρίσκεται σίγουρα σε σημείο του ημικυκλίου με κατεύθυνση διαφορετική από αυτή των μετρήσεων, πράγμα το οποίο φυσικά δεν είναι αποδεκτό. Όπως φαίνεται και στο Διάγραμμα 1-4 ο αριθμητικός μέσος των δέκα παρατηρήσεων προκύπτει σε κατεύθυνση που δεν αντιστοιχεί σε καμία από αυτές των επιμέρους μετρήσεων.



Διάγραμμα 1-4: Αριθμητικός μέσος και συνιστάμενο διάνυσμα δείγματος με αντιδιαμετρικές παρατηρήσεις στον άξονα των X

Ένα άλλο παράδειγμα που επιβεβαιώνει το γεγονός ότι η κλασική στατιστική ανάλυση στην περίπτωση των κυκλικών και σφαιρικών δεδομένων οδηγεί σε εντελώς ανασφαλής συμπεράσματα είναι αυτό της συλλογής δεδομένων γεωγραφικής κατεύθυνσης. Έστω ότι έχουμε τέσσερις παρατηρήσεις μετρούμενες βάσει του κλασικού ορισμού μηδενικής κατεύθυνσης ως ανατολή και θετικής κατεύθυνσης φοράς, της αντίθετης των δεικτών του ωρολογίου. Πιο συγκεκριμένα έστω ότι οι παρατηρήσεις μας παίρνουν τιμές στο διάστημα $[0^\circ, 360^\circ]$ και προκύπτουν $[50^\circ, 160^\circ, 200^\circ, 310^\circ]$. Η μέση τιμή τους λοιπόν προκύπτει 180° και ταυτίζεται με την κατεύθυνση που έχουμε ορίσει ως Δύση. Η παραμικρή αλλαγή όμως σε αυτές τις υποκειμενικές παραδοχές που κάνουμε μπορεί να αλλάξει εντελώς τα συμπεράσματά μας. Συγκεκριμένα έστω ότι κρατάμε σταθερές τις αξιωματικές παραδοχές που έχουμε κάνει και αλλάζουμε απλώς το πεδίο τιμών των παρατηρήσεων μας στο $[-180^\circ, 180^\circ]$. Αυτόματα οι παρατηρήσεις μας γίνονται $[-50^\circ, -160^\circ, 150^\circ, 60^\circ]$ και η μέση κατεύθυνση είναι αυτή των 0° δηλαδή ταυτίζεται με την Ανατολή εντελώς αντιδιαμετρικά με την προηγούμενη ανάλυση.

Τα παραπάνω παραδείγματα μας δείχνουν μερικές από τις εγγενείς δυσκολίες της προσέγγισης των κυκλικών δεδομένων και καταδεικνύουν την ανάγκη εύρεσης ενός διαφορετικού μέτρου της μέσης κατεύθυνσης που έχουν τα δεδομένα μας που δεν θα εξαρτάται από την εκάστοτε επιλογή της μηδενικής κατεύθυνσης και της θετικής φοράς περιστροφής. Μια τέτοια προσέγγιση αναλύεται ακολούθως.

Μέση κατεύθυνση και συνιστάμενο διάνυσμα

Μέσω της αναπαράστασης των δεδομένων ως διανύσματα θέσης, προκύπτει ότι μια σωστή προσέγγιση της μέσης κατεύθυνσης είναι η κατεύθυνση του συνισταμένου διανύσματος όλων των παρατηρήσεων.

- *Ορισμός συνισταμένου διανύσματος*

Έστω $(\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n)$ σύνολο κυκλικών παρατηρήσεων όπου $\alpha_i, \forall i = 1, 2, \dots, n$ είναι οι αντίστοιχες κατευθύνσεις τους. Τότε, το συνιστάμενο διάνυσμα \vec{R} προκύπτει μέσω του ακόλουθου τύπου:

$$\vec{R} = \left(\sum_{i=1}^n \cos \alpha_i, \sum_{i=1}^n \sin \alpha_i \right) \equiv (C, S). \quad (1.4)$$

Το μήκος του δε υπολογίζεται ως εξής :

$$\|\vec{R}\| = R = \sqrt{C^2 + S^2}. \quad (1.5)$$

Βάσει των παραπάνω ορίζουμε ως μέση κατεύθυνση του συνόλου των παρατηρήσεών μας την κατεύθυνση του R . Εάν συμβολίσουμε αυτή την κατεύθυνση ως $\bar{\alpha}_0$ τότε ορίζεται ως εξής:

$$\bar{\alpha}_0 = \arg \left\{ \sum_{j=1}^n \cos \alpha_j + i \sum_{j=1}^n \sin \alpha_j \right\} \text{ ή διαφορετικά } \begin{cases} \cos \bar{\alpha}_0 = \frac{C}{R} \\ \sin \bar{\alpha}_0 = \frac{S}{R} \end{cases}, \quad (1.6)$$

όπου \arg η συνάρτηση που μας επιστρέφει την γωνία που σχηματίζει ένας μιγαδικός αριθμός με τον άξονα των πραγματικών αριθμών.

Στο σημείο αυτό παρουσιάζεται μία ιδιαιτερότητα στον ορισμό της μέσης κατεύθυνσης. Καθώς οι συναρτήσεις του ημιτόνου και του συνημιτόνου που εμπλέκονται στον ορισμό του $\bar{\alpha}_0$ είναι περιοδικές πρέπει να δίνεται βάσει στα πρόσημα των (C, S) δηλαδή των αθροισμάτων των δύο τριγωνομετρικών συναρτήσεων για τις παρατηρήσεις μας. Για αυτό τον λόγο προχωράμε σε μία πιο αναλυτική μορφή του παραπάνω ορισμού όπως περιγράφεται παρακάτω:

$$\bar{\alpha}_0 = \arctan^* \left(\frac{S}{C} \right), \quad (1.7)$$

$$\text{όπου } \arctan^* \left(\frac{S}{C} \right) = \begin{cases} \arctan \left(\frac{S}{C} \right), & \alpha \nu C > 0 \text{ και } S \geq 0 \\ \frac{\pi}{2}, & \alpha \nu C = 0 \text{ και } S > 0 \\ \arctan \left(\frac{S}{C} \right) + \pi, & \alpha \nu C < 0 \\ \arctan \left(\frac{S}{C} \right) + 2\pi, & \alpha \nu C \geq 0 \text{ και } S < 0 \\ \text{αόριστο}, & \alpha \nu C = S = 0 \end{cases}. \quad (1.8)$$

Ο ορισμός της μέσης κατεύθυνσης με αυτό τον τρόπο μας επιτρέπει να δείξουμε ότι ξεφεύγουμε από την υποκειμενική επιλογή της μηδενικής διεύθυνσης και της θετικής φοράς περιστροφής. Αυτό καθίσταται ξεκάθαρο μέσα από τις ακόλουθες ιδιότητες του συνισταμένου διανύσματος.

■ *Ιδιότητες μέσου διανύσματος*

1. Έστω, οι παρατηρήσεις με κατευθύνσεις $(\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n)$ και μέση κατεύθυνση $\bar{\alpha}_0$. Θα αποδείξουμε ότι οι παρατηρήσεις $(\alpha_1 + c, \alpha_2 + c, \alpha_3 + c, \dots, \alpha_n + c)$ έχουν μέση κατεύθυνση $\bar{\alpha}_0 + c$ όπου c μια τυχαία σταθερά κατεύθυνσης.
- 2.

Έστω ότι ορίζουμε το συνιστάμενο διάνυσμα του δεύτερου συνόλου παρατηρήσεων ως:

$$\vec{R} = \left(\sum_{i=1}^n \cos(\alpha_i + c), \sum_{i=1}^n \sin(\alpha_i + c) \right) \equiv (C', S').$$

Τότε ισχύει:

$$\begin{aligned} C' &= \sum_{i=1}^n \cos(\alpha_i + c) = \sum_{i=1}^n (\cos \alpha_i \cos c - \sin \alpha_i \sin c) = \\ &= C \cos c - S \sin c = R \cos \bar{\alpha}_0 \cos c - R \sin \bar{\alpha}_0 \sin c = R \cos(\bar{\alpha}_0 + c). \end{aligned}$$

Ομοίως, $S' = R \sin(\bar{\alpha}_0 + c)$. Τότε $R' = \|\vec{R}'\| = \sqrt{C'^2 + S'^2} = \sqrt{C^2 + S^2} = R$.

Συνεπώς, η τιμή της μέσης κατεύθυνσης αλλάζει κατά ποσότητα όμοια με την περιστροφή των δεδομένων και το μήκος του μέσου παραμένει αναλλοίωτο ανεξάρτητα της σταθεράς c .

3. Με παρεμφερή τρόπο θα αποδείξουμε ότι η μέση κατεύθυνση όπως την ορίσαμε στην σχέση (1.7) παραμένει αναλλοίωτη ανεξάρτητα από την επιλογή της θετικής φοράς περιστροφής.

Έστω λοιπόν σύνολο κυκλικών δεδομένων με κατευθύνσεις $(\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n)$ με μέση κατεύθυνση $\bar{\alpha}_0$. Εάν αλλάξουμε την θετική φορά περιστροφής από ωρολογιακή σε αντι-ωρολογιακή τα δεδομένα μας γίνονται $2\pi - \alpha_i, \forall i = 1, 2, \dots, n$.

Θα αποδείξουμε λοιπόν ότι η μέση κατεύθυνση του εν λόγω συνόλου γίνεται σε αντιστοιχία $2\pi - \bar{\alpha}_0$.

Ορίζουμε το συνιστάμενο διάνυσμα της δεύτερης μορφής των δεδομένων μας

$$\text{ως: } \vec{R}'' = \left(\sum_{i=1}^n \cos(2\pi - \alpha_i), \sum_{i=1}^n \sin(2\pi - \alpha_i) \right) \equiv (C'', S'').$$

Τότε ισχύει:

$$\begin{aligned} C'' &= \sum_{i=1}^n \cos(2\pi - \alpha_i) = \sum_{i=1}^n (\cos 2\pi \cos \alpha_i + \sin 2\pi \sin \alpha_i) = \\ &= \sum_{i=1}^n \cos \alpha_i = C = R \cos \bar{\alpha}_0 = R \cos(-\bar{\alpha}_0) = R \cos(2\pi - \bar{\alpha}_0). \end{aligned}$$

Ομοίως έχουμε:

$$\begin{aligned} S'' &= \sum_{i=1}^n \sin(2\pi - \alpha_i) = \sum_{i=1}^n (\sin 2\pi \cos \alpha_i - \cos 2\pi \sin \alpha_i) = \\ &= -\sum_{i=1}^n \sin \alpha_i = -S = -R \sin \bar{\alpha}_0 = R \sin(-\bar{\alpha}_0) = R \sin(2\pi - \bar{\alpha}_0). \end{aligned}$$

Συνεπώς:

$$R'' = \sqrt{C''^2 + S''^2} = \sqrt{C^2 + S^2} = R \Rightarrow \begin{cases} \frac{C''}{R''} = \frac{C}{R} = \cos(2\pi - \bar{\alpha}_0) \\ \frac{S''}{R''} = \frac{S}{R} = \sin(2\pi - \bar{\alpha}_0) \end{cases}.$$

Οι δυο παραπάνω αποδείξεις των ιδιοτήτων του συνιστάμενου διανύσματος επιβεβαιώνουν την υπεροχή του ως «δείκτης» της μέσης κατεύθυνσης στην περίπτωση κυκλικών δεδομένων καθώς ορίζεται χωρίς να εξαρτάται από υποκειμενικά κριτήρια επιλογής φοράς και μηδενικής κατεύθυνσης.

4. Οι δύο παρακάτω σχέσεις προκύπτουν εάν αντικαταστήσουμε την τυχαία σταθερά κατεύθυνσης c της πρώτης ιδιότητας με την $-\bar{\alpha}_0$. Μας επιβεβαιώνουν μάλιστα ότι η μέση κατεύθυνση έχει κατασκευαστεί με ορθό τρόπο καθώς οι παρατηρήσεις δείχνουν να απέχουν ανάλογα ίσα από αυτή.

$$\begin{aligned} \sum_{i=1}^n \sin(\alpha_i - \bar{\alpha}_0) &= \sum_{i=1}^n (\sin \alpha_i \cos \bar{\alpha}_0 - \cos \alpha_i \sin \bar{\alpha}_0) = S \cos \bar{\alpha}_0 - C \sin \bar{\alpha}_0 = \\ &= R \cos \bar{\alpha}_0 \sin \bar{\alpha}_0 - R \cos \bar{\alpha}_0 \sin \bar{\alpha}_0 = 0. \end{aligned}$$

5.
$$\begin{aligned} \sum_{i=1}^n \cos(\alpha_i - \bar{\alpha}_0) &= \sum_{i=1}^n (\cos \alpha_i \cos \bar{\alpha}_0 + \sin \alpha_i \sin \bar{\alpha}_0) = C \cos \bar{\alpha}_0 + S \sin \bar{\alpha}_0 = \\ &= R \cos^2 \bar{\alpha}_0 + R \sin^2 \bar{\alpha}_0 = R. \end{aligned}$$

➤ Μέτρα μεταβλητότητας

Μία άλλη χρήσιμη ιδιότητα του συνιστάμενου διανύσματος είναι ότι το μέτρο του μπορεί να χρησιμοποιηθεί ως ένδειξη της διασποράς του δείγματος. Σε πακέτο κυκλικών δεδομένων με μία κορυφή, εάν οι παρατηρήσεις είναι αρκετά συγκεντρωμένες, τότε το R τείνει να γίνει ίσο με το n , ενώ αντίθετα, εάν οι παρατηρήσεις είναι ισόμορφα κατανεμημένες, τότε τείνει στο μηδέν.

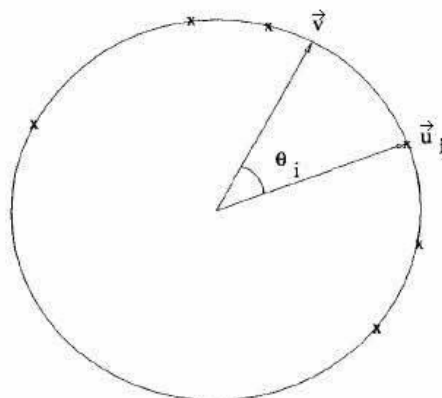
Θα αποδείξουμε ότι η ποσότητα $n - R$ είναι το αντίστοιχο της συνηθισμένης διασποράς σε δείγμα αποτελούμενο από κυκλικά δεδομένα.

Απόδειξη: Προτού προχωρήσουμε στη μαθηματική απόδειξη, πρέπει να ορίσουμε με ορθό τρόπο την έννοια της *κυκλικής απόστασης*. Ως κυκλική απόσταση, ονομάζουμε το μικρότερο των δύο τόξων μεταξύ δύο σημείων στην περιφέρεια του κύκλου. Ορίζεται δε μαθηματικά ως: $d(\vec{w}, \vec{z}) = 1 - \cos(\delta - \varepsilon)$, όπου \vec{w}, \vec{z} μοναδιαία διανύσματα δύο τυχαίων παρατηρήσεων και δ, ε οι αντίστοιχες γωνίες τους.

Έστω $\{\vec{u}_i, i = 1, 2, \dots, n\}$ το σύνολο μοναδιαίων διανυσμάτων που αναπαριστούν τις παρατηρήσεις μας, $\vec{v} = (p, q)$ τυχαίο μοναδιαίο διάνυσμα, και θ_i η γωνία που σχηματίζει η κάθε παρατήρηση με αυτό. Όπως παρατηρούμε και στο Διάγραμμα 1-6 η γωνία θ_i λαμβάνει τιμές στο διάστημα $[0, \pi]$ και βάσει του ορισμού της κυκλικής απόστασης ισχύει: $d(\vec{v}, \vec{u}_i) = 1 - \cos\theta_i$.

Συνεπώς εάν ορίσουμε το άθροισμα των επιμέρους κυκλικών αποστάσεων ως την διασπορά του δείγματος μας σε σχέση με το τυχαίο διάνυσμα \vec{v} έχουμε:

$$D_{\vec{v}}(\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n) = \sum_{i=1}^n d(\vec{v}, \vec{u}_i) = n - \sum_{i=1}^n (\vec{v} \cdot \vec{u}_i) = n - \sum_{i=1}^n \cos\theta_i. \quad (1.9)$$



Διάγραμμα 1-5: Κυκλική απόσταση του διανύσματος \vec{v} και τυχούσας παρατήρησης

Θα αποδείξουμε ότι αυτή η παραπάνω απόσταση, ελαχιστοποιείται με την ταύτιση του τυχαίου διανύσματος \vec{v} με το νορμαλισμένο συνισταμένο διάνυσμα και μάλιστα σε αυτή την περίπτωση λαμβάνει την τιμή $(n - R)$ όπως προείπαμε.

Έστω ότι τα μοναδιαία διανύσματα των παρατηρήσεων μας είναι της μορφής $\vec{u}_i = (\cos \alpha_i, \sin \alpha_i)$ και το τυχαίο μοναδιαίο διάνυσμα $v = (p, q)$.

Σε αυτή την περίπτωση ισχύει:

$$p^2 + q^2 = 1 \Rightarrow 2p(dp) + 2q(dq) = 0 \Rightarrow \frac{dq}{dp} = -\frac{p}{q}.$$

Παραγωγίζοντας λοιπόν την ποσότητα D_v^- , όπως ορίστηκε στην σχέση (1.9), ως προς p και εξισώνοντας με το μηδέν, σύμφωνα με τον παραπάνω λόγο προκύπτει:

$$\begin{aligned} \frac{d(D_v^-)}{dp} &= \frac{d\left(n - p \sum_{i=1}^n \cos \alpha_i - q \sum_{i=1}^n \sin \alpha_i\right)}{dp} = 0 \Rightarrow -\sum_{i=1}^n \cos \alpha_i - \frac{dq}{dp} \sum_{i=1}^n \sin \alpha_i = 0 \Rightarrow \\ &\Rightarrow -\frac{dq}{dp} = \frac{\sum_{i=1}^n \cos \alpha_i}{\sum_{i=1}^n \sin \alpha_i} \stackrel{(1.4)}{\Rightarrow} \frac{p}{q} = \frac{C}{S}. \end{aligned}$$

Επιστρέφοντας στην σχέση $p^2 + q^2 = 1$ έχουμε με αντικατάσταση :

$$\frac{C^2}{S^2} = \frac{p^2}{1 - p^2} \Rightarrow \begin{cases} p = \frac{C}{R} \\ q = \frac{S}{R} \end{cases}.$$

Άρα, έχουμε την ελάχιστη διασπορά για το τυχαίο διάνυσμα $\vec{v}^* = \left(\frac{C}{R}, \frac{S}{R}\right)$ που μάλιστα μας δίνει:

$$\begin{aligned} D_{\vec{v}^*}^- &= \sum_{i=1}^n d(\vec{v}^*, \vec{u}_i) = n - \sum_{i=1}^n \cos \theta_i = n - \sum_{i=1}^n (\vec{v}^* \cdot \vec{u}_i) = \\ &= n - \sum_{i=1}^n \left(\frac{C}{R} \cos \alpha_i + \frac{S}{R} \sin \alpha_i\right) = n - \left(\frac{C^2}{R} + \frac{S^2}{R}\right) = n - R. \end{aligned} \tag{1.10}$$

Αυτό δείχνει ότι όντως η ποσότητα $(n-R)$ είναι κατάλληλο μέτρο διασποράς του δείγματος.

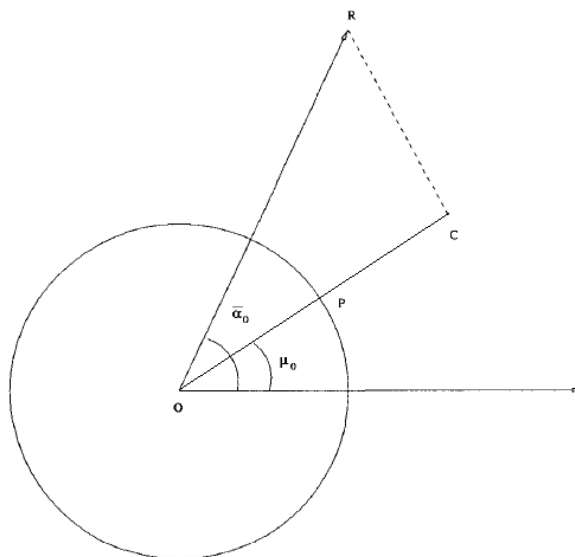
Σημείωση: Στο σημείο αυτό πρέπει να τονίσουμε ότι οι παραπάνω ορισμοί αφορούν τη διασπορά γύρω από τη μέση κατεύθυνση χωρίς να λαμβάνεται καθόλου υπόψη ο παράγοντας του πληθυσμού. Στην περίπτωση που μας δίνεται η μέση κατεύθυνση μ_0 ολόκληρου του πληθυσμού, η οποία καλείται πολική κατεύθυνση ορίζουμε το πολικό διάνυσμα $\vec{P} = (\cos \mu_0, \sin \mu_0)$ και υπολογίζουμε την ποσότητα:

$$D_{\vec{p}} = n - \sum_{i=1}^n \cos(\alpha_i - \mu_0) = n - V_0. \quad (1.11)$$

Ως V ορίζουμε το μήκος της προβολής του συνιστάμενου διανύσματος στην πολική κατεύθυνση. Όπως παρατηρούμε και στο Διάγραμμα 1-6 η ποσότητα αυτή είναι το μήκος του ευθύγραμμου τμήματος (OC) .

Πιο συγκεκριμένα ισούται με:

$$\begin{aligned} V_0 &= \sum_{i=1}^n (\cos \alpha_i \cos \mu_0 + \sin \alpha_i \sin \mu_0) = C \cos \mu_0 + S \sin \mu_0 = \\ &= R \cos \bar{\alpha}_0 \cos \mu_0 + R \sin \bar{\alpha}_0 \sin \mu_0 = R \cos(\bar{\alpha}_0 - \mu_0) \equiv Rc. \end{aligned}$$



Διάγραμμα 1-6: Προβολή συνιστάμενου διανύσματος στην πολική κατεύθυνση

Γνωρίζοντας μάλιστα ότι η ποσότητα c είναι φραγμένη από τη μονάδα, συμπεραίνουμε ότι $V_0 \leq R$ με την περίπτωση της ισότητας να ισχύει αν και μόνο αν η δειγματική μέση κατεύθυνση ταυτίζεται με την πολική κατεύθυνση, δηλαδή $\bar{\alpha}_0 = \mu_0$.

Κεφάλαιο 2

Κυκλικές Κατανομές

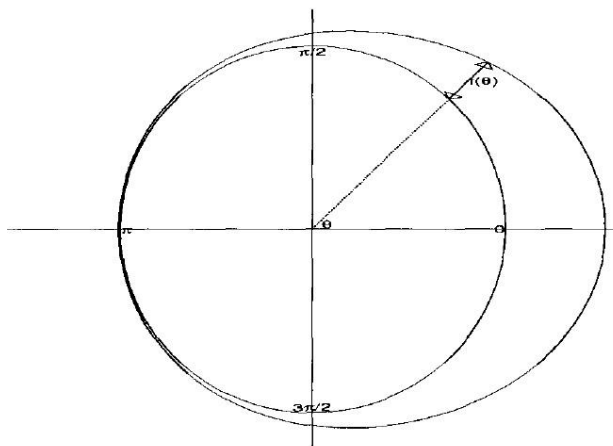
2.1 Εισαγωγή

Στην στατιστική ανάλυση των κυκλικών δεδομένων η λογική που ακολουθείται δεν παρουσιάζει διαφορές με αυτή των γραμμικών δεδομένων. Στο προηγούμενο κεφάλαιο αναλύσαμε τον τρόπο με τον οποίο ένα δείγμα κυκλικών δεδομένων ερμηνεύεται γραφικά και το κατά πόσο τα βασικά μέτρα θέσης και μεταβλητότητας διαφέρουν με την προσέγγιση των αντίστοιχων γραμμικών. Σε αυτό το κεφάλαιο καλούμαστε να αναλύσουμε τις διάφορες κυκλικές κατανομές πιθανότητας, τους μηχανισμούς με τους οποίους προκύπτουν και φυσικά τους βασικούς εκπροσώπους τους, που χρησιμοποιούνται πιο συχνά σε μελέτες που εμπεριέχουν κυκλικές μεταβλητές.

2.2 Συνάρτηση πιθανότητας κυκλικών μεταβλητών

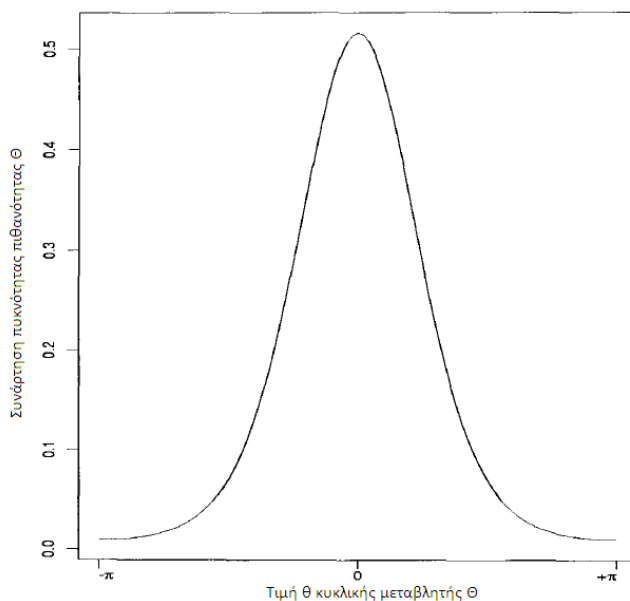
Η κυκλική κατανομή πιθανότητας είναι μια κατανομή στην οποία η συνολική πιθανότητα είναι συγκεντρωμένη στην περιφέρεια ενός μοναδιαίου κύκλου. Όπως έχουμε ήδη αναφέρει στο προηγούμενο κεφάλαιο κάθε σημείο της περιφέρειας του κύκλου αναπαριστά μία συγκεκριμένη κατεύθυνση και έτσι η κυκλική κατανομή πιθανότητας είναι στην ουσία ο μηχανισμός που αντιστοιχεί σε κάθε κατεύθυνση την πιθανότητα να παρουσιαστεί στις παρατηρήσεις μας. Οι κυκλικές κατανομές, όπως και οι γραμμικές, μπορούν να είναι είτε διακριτές είτε συνεχείς. Στην τελευταία περίπτωση, εάν Θ μια κυκλική τυχαία μεταβλητή η οποία παίρνει τιμές στο διάστημα $[0, 2\pi]$, η συνάρτηση πυκνότητας πιθανότητας f οφείλει να υπακούει στις ακόλουθες ιδιότητες:

- $f(\theta) \geq 0, \forall \theta \in [0, 2\pi]$
- $\int_0^{2\pi} f(\theta) d\theta = 1$
- $f(\theta) = f(\theta + 2k\pi), \forall k \text{ ακέραιο}$



Διάγραμμα 2-1(α): Κυκλική αναπαράσταση συνεχούς κυκλικής κατανομής

Στο Διάγραμμα 2-1, παρατηρούμε τη συνάρτηση πυκνότητας πιθανότητας μιας τέτοιας συνεχούς κυκλικής κατανομής, τόσο σε γραμμική αναπαράσταση [2-1(β)] όσο και σε κυκλική [2-1(α)]. Στα δύο αυτά διαγράμματα αποτυπώνεται η εφαρμογή των τριών παραπάνω ιδιοτήτων ενώ εύκολα μπορούμε να παρατηρήσουμε ποια κατεύθυνση φαίνεται να είναι η «επικρατέστερη».



Διάγραμμα 2-1(β): Γραμμική αναπαράσταση συνεχούς κυκλικής κατανομής

2.3 Χαρακτηριστική συνάρτηση

Σε πλήρη αντιστοίχιση με την γραμμική περίπτωση, μια κυκλική κατανομή μπορεί να περιγραφεί από ένα από τα πιο ισχυρά εργαλεία της Θεωρίας Πιθανοτήτων, την χαρακτηριστική της συνάρτηση. Εάν Θ κυκλική συνεχής τυχαία μεταβλητή, θ τυχαία τιμή της που λαμβάνει τιμές στο διάστημα $[0, 2\pi]$ και $F(\theta)$ η αθροιστική κατανομή πιθανότητας της ορίζουμε τις παρακάτω ποσότητες:

- Μέση τιμή της τυχαίας μεταβλητής Θ :

$$E_{\Theta}(\theta) = \int_0^{2\pi} \theta dF(\theta).$$

- Διασπορά της τυχαίας μεταβλητής Θ :

$$V_{\Theta}(\theta) = E_{\Theta}[(\theta - E_{\Theta}(\theta))^2].$$

Ορίζουμε βάσει των παραπάνω όπως και για τις «γραμμικές μεταβλητές», την ακόλουθη συνάρτηση ως χαρακτηριστική συνάρτηση της μεταβλητής μας:

$$\varphi_{\Theta}(p) = E_{\Theta}[e^{ip\theta}], \forall p \in \mathbb{R}.$$

Στην περίπτωση των κυκλικών μεταβλητών όμως παίζει σημαντικό ρόλο η περιοδικότητα και συνεπώς η χαρακτηριστική συνάρτηση της μεταβλητής Θ δεν μπορεί παρά να ταυτίζεται με αυτή της $\Theta' = \Theta + 2\pi$. Κατά συνέπεια έχουμε:

$$\varphi_{\Theta}(p) = E_{\Theta}[e^{ip\theta}] = E_{\Theta}[e^{ip\theta'}] = E_{\Theta}[e^{ip(\theta+2\pi)}] = e^{ip2\pi} \cdot \varphi_{\Theta}(p).$$

Αυτό συμβαίνει όταν $e^{ip2\pi} = 1$ ή με άλλα λόγια μόνο για ακέραιες τιμές του p .

Συνεπώς για κάθε τυχαία κυκλική μεταβλητή Θ ορίζουμε την τιμή της χαρακτηριστικής συνάρτησης στον ακέραιο p ως την p -τάξης τριγωνομετρική ροπή της Θ .

$$\varphi_{\Theta}(p) = E_{\Theta}(e^{ip\theta}) = \int_0^{2\pi} e^{ip\theta} dF(\theta) \equiv \rho_p e^{i\mu_p}, \quad p = 0, \pm 1, \pm 2, \dots \quad (2.1)$$

Από την σχέση (2.1) μπορούμε να δούμε ότι ισχύουν τα παρακάτω:

$$\varphi_{\Theta}(0) = 1, |\varphi_{\Theta}(p)| = \rho_p = \|E(e^{ip\theta})\| \leq E(\|e^{ip\theta}\|) = 1.$$

Μια άλλη γραφή των τριγωνομετρικών ροπών είναι η εξής:

$$\begin{aligned} \varphi_{\Theta}(p) &= E_{\Theta}[\cos p\theta + i \sin p\theta] \equiv \alpha_p + i\beta_p, \\ \text{όπου } \alpha_p &\equiv \alpha_{\Theta}(p) = E_{\Theta}(\cos p\theta), \quad \beta_p \equiv \beta_{\Theta}(p) = E_{\Theta}(\sin p\theta). \end{aligned} \quad (2.2)$$

Από την σχέση (2.1) προκύπτει φυσικά ότι:

$$\rho_p = \sqrt{\alpha_p^2 + \beta_p^2} \quad \text{και} \quad \mu_p = \arctan^* \left(\frac{\beta_p}{\alpha_p} \right).$$

Ιδιαίτερη σημασία έχει η τριγωνομετρική ροπή πρώτης τάξης:

$$\varphi_{\Theta}(1) = \alpha_1 + i\beta_1 = \rho_1 e^{i\mu_1}. \quad (2.3)$$

Συμβολίζουμε τότε ως $\rho_1 \equiv \rho$ και $\mu_1 \equiv \mu$ και αυτές οι ποσότητες αντιστοιχούν στη συγκέντρωση και μέση κατεύθυνση του πληθυσμού. Στην ειδική περίπτωση που το ρ προσεγγίζει την μονάδα, έχουμε εκφυλισμένη κατανομή με συγκεντρωμένη πιθανότητα στην κατεύθυνση μ , ενώ στην αντίθετη περίπτωση που το ρ τείνει στο 0, έχουμε ισοτροπική ή ομοιόμορφη κατανομή. Τέλος, το εμπειρικό ανάλογο αυτών των δύο ποσοτήτων για το δείγμα μας είναι το $\left(\frac{R}{n}\right)$ και το $\bar{\alpha}_0$ αντίστοιχα.

Η σημασία των τριγωνομετρικών ροπών p -τάξης, δηλαδή της χαρακτηριστικής συνάρτησης της κυκλικής μεταβλητής μας έγκειται, στο γεγονός ότι συμπίπτουν με τους συντελεστές της σειράς Fourier, στην οποία αναπτύσσεται η συνάρτηση κατανομής πιθανότητας της Θ .

Συγκεκριμένα η κυκλική κατανομή πιθανότητας είναι διπλά παραγωγίσιμη στο διάστημα $[0, 2\pi)$, μπορούμε γνωρίζοντας τις τριγωνομετρικές ροπές να προσδιορίσουμε επακριβώς την συνάρτηση κατανομής πιθανότητας της κυκλικής μεταβλητής μας μέσω της ακόλουθης σειράς Fourier:

$$f(\theta) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \varphi_p e^{ip\theta} = \frac{1}{2\pi} \left[1 + 2 \sum_{i=1}^{\infty} (\alpha_p \cos p\theta + \beta_p \sin p\theta) \right]. \quad (2.4)$$

2.4 Μέθοδοι κατασκευής κυκλικών κατανομών

Όπως και στην κλασική στατιστική το βασικό εργαλείο για την ανάλυση των κυκλικών δεδομένων είναι ο προσδιορισμός της κατανομής που ακολουθούν. Στο πρώτο κεφάλαιο της διπλωματικής καταστήσαμε σαφές ότι μια κυκλική τυχαία μεταβλητή μπορεί να αναπαρασταθεί μοναδικά είτε σε πολικές συντεταγμένες είτε στο καρτεσιανό σύστημα μέσω της σχέσης (1.1). Οι κυκλικές κατανομές κατά συνέπεια δεν προκύπτουν με αυθαίρετο τρόπο αλλά παράγονται βάσει των αντίστοιχων κατανομών γραμμικών μεταβλητών.

Οι βασικές και συνηθέστερες μέθοδοι για την μετάβαση αυτή είναι:

- Περίελιξη γραμμικής κατανομής σε μοναδιαίο κύκλο

Κάθε γραμμική τυχαία μεταβλητή X μπορεί να μετατραπεί σε κυκλική τυχαία μεταβλητή Θ μέσω του ακόλουθου τύπου: $\Theta = X \pmod{2\pi}$.

Αυτή η διαδικασία αντιστοιχεί στην περιέλιξη του άξονα των πραγματικών αριθμών γύρω από μοναδιαίο κύκλο, αποδίδοντας πιθανότητες πάνω σε όλα τα υπερκαλυπτόμενα σημεία $x = \theta, \theta \pm 2\pi, \theta \pm 4\pi, \dots$. Δημιουργείται συνεπώς μία πολλά-προς-ένα συνάρτηση τέτοια ώστε, εάν $g(\theta)$ η συνάρτηση πυκνότητας πιθανότητας της κυκλικής μεταβλητής και $f(x)$ η αντίστοιχη γραμμική, τότε να ισχύει:

$$g(\theta) = \sum_{m=-\infty}^{\infty} f(\theta + 2\pi m), 0 \leq \theta \leq 2\pi. \quad (2.5)$$

Η συνάρτηση αυτή που συσχετίζει τις δύο μεταβλητές «μεταφέρει» και μια χρήσιμη ιδιότητα που μπορεί να μας βοηθήσει στη στατιστική ανάλυση της κατανομής που ακολουθεί η κυκλική μεταβλητή.

Η τριγωνομετρική ροπή p -τάξης της περιελιγμένης κυκλικής κατανομής αντιστοιχεί στην τιμή της χαρακτηριστικής συνάρτησης της γραμμικής μεταβλητής X .

Απόδειξη: Εάν G, F οι αθροιστικές συναρτήσεις κατανομής των Θ και X αντίστοιχα και φ_x η χαρακτηριστική συνάρτηση της X , τότε ισχύει:

$$\varphi_{\Theta}(p) = \int_0^{2\pi} e^{ip\theta} dG_{\Theta}(\theta) = \sum_{k=-\infty}^{\infty} \int_{2\pi k}^{2\pi(k+1)} e^{ip\theta} dF_X(\theta) = \int_{-\infty}^{\infty} e^{ip\theta} dF_X(x) = \varphi_x(p).$$

Η συγκεκριμένη ιδιότητα μάλιστα μας επιβεβαιώνει το γεγονός ότι στις κυκλικές μεταβλητές η συνάρτηση κατανομής πιθανότητας καθορίζεται πλήρως από την χαρακτηριστική συνάρτηση. Αυτό εξηγείται διαισθητικά από το γεγονός ότι στις κυκλικές μεταβλητές η πυκνότητα πιθανότητας δεν μπορεί να απειριστεί.

- Βάσει χρήσιμων ιδιοτήτων

Είναι αρκετά χρήσιμο να κατηγοριοποιούνται οι κυκλικές κατανομές βάσει χαρακτηριστικών ιδιοτήτων. Κάποιες αρκετά συχνές και ενδιαφέρουσες ιδιότητες είναι αυτές της μέγιστης εντροπίας (μη μηδενική τριγωνομετρική ροπή πρώτης τάξης), και της ανεξαρτησίας του μήκους του συνιστάμενου διανύσματος και της δειγματικής μέσης κατεύθυνσης. Τέλος, ιδιαίτερη προσοχή δίνεται εφόσον η μέση κατεύθυνση εκτιμάται με μέγιστη πιθανοφάνεια από το εν λόγω μήκος.

- Μετατροπή διμετάβλητης κατανομής τυχαίας γραμμική μεταβλητής στην κατευθυνόμενη συνιστώσα της.

Η διαδικασία της μετατροπής αυτής γίνεται μετατρέποντας τις ορθογώνιες συντεταγμένες του διμετάβλητου τυχαίου διανύσματος σε πολικές και ολοκληρώνοντας ως προς R για δεδομένο θ . Εάν $f(X, Y)$ η από κοινού κατανομή πιθανότητας της διμετάβλητης κατανομής, τότε η συνάρτηση κατανομής πιθανότητας της θ που προκύπτει:

$$g(\theta) = \int_0^{\infty} f(r \cos \theta, r \sin \theta) r dr. \quad (2.6)$$

2.5 Ομοιόμορφη κατανομή

Η κατανομή αυτή είναι η πιο απλή κυκλική κατανομή και αναφέρεται συχνά στην βιβλιογραφία ως το μηδενικό υπόδειγμα (null model). Στην περίπτωση της ομοιόμορφης κατανομής, η συνολική πιθανότητα είναι όμοια κατανεμημένη σε όλη την περιφέρεια του κύκλου, με αποτέλεσμα να μην υπάρχει «επικρατέστερη» κατεύθυνση. Αυτό φυσικά σημαίνει ότι η συγκέντρωση του πληθυσμού όπως ορίστηκε στην σχέση (2.3) είναι μηδενική, $\rho = 0$, και προκύπτει ότι όλοι οι συντελεστές των τριγωνομετρικών ροπών κάθε τάξης είναι μηδενικοί, εκτός του α_0 που ισούται με την μονάδα.

Η συνάρτηση πυκνότητας πιθανότητας είναι φυσικά ανεξάρτητη της τιμής της κυκλικής μεταβλητής Θ και δίνεται από τον τύπο:

$$f(\theta) = \frac{1}{2\pi}, 0 \leq \theta < 2\pi. \quad (2.7)$$

Οφείλουμε να τονίσουμε, ότι παρόλη την απλότητα που εμφανίζει αυτή η κατανομή συνδυάζει μια σειρά από χρήσιμες ιδιότητες όπως, της μέγιστης εντροπίας ενώ είναι η μοναδική κατανομή, στην οποία το μήκος του συνιστάμενου διανύσματος του δείγματος R είναι ανεξάρτητο της μέσης κατεύθυνσης $\bar{\alpha}_0$.

2.6 Καρδιοειδής κατανομή

Η κατανομή αυτή είναι μονοκόρυφη και συμμετρική γύρω από την παράμετρο μ . Έχει δε την ακόλουθη συνάρτηση πυκνότητας πιθανότητας:

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} \{1 + 2\rho \cos(\theta - \mu)\}, 0 \leq \mu < 2\pi, -\frac{1}{2} \leq \rho < \frac{1}{2}. \quad (2.8)$$

Οι τριγωνομετρικές ροπές της δίνονται από τους ακόλουθους τύπους:

$$\begin{cases} \varphi_{\ominus}(1) = |\rho| e^{i\mu} \\ \varphi_{\ominus}(p) = 0, p \geq 2 \end{cases},$$

έτσι ώστε η παράμετρος ρ να υποδηλώνει την συγκέντρωση και η μ την μέση κατεύθυνση αντίστοιχα.

2.7 Τριγωνική κατανομή

Αυτή η κατανομή, όπως υποδηλώνει και το όνομα της, έχει συνάρτηση πυκνότητας πιθανότητας σε σχήμα τριγώνου και συμμετρική γύρω από το σημείο $\theta = 0$. Δίνεται δε από τον εξής τύπο:

$$f(\theta; \rho) = \frac{1}{8\pi} \{4 - \pi^2 \rho + 2\pi\rho |\pi - \theta|\}, \quad 0 \leq \theta < 2\pi, \quad 0 \leq \rho \leq \frac{4}{\pi^2}. \quad (2.9)$$

Οι τριγωνομετρικές ροπές της συγκεκριμένης κατανομής είναι:

$$\begin{cases} \varphi_{\ominus}(1) = \rho \\ \varphi_{\ominus}(2p-1) = \frac{\rho}{(2p-1)^2} \end{cases}.$$

Έτσι, παρατηρούμε ότι η μέση κατεύθυνση είναι η μηδενική και η μεταβλητή ρ είναι δείκτης της συγκέντρωσης γύρω από αυτή.

2.8 Κυκλική κανονική κατανομή (VonMises)

Η πιο διαδεδομένη κατανομή στα προβλήματα κυκλικών δεδομένων είναι η κυκλική κανονική κατανομή (CN ή VonMises). Η συνάρτηση κατανομής της είναι:

$$f(\theta; \mu, \kappa) = \left(\frac{1}{2\pi I_0(\kappa)} \right) \exp(\kappa \cos(\theta - \mu)), \quad 0 \leq \theta \leq 2\pi, \quad 0 \leq \mu \leq 2\pi, \quad \kappa \geq 0. \quad (2.10)$$

Η εν λόγω κατανομή εξαρτάται από δύο παραμέτρους, την μέση κατεύθυνση μ και την παράμετρο συγκέντρωσης κ , ενώ το $I_0(\kappa)$ είναι η σταθερά κανονικοποίησης της τροποποιημένης συνάρτησης Bessel (βλέπε παράρτημα) πρώτου είδους και μηδενικής τάξης και δίνεται από τον ακόλουθο τύπο:

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos \theta) d\theta = \sum_{r=0}^{\infty} \left(\frac{\kappa}{2} \right)^{2r} \left(\frac{1}{r!} \right)^2. \quad (2.11)$$

Η κατανομή αυτή είναι έχει μελετηθεί περισσότερο από οποιαδήποτε άλλη κυκλική κατανομή και χρησιμοποιείται ευρέως, καθώς συγκεντρώνει μια σειρά από σημαντικές ιδιότητες:

1. Είναι συμμετρική γύρω από την μέση κατεύθυνση μ και φυσικά από την $\mu - \pi$.

2. Είναι μονοκόρυφη κατανομή με κορυφή το μ και μέγιστη τιμή εκεί την:

$$f(\mu) = \frac{e^{\kappa}}{2\pi I_0(\kappa)}.$$

3. Εμφανίζει λόγω της εξάρτησης από την συνάρτηση του συνημιτόνου «αντίθετη» μέση κατεύθυνση και κορυφή στα σημεία $\mu \pm \pi$ με ελάχιστη

$$\text{πυκνότητα εκεί την: } f(\mu \pm \pi) = \frac{e^{-\kappa}}{2\pi I_0(\kappa)}.$$

4. Από τις δύο παραπάνω ιδιότητες προκύπτει και η σημασία της παραμέτρου συγκέντρωσης κ , καθώς ισχύει: $\frac{f(\mu)}{f(\mu \pm \pi)} = e^{2\kappa}$.

Συνεπώς, όσο μεγαλύτερη η τιμή του κ τόσο μεγαλύτερη συγκέντρωση εμφανίζει η κατανομή γύρω από την μέση κατεύθυνση του πληθυσμού.

5. Γνωρίζουμε ότι εάν $\theta = \alpha - \mu$, ισχύουν οι ακόλουθες σχέσεις:

$$\frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) \exp(\kappa \cos\theta) d\theta = I_p(\kappa),$$

$$\frac{1}{2\pi} \int_0^{2\pi} \sin(n\theta) \exp(\kappa \cos\theta) d\theta = 0, \quad \forall \kappa \geq 0, 0 \leq \theta < 2\pi.$$

Από τις δύο αυτές σχέσεις μπορούμε να υπολογίσουμε τις τριγωνομετρικές ροπές ως εξής:

$$\begin{aligned} \varphi_{\Theta}(p) &= \left(\frac{1}{2\pi I_0(\kappa)} \right) \int_0^{2\pi} e^{ip\alpha} \exp(\kappa \cos(\alpha - \mu)) d\alpha = \\ &= \left(\frac{1}{2\pi I_0(\kappa)} \right) \int_0^{2\pi} e^{ip(\theta + \mu)} \exp(\kappa \cos\theta) d\theta = \\ &= \left(\frac{e^{ip\mu}}{2\pi I_0(\kappa)} \right) \int_0^{2\pi} [\cos(p\theta) + i\sin(p\theta)] \exp(\kappa \cos\theta) d\theta = \\ &= \left[\begin{array}{l} I_p(\kappa) \\ I_0(\kappa) \end{array} \right] e^{ip\mu}, \quad p = 0, \pm 1, \pm 2, \pm 3, \dots \end{aligned}$$

Από αυτή τη σχέση βλέπουμε ότι το μήκος ρ του πολικού διανύσματος (μέσω της τριγωνομετρικής ροπής πρώτης τάξης) είναι μια συνάρτηση του συντελεστή συγκέντρωσης:

$$\frac{I_1(\kappa)}{I_0(\kappa)} \equiv A(\kappa), \quad (2.12)$$

όπου $I_1(\kappa) = \frac{d}{d\kappa} I_0(\kappa)$. (Η απόδειξη της σχέσης παρουσιάζεται στο παράρτημα)

Η $A(\kappa)$ είναι μια γνησίως αύξουσα συνάρτηση που εμφανίζει ελάχιστο για $\kappa = 0$ στο μηδέν και απειρίζεται όσο το κ πλησιάζει το άπειρο.

6. Η συνέλιξη δύο vonMises κατανομών δεν είναι vonMises κατανομή.
7. Από την κυκλική κανονική κατανομή μπορούμε να περάσουμε σε άλλες κυκλικές κατανομές με κατάλληλη επιλογή παραμέτρου συγκέντρωσης. Για παράδειγμα, για $\kappa = 0$ παίρνουμε την κυκλική ομοιόμορφη κατανομή. Θα αποδείξουμε ότι για κ να τείνει στο άπειρο μπορούμε να πετύχουμε προσέγγιση με την «γραμμική» κανονική κατανομή.

Απόδειξη: Έστω $\beta = \sqrt{\kappa}(\alpha - \gamma)$. Τότε για μεγάλο κ ισχύει $\cos(\alpha - \gamma) = \cos\left(\frac{\beta}{\sqrt{\kappa}}\right) \cong 1 - \frac{\beta^2}{2\kappa}$. Χρησιμοποιώντας την προσέγγιση, ξανά για μεγάλες τιμές της παραμέτρου συγκέντρωσης, $I_0(\kappa) \cong \frac{\exp(\kappa)}{\sqrt{2\pi\kappa}}$ έχουμε την ακόλουθη συνάρτηση πυκνότητας πιθανότητας της β :

$$\begin{aligned} g(\beta) &= \frac{\exp\left(\kappa \cos\left(\frac{\beta}{\sqrt{\kappa}}\right)\right)}{2\pi I_0(\kappa)} \cdot \frac{1}{\sqrt{\kappa}} \cong \\ &= \frac{\exp\left(\kappa \cos\left(\frac{\beta}{\sqrt{\kappa}}\right)\right)}{2\pi \frac{\exp(\kappa)}{\sqrt{2\pi\kappa}}} \cdot \frac{1}{\sqrt{\kappa}} \cong \\ &= \frac{\exp\left(\kappa \left(1 - \frac{\beta^2}{2\kappa}\right)\right)}{e^\kappa \sqrt{2\pi}} = \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\beta^2}{2}\right). \end{aligned} \quad (2.13)$$

2.9 Παραλλαγές κυκλικής κανονικής κατανομής

- Von Mises κατανομή σε κομμάτι της περιφέρειας του κύκλου

Μια κυκλική κανονική κατανομή μπορεί να περιοριστεί σε ένα συγκεκριμένο τόξο της περιφέρειας του κύκλου. Εάν w είναι αυθαίρετη ακέραιη σταθερά τότε για κάθε τόξο μήκους $\frac{2\pi}{w}$ η συνάρτηση πυκνότητας πιθανότητας της κυκλικής κανονικής κατανομής περιορισμένης στο συγκεκριμένο τόξο βάσει της εξίσωσης (2.10) γίνεται:

$$g(\theta) = \left(\frac{w}{2\pi I_0(\kappa)} \right) \exp \{ \kappa \cos w(\theta - \mu) \}, \begin{cases} 0 \leq \theta < \frac{2\pi}{w} \\ 0 \leq \mu < \frac{2\pi}{w} \end{cases}. \quad (2.14)$$

Ιδιαίτερο ενδιαφέρον παρουσιάζει η περίπτωση που η σταθερά w παίρνει την τιμή δύο, δηλαδή μιλάμε για μια κατανομή περιορισμένη σε ημικύκλιο. Αυτού του είδους η κατανομή καλείται *αξονική* και τα δεδομένα που την ακολουθούν αξονικά και αποτελούν υποκατηγορία των κυκλικών δεδομένων.

- Κυκλική κανονική κατανομή με w κόμβους

Οι κατανομές που έχουμε παρουσιάσει μέχρι στιγμής ήταν όλες συμμετρικές και παρουσίαζαν έναν κόμβο. Η κυκλική κανονική κατανομή με μία μετατροπή στην εξίσωση (2.10) μπορεί να αποκτήσει w ισαπέχοντες κόμβους όπου w η αυθαίρετη ακέραια σταθερά που ορίσαμε παραπάνω. Η συνάρτηση πυκνότητας πιθανότητάς της τότε λαμβάνει την ακόλουθη μορφή:

$$g(\theta) = \left(\frac{1}{2\pi I_0(\kappa)} \right) \exp \{ \kappa \cos w(\theta - \mu) \}, \begin{cases} 0 \leq \theta < \frac{2\pi}{w} \\ 0 \leq \mu < \frac{2\pi}{w} \end{cases}. \quad (2.15)$$

- Μίξεις κυκλικών κανονικών κατανομών

Κυκλικές κατανομές πιθανότητας με δύο ή και παραπάνω κόμβους μπορούν να προκύψουν από την μίξη διαφορετικών κατανομών ή ακόμη και από μίξη όμοιων κατανομών με διαφορετικές παραμέτρους. Παρότι με αυτή την μέθοδο έχουμε αυξημένη ευελιξία αλλά και εμπλουτισμό των διαθέσιμων κυκλικών κατανομών, συναντάτε ιδιαίτερη δυσκολία στη στατιστική συμπερασματολογία λόγω αυξημένου αριθμού παραμέτρων. Στην περίπτωση που έχουμε μίξη συγκεκριμένα δύο κατανομών Von Mises η συνάρτηση πυκνότητας πιθανότητας γίνεται:

$$\begin{aligned}
f(\theta; \mu, \kappa) &= \frac{1}{2} \{CN(\theta; \mu, \kappa) + CN(\theta; -\mu, \kappa)\} = \\
&= \frac{1}{4\pi I_0(\kappa)} \{e^{\kappa \cos(\theta-\mu)} + e^{\kappa \cos(\theta+\mu)}\}.
\end{aligned} \tag{2.16}$$

Αυτή η κατανομή δεν έχει απαραίτητα δύο κορυφές αλλά έχει αποδειχθεί ότι έχει μοναδική κορυφή για τις ακόλουθες τιμές της μέσης κατεύθυνσης:

$$\mu \leq \arccos \left\{ \left[-1 + \left(1 + 4\kappa^2\right)^{\frac{1}{2}} \right] \cdot (2\kappa)^{-1} \right\}.$$

2.10 Περιελιγμένες κυκλικές κατανομές

2.10.1 Περιελιγμένη κανονική κατανομή

Η περιελιγμένη κανονική κατανομή προκύπτει από την «γραμμική» κανονική κατανομή $N(\mu, \sigma^2)$ και η συνάρτηση πυκνότητας πιθανότητας της είναι:

$$\begin{aligned}
g(\theta) &= \sum_{m=-\infty}^{\infty} f(\alpha + 2m\pi) = \\
&= \frac{1}{\sigma\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} \exp \left[\frac{-(\theta - \mu - 2\pi m)^2}{2\sigma^2} \right].
\end{aligned} \tag{2.17}$$

Η συγκεκριμένη κατανομή έχει μοναδική κορυφή, εμφανίζεται συμμετρική γύρω από την τιμή μ και έχει την αθροιστική ιδιότητα, δηλαδή η συνέλιξη δύο περιελιγμένων κυκλικών κατανομών είναι περιελιγμένη κυκλική κατανομή.

2.10.2 Περιελιγμένη Cauchy κατανομή

Η περιελιγμένη Cauchy κατανομή προκύπτει από την Cauchy κατανομή στην ευθεία των πραγματικών αριθμών με συνάρτηση πυκνότητας πιθανότητας την:

$$f(x) = \left(\frac{1}{\pi} \right) \left[\frac{\sigma}{\sigma^2 + (x - \mu)^2} \right].$$

Πραγματοποιώντας την περιέλιξη της παραπάνω κατανομής στη περιφέρεια μοναδιαίου κύκλου λαμβάνουμε την ακόλουθη συνάρτηση πυκνότητας πιθανότητας:

$$\begin{aligned}
g(\theta) &= \frac{1}{2\pi} \left(1 + 2 \sum_{k=1}^{\infty} \rho^k \cos k(\theta - \mu) \right) = \\
&= \frac{1}{2\pi} \cdot \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}, \quad 0 \leq \theta < 2\pi, \rho = e^{-\sigma}.
\end{aligned} \tag{2.18}$$

Η εν λόγω κατανομή έχει μοναδική κορυφή, είναι συμμετρική και εμφανίζει και την προσθετική ιδιότητα.

2.10.3 Περιελιγμένες α -σταθερές κατανομές

Η οικογένεια των περιελιγμένων α -σταθερών κατανομών προκύπτει από την οικογένεια των α -σταθερών γραμμικών κατανομών και έχει συνάρτηση πυκνότητας πιθανότητας την:

$$\begin{aligned}
f(\theta) &= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \exp\{-\tau^\alpha k^\alpha\} \cos \left\{ k(\theta - \mu) - \tau^\alpha k^\alpha \beta \tan \left(\frac{\alpha\pi}{2} \right) \right\} \\
\tau &\geq 0, \quad |\beta| \leq 1, \quad \alpha \in (0,1) \cup (1,2], \quad \mu \text{ πραγματική σταθερά,}
\end{aligned} \tag{2.19}$$

όπου θ τιμή κυκλικής τυχαίας μεταβλητής στο διάστημα $[0, 2\pi)$.

Στην περίπτωση που ισχύει $\beta = 0$, τότε προκύπτει η οικογένεια των *συμμετρικών περιελιγμένων σταθερών κατανομών (WS)*, στην οποία ανήκουν και οι παραπάνω περιελιγμένες κατανομές που αναφέραμε. Η μεν κανονική για $\alpha = 2$, $\rho = e^{-\frac{\sigma^2}{2}}$ η δε Cauchy για $\alpha = 1$, $\rho = e^{-\tau}$.

Η οικογένεια των περιελιγμένων α -σταθερών κατανομών είναι μια πολύ σημαντική υπό-κατηγορία των κυκλικών κατανομών καθώς μεταφέρει τις ιδιαίτερα χρήσιμες ιδιότητες των α -σταθερών «γραμμικών» κατανομών λόγω της σχέσης της χαρακτηριστικής συνάρτησης με τους συντελεστές Fourier καθώς και επειδή μας δίνει την δυνατότητα επιλογής ενός ιδιαίτερα προσαρμόσιμου μοντέλου, πράγμα που η περιελιγμένη κανονική και Cauchy δεν μας επιτρέπουν ακόμη και στην περίπτωση επιλογής συμμετρικών συναρτήσεων πυκνότητας πιθανότητας.

2.11 Κυκλική βήτα κατανομή

Έστω μια συνεχής «γραμμική» τυχαία μεταβλητή X . Λέμε ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή βήτα με παραμέτρους α, β εάν για κάθε τιμή της x ισχύει:

$$\begin{cases} f(x) = \frac{1}{B(\alpha, \beta)} x^\alpha (1-x)^{\beta-1}, \quad \forall 0 < x < 1, \alpha > 0, \beta > 0, \\ f(x) = 0 & \forall x < 0, x > 1 \end{cases},$$

όπου $B(\alpha, \beta)$ η συνάρτηση βήτα που ορίζεται με τον εξής τύπο:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \mu\epsilon \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \quad \forall z > 0.$$

Με βάση αυτή την κατανομή πιθανότητας λαμβάνουμε και την κυκλική βήτα κατανομή. Έστω κυκλική τυχαία μεταβλητή Θ που ακολουθεί την κυκλική βήτα κατανομή και τυχαίο δείγμα παρατηρήσεων της $\{\theta_1, \theta_2, \dots, \theta_n\}$. Υποθέτοντας ότι $\bar{\mu}, \bar{\theta}$ είναι τα διανύσματα της μέσης κατεύθυνσης και των παρατηρήσεων, υπό την προϋπόθεση ότι $\alpha \geq \beta$, η συνάρτηση πυκνότητας πιθανότητας της κυκλικής βήτα κατανομής είναι:

$$f(\bar{\theta}) = \frac{1}{2^{\alpha+\beta} B(\alpha, \beta)} (1 + \bar{\mu}'\bar{\theta})^{\alpha-\frac{1}{2}} (1 - \bar{\mu}'\bar{\theta})^{\beta-\frac{1}{2}}, \quad \bar{\theta} \in S^1. \quad (2.20)$$

Η συγκεκριμένη κυκλική κατανομή πιθανότητας έχει τη ιδιαιτερότητα ότι για διαφορετικές τιμές των παραμέτρων της συνάρτησης βήτα εμφανίζει διαφορετικά ποιοτικά χαρακτηριστικά. Συγκεκριμένα για τιμές $\alpha, \beta \geq 0.5$, η συνάρτηση πυκνότητας πιθανότητας είναι μη-φραγμένη και εμφανίζει δύο κορυφές στις κατευθύνσεις 0° και 180° ενώ για τις τιμές $\alpha = 0.5, \beta \leq 0.5$ η συνάρτηση γίνεται φραγμένη με μόνο μία κορυφή στις 0° . Τέλος εάν ισχύει $\alpha = \beta = 0.5$, τότε η συνάρτηση πυκνότητας πιθανότητας εκφυλίζεται στην αντίστοιχη της κυκλικής ομοιόμορφης κατανομής.

Κεφάλαιο 3

Κατανομές δειγματικών ποσοτήτων

3.1 Εισαγωγή

Στα πλαίσια της ορθής στατιστικής συμπερασματολογίας, βάσει γνωστού στατιστικού μοντέλου, είναι απαραίτητη η συγκέντρωση επαρκών πληροφοριών. Αυτή η πληροφορία αποκτάται από δείγμα, προερχόμενο από το εν λόγω μοντέλο και ακολούθως από τον προσδιορισμό της κατανομής που ακολουθούν βασικές στατιστικές του συναρτήσεις.

Τα πιο διαδεδομένα μοντέλα για τα οποία μπορούμε να αποκτήσουμε την παραπάνω πληροφορία, όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, είναι αυτά της κυκλικής ομοιόμορφης και κυκλικής κανονικής κατανομής. Βάσει της ανάλυσης του προηγούμενου κεφαλαίου, δεδομένου ότι έχουμε ένα δείγμα ανεξάρτητων και ομοιόμορφα κατανεμημένων παρατηρήσεων από κάποια από τις δύο κατανομές, οι ποσότητες που είναι καθόλα χρήσιμες στην στατιστική συμπερασματολογία και για τις οποίες θα χρειαστεί να προσδιορίσουμε την κατανομή τους είναι οι C, S, R , όπως ορίστηκαν στις σχέσεις (1.4) και (1.5). Μάλιστα, προκύπτει ότι η από κοινού κατανομή των (C, S) μας δίνει την αντίστοιχη των $(\bar{\alpha}_0, R)$, λόγω του ένα προς ένα μετασχηματισμού που ορίζεται στην σχέση (1.6).

3.2 Εύρεση δειγματικών κατανομών

3.2.1 Δείγμα από κυκλική ομοιόμορφη κατανομή

- **Η κατανομή του R**

Αφετηρία για τον προσδιορισμό της κατανομής αυτής αποτέλεσε, ο λεγόμενος τυχαίος περίπατος του Pearson. Εκεί ένας τυχαίος περίπατος ξεκινούσε από δοθέν σημείο ακολουθώντας διαδρομές μήκους $(\alpha_1, \alpha_2, \dots, \alpha_n)$ σε τυχαίες διακριτές κατευθύνσεις. Εάν θέσουμε την ποσότητα R ως το μήκος από το αρχικό στο τελικό σημείο και στην περίπτωση που οι n διαδρομές προέρχονται από την κυκλική κανονική κατανομή, τότε προκύπτει με χρήση συναρτήσεων Bessel η εξής συνάρτηση κατανομής:

$$f_0(r) = r \int_0^{\infty} J_0(rt) J_0^n(t) dt \equiv r \psi_n(r), \quad 0 \leq r \leq n,^2 \quad (3.1)$$

όπου $J_0(\cdot)$ η συνάρτηση Bessel μηδενικής τάξης όπως ορίζεται στην σχέση (A.1.1) του παραρτήματος.

• Η κατανομή του (C,S)

Βάσει του ορισμού των C,S στην σχέση (1.4), και των ποσοτήτων $R, \bar{\alpha}_0$ ως το συνιστάμενο διάνυσμα και η κατεύθυνση του αντίστοιχα, γνωρίζουμε ότι ισχύει ο ακόλουθος ένα προς ένα μετασχηματισμός: $C = R \cos \bar{\alpha}_0, S = R \sin \bar{\alpha}_0$. Λαμβάνοντας υπόψη ότι $dc ds = r dr d\bar{\alpha}_0$, έχουμε, βάσει της κατανομής του R , την ακόλουθη κατανομή :

$$f_0(c, s) = \frac{1}{2\pi} \psi_n(\sqrt{c^2 + s^2}). \quad (3.2)$$

• Η κατανομή του V

Έστω το μήκος της προβολής του συνιστάμενου διανύσματος στην πολική κατεύθυνση του δείγματος, ορισμένη ως εξής:

$$V = \sum_{i=1}^n \cos(\alpha_i - \mu), \quad (3.3)$$

όπου μ η μέση κατεύθυνση του δείγματος μας.

Γνωρίζουμε ότι η χαρακτηριστική συνάρτηση του συνημιτόνου είναι: $\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{it \cos \alpha} da = \int_0^{2\pi} \frac{1}{2\pi} e^{it \cos \alpha} da = J_0(t)$. Όπως έχουμε παραδεχθεί στην εισαγωγή του κεφαλαίου οι παρατηρήσεις μας είναι ανεξάρτητες και ισόνομα κατανεμημένες, άρα η χαρακτηριστική συνάρτηση μιας τιμής v της τυχαίας μεταβλητής V , είναι το γινόμενο $J_0^n(t)$. Βάσει αυτών, η συνάρτηση πυκνότητας πιθανότητας του V προκύπτει:

$$f_0(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itv} J_0^n(t) dt = \frac{1}{\pi} \int_0^{\infty} \frac{e^{itv} + e^{-itv}}{2} J_0^n(t) dt = \frac{1}{\pi} \int_0^{\infty} J_0^n(t) \cos(vt) dt. \quad (3.4)$$

² Η πλήρης απόδειξη καθώς και κάποιες βασικές ιδιότητες των συναρτήσεων Bessel παρατίθενται στο παράρτημα στο τέλος της διπλωματικής εργασίας.

3.2.2 Δείγμα από κυκλική κανονική κατανομή

Έστω κατευθύνσεις $(\alpha_1, \alpha_2, \dots, \alpha_n)$ από την κυκλική κανονική κατανομή, με παραμέτρους μ και κ . Σε αυτή την περίπτωση, η από κοινού συνάρτηση πυκνότητας πιθανότητας των παρατηρήσεων είναι:

$$\begin{aligned} f_{\kappa}(\alpha_1, \alpha_2, \dots, \alpha_n) &= \frac{1}{I_0^n(\kappa)(2\pi)^n} \exp\left(\kappa \cos \mu \sum_{i=1}^n \cos \alpha_i + \kappa \sin \mu \sum_{i=1}^n \sin \alpha_i\right) = \\ &= \frac{1}{I_0^n(\kappa)} \exp\left(\kappa \cos \mu \sum_{i=1}^n \cos \alpha_i + \kappa \sin \mu \sum_{i=1}^n \sin \alpha_i\right) f_0(\alpha_1, \alpha_2, \dots, \alpha_n), \end{aligned} \quad (3.5)$$

όπου $f_0(\alpha_1, \alpha_2, \dots, \alpha_n)$ η από κοινού συνάρτηση πυκνότητας πιθανότητας για τα δεδομένα υπό την κυκλική ομοιόμορφη κατανομή. Βάσει αυτής της σχέσης μπορούμε να προσδιορίσουμε τις επιθυμητές κατανομές χρησιμοποιώντας τις αντίστοιχες της ομοιόμορφης κατανομής.

• Η κατανομή του (C,S)

Χρησιμοποιώντας την ιδιότητα της σχέσης (3.4) έχουμε:

$$\begin{aligned} f_{\kappa}(c, s) &= \int_F f_{\kappa}(\alpha_1, \alpha_2, \dots, \alpha_n) d\alpha_1 d\alpha_2 \dots d\alpha_n = \\ &= \frac{\exp[(\kappa \cos \mu)c + (\kappa \sin \mu)s]}{I_0^n(\kappa)} \int_F f_0(\alpha_1, \alpha_2, \dots, \alpha_n) d\alpha_1 d\alpha_2 \dots d\alpha_n = \\ &= \frac{\exp[(\kappa \cos \mu)c + (\kappa \sin \mu)s]}{I_0^n(\kappa)} f_0(c, s), \end{aligned} \quad (3.6)$$

όπου $f_0(c, s)$ όπως ορίστηκε στην σχέση (3.2) και

$$F = \left\{ (\alpha_1, \alpha_2, \dots, \alpha_n) : \sum_{i=1}^n \cos \alpha_i = c, \sum_{i=1}^n \sin \alpha_i = s \right\}.$$

• Η κατανομή του R

Έστω $\beta = \bar{\alpha}_0 - \mu$, δηλαδή η γωνία που σχηματίζει το συνιστάμενο διάνυσμα με την μέση κατεύθυνση του δείγματος μας μ .

Από την εργασία των Kent et al. (1979) γνωρίζουμε ότι για δεδομένα υπό την ομοιόμορφη κατανομή η γωνία β ακολουθεί και αυτή την ομοιόμορφη κατανομή και μάλιστα είναι ανεξάρτητη του συνιστάμενου διανύσματος R . Αν θέσουμε ως c το συνημίτονο της εν λόγω γωνίας ($c = \cos \beta$) τότε έχουμε:

$$f_0(c) = \frac{1}{\pi\sqrt{1-c^2}}, \quad -1 \leq c \leq 1. \quad (3.7)$$

Όπως αναφέραμε παραπάνω οι ποσότητες R, β είναι ανεξάρτητες μεταξύ τους και συνεπώς ισχύει:

$$f_0(r, c) = f_0(r)f_0(c) = \frac{r\psi_n(r)}{\pi\sqrt{1-c^2}}. \quad (3.8)$$

Βάσει των σχέσεων (3.2) και (3.6) προκύπτει για δείγμα υπό την κυκλική κανονική κατανομή:

$$f_\kappa(r, c) = \frac{e^{\kappa r c}}{I_0^n(\kappa)} \cdot \frac{r\psi_n(r)}{\pi\sqrt{1-c^2}}. \quad (3.9)$$

Από αυτή την από κοινού συνάρτηση πυκνότητας πιθανότητας μπορούμε να ολοκληρώσουμε ως προς το c και να λάβουμε την παρακάτω επιθυμητή συνάρτηση πυκνότητας πιθανότητας του R για δεδομένα υπό την VonMises κατανομή:

$$f_\kappa(r) = \int_{-1}^1 f_\kappa(r, c) dc = \left(\frac{I_0(\kappa r)}{I_0^n(\kappa)} \right) r\psi_n(r). \quad (3.10)$$

• Η κατανομή του V

Αντίστοιχα η κατανομή του V για δεδομένα υπό την κυκλική κανονική κατανομή προκύπτει:

$$f_\kappa(v) = \frac{e^{\kappa v}}{I_0^n} f_0(v) = \frac{e^{\kappa v}}{\pi I_0^n} \int_0^\infty J_0^n(t) \cos(vt) dt. \quad (3.11)$$

3.3 Προσεγγίσεις σε Δείγματα μεγάλου μεγέθους

3.3.1 Προσεγγίσεις με σειρές

Έστω τυχαία μεταβλητή $A \sim U(0, 2\pi)$ και έστω $(\alpha_1, \alpha_2, \dots, \alpha_n)$ τυχαίο δείγμα από την εν λόγω κατανομή.

Υποθέτοντας για το συγκεκριμένο δείγμα ότι ισχύει $c = \cos \alpha_i, s = \sin \alpha_i \quad \forall i \in [1, n]$, έχουμε:

$$\left\{ \begin{array}{l} E_A(c) = \int_0^{2\pi} \cos \alpha_i \frac{1}{2\pi} d\alpha_i = 0 \\ 0 \\ E_A(s) = \int_0^{2\pi} \sin \alpha_i \frac{1}{2\pi} d\alpha_i = 0 \\ 0 \end{array} \right. , \quad \left\{ \begin{array}{l} E_A(c^2) = \frac{1}{2} = E(s^2) \\ E_A(c \cdot s) = 0 \end{array} \right. .$$

Αφού ο πίνακας ροπών δευτέρας τάξης είναι πεπερασμένος ικανοποιούνται οι προϋποθέσεις του δισδιάστατου κεντρικού οριακού θεωρήματος και συνεπώς:

$$(\bar{C}, \bar{S}) \xrightarrow{d} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ με } \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{2n} & 0 \\ 0 & \frac{1}{2n} \end{pmatrix} \text{ για μεγάλο } n,$$

όπου $(\bar{C}, \bar{S}) \equiv \left(\frac{1}{n} \sum_{j=1}^n \cos \alpha_j, \frac{1}{n} \sum_{j=1}^n \sin \alpha_j \right)$ ο μέσος όρος των ποσοτήτων που ορίστηκαν στην σχέση (1.4).

Αξιοσημείωτο είναι, ότι ενώ τα (c, s) είναι εξαρτημένα μεταξύ τους, τα (\bar{C}, \bar{S}) γίνονται ασυμπτωματικά ανεξάρτητα. Αφού λοιπόν δείξαμε ότι τα (\bar{C}, \bar{S}) συγκλίνουν κατά κατανομή στην παραπάνω κανονική, τότε τα $(\sqrt{2n}\bar{C})^2, (\sqrt{2n}\bar{S})^2$ συγκλίνουν κατά κατανομή στην X_1^2 και συνεπώς το άθροισμά τους για μεγάλες τιμές του n

$$Z \equiv \frac{2R^2}{n} = 2n(\bar{C}^2 + \bar{S}^2) \sim X_2^2.$$

Χρησιμοποιώντας το παραπάνω αποτέλεσμα μπορούμε να γράψουμε την προσεγγιστική συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής R για

μεγάλα δείγματα ως $f(r) = g(z) \left| \frac{dz}{dr} \right| = \frac{1}{2} e^{-\frac{r^2}{n}} \frac{4r}{n} = \frac{2r}{n} e^{-\frac{r^2}{n}}$, με αντίστοιχη συνάρτηση

κατανομής την $F_n(r) = 1 - e^{-\frac{r^2}{n}}, 0 \leq r \leq n$. Μια αρκετά πιο ακριβής αναπαράσταση της είναι το ανάπτυγμα σε σειρά Laguerre ως εξής:

$$F_n(r) = 1 - e^{-z} \left[1 + \frac{1}{2n} \left(z - \frac{z^2}{2!} \right) + \frac{1}{12n^2} \left(-z + \frac{11z^2}{2!} - \frac{19z^3}{3!} + 9 \frac{z^4}{4!} \right) + \frac{1}{24n^3} \left(-2z - 4 \frac{z^2}{2!} + 69 \frac{z^3}{3!} - 163 \frac{z^4}{4!} + 145 \frac{z^5}{5!} - 45 \frac{z^6}{6!} \right) \right],$$

$$\text{όπου } z = \frac{r^2}{n}.$$

Στην περίπτωση που η διασπορά υπολογίζεται γύρω από συγκεκριμένη υποθετική μέση κατεύθυνση μ , μπορούμε χωρίς βλάβη της γενικότητας να την υποθέσουμε ως μηδενική και συνεπώς λαμβάνουμε από την σχέση (3.3) ότι $V = C$. Τότε η ασυμπτωματική κατανομή της τυχαίας μεταβλητής V υπό την ομοιόμορφη κυκλική κατανομή είναι η $N(0, \frac{n}{2})$.

Μια προσέγγιση της με σειρές βασισμένες σε ερμιτιανά πολυώνυμα είναι:

$$F_n(v) = P(V \leq v) = \Phi(z) + \frac{(z^2 - 3z)}{16n} \varphi(z) - \frac{(15z + 305z^3 - 125z^5 + 9z^7)}{4608n^2} \varphi(z), \quad \text{όπου}$$

$z = \sqrt{\frac{2}{n}}v$ και $\varphi(\cdot), \Phi(\cdot)$ οι συναρτήσεις πυκνότητας πιθανότητας και κατανομής της τυποποιημένης κανονικής κατανομής αντίστοιχα.

3.3.2 Προσεγγίσεις κεντρικού οριακού θεωρήματος

Έστω δείγμα $(\alpha_1, \alpha_2, \dots, \alpha_n)$ ανεξάρτητων και ισόνομων παρατηρήσεων από μια κυκλική κατανομή με συνάρτηση κατανομής $G(\theta)$. Θα δείξουμε ότι η κατανομή του αθροίσματος $S_n = (\alpha_1 + \alpha_2 + \dots + \alpha_n) \pmod{2\pi}$ συγκλίνει στην ομοιόμορφη κυκλική κατανομή όταν $n \rightarrow \infty$.

Ας υποθέσουμε ότι χαρακτηριστική συνάρτηση που αντιστοιχεί στην $G(\theta)$ είναι η $\{\varphi_\theta(p)\}$, για την οποία έχουμε αποδείξει ότι ισχύει $\varphi_\theta(0) = 1, |\varphi_\theta(p)| < 1$ για $p \neq 0$. Έτσι, για το άθροισμα S_n έχουμε $\varphi_{S_n}(0) = 1, \varphi_{S_n}(p) = \varphi_\theta^n(p) \rightarrow 0$ για $p \neq 0$, που όπως ξέρουμε είναι η χαρακτηριστική συνάρτηση της ομοιόμορφης κατανομής.

Με μια διαφορετική προσέγγιση του αθροίσματος βέβαια μπορούμε να καταλήξουμε σε σύγκλιση στην περιελιγμένη κανονική. Για παράδειγμα, έστω α_i που ανήκουν στο διάστημα $(-\pi, \pi)$ με $E(\alpha_i) = 0, E(\alpha_i^2) = \sigma^2$. Τότε $n^{-\frac{1}{2}} \sum_{i=1}^n \alpha_i \xrightarrow{d} N(0, \sigma^2)$ και συνεπώς

$$S_n^* = \frac{\sum_{i=1}^n \alpha_i}{\sqrt{n}} \pmod{2\pi} \xrightarrow{d} WN(0, \sigma^2).$$

3.3.3 Προσεγγίσεις με την μέθοδο των ροπών

Δεδομένου ενός δείγματος $(\alpha_1, \alpha_2, \dots, \alpha_n)$, όπως έχουμε παρατηρήσει, οι περισσότερες στατιστικά συναρτήσεις που αντιμετωπίζουμε εκφράζονται από τις ποσότητες (C, S) ή (\bar{C}, \bar{S}) .

Επειδή όλες οι ροπές είναι φραγμένες μπορούμε να εφαρμόσουμε το Κεντρικό Οριακό Θεώρημα στην από κοινού κατανομή των (\bar{C}, \bar{S}) .

Συγκεκριμένα έστω $(\theta_1, \theta_2, \dots, \theta_n)$ ανεξάρτητες και ισόνομες παρατηρήσεις στο $[0, 2\pi)$. Σύμφωνα με τις σχέσεις (2.1), (2.2) και (2.3) λαμβάνουμε τις ροπές πρώτης και δεύτερης τάξης αντίστοιχα:

$$E_{\Theta}(e^{i\theta}) = \rho e^{i\mu} = (\alpha_1 + i\beta_1) \quad E_{\Theta}(e^{i2\theta}) = \rho_2 e^{i\mu_2} = (\alpha_2 + i\beta_2).$$

Το πολυδιάστατο κεντρικό οριακό θεώρημα εφαρμοσμένο στα ανεξάρτητα και ισόνομα μοναδιαία διανύσματα $(\cos \theta_j, \sin \theta_j)$, $j = 1, \dots, n$ ως ακόλουθα:

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^n (\cos \theta_j - E(\cos \theta_1)) \\ \sum_{j=1}^n (\sin \theta_j - E(\sin \theta_1)) \end{pmatrix} \sim N(0, \Sigma),$$

όπου ο πίνακας διακύμανσης- συνδιακύμανσης δίνεται ως εξής:

$$\Sigma = \begin{pmatrix} \sigma_{cc} & \sigma_{cs} \\ \sigma_{sc} & \sigma_{ss} \end{pmatrix}.$$

Χρησιμοποιώντας τον ορισμό των ροπών πρώτης τάξης και τις τριγωνομετρικές ιδιότητες, $\cos 2\theta = 2\cos^2 \theta - 1 = 1 - 2\sin^2 \theta$ και $\sin 2\theta = 2\sin \theta \cos \theta$ έχουμε:

$$\begin{cases} \sigma_{cc} = \frac{1}{2} + \frac{1}{2} \rho_2 \cos \mu_2 - \rho^2 \cos^2 \mu = \frac{1}{2} (1 + \alpha_2 - 2\alpha_1^2) \\ \sigma_{cs} = \sigma_{sc} = \frac{1}{2} \rho_2 \sin \mu_2 - \rho^2 \cos \mu \sin \mu = \frac{1}{2} (\beta_2 - 2\alpha_1 \beta_1) \\ \sigma_{ss} = \frac{1}{2} - \frac{1}{2} \rho_2 \cos \mu_2 - \rho^2 \sin^2 \mu = \frac{1}{2} (1 - \alpha_2 - 2\beta_1^2) \end{cases}$$

Για να προσδιορίσουμε την ασυμπτωτική κατανομή των στατιστικών ποσοτήτων που μας ενδιαφέρουν θα χρησιμοποιήσουμε την λεγόμενη δ-μέθοδο την οποία θα ορίσουμε παρακάτω μέσω του βασικού λήμματος της.

Λήμμα δ-μεθόδου

Έστω $(\sqrt{n}(T_{1n} - \theta_1), \dots, \sqrt{n}(T_{kn} - \theta_k))$ τυχαίο διάνυσμα που ακολουθεί ασυμπτωτικά την κανονική κατανομή k διάστασης με μέση τιμή μηδέν και πίνακα διακύμανσης $\Sigma = (\sigma_{ij})$ όπου $\sigma_{ij} = Cov(T_i, T_j)$ $i = 1, \dots, k$, $j = 1, \dots, k$. Έστω επίσης συνάρτηση g, k μεταβλητών πλήρως διαφορίσιμη. Τότε το τυχαίο διάνυσμα:

$$\left(\sqrt{n} \left[g(T_{1n}, \dots, T_{kn}) - g(\theta_1, \dots, \theta_k) \right]\right)$$

ακολουθεί ασυμπτωτικά την κανονική κατανομή με μέση τιμή μηδέν και διακύμανση:

$$\sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} \frac{dg}{d\theta_i} \frac{dg}{d\theta_j}.$$

Ένα παράδειγμα της μεθόδου αυτής μας δίνει το ακόλουθο αποτέλεσμα για την δειγματική μέση κατεύθυνση $\bar{\alpha}_0$:

Χρησιμοποιώντας ως στατιστική συνάρτηση την εφαπτομένη της $\bar{\alpha}_0$ έχουμε:

$$\sqrt{n} (\tan \bar{\alpha}_0 - \tan \mu) \sim N(0, \sigma^2),$$

$$\begin{aligned} \sigma^2 &= \frac{1}{\alpha_1^2} \left\{ \sigma_{cc} \left(\frac{\beta_1}{\alpha_1} \right)^2 + \sigma_{ss} - 2\sigma_{cs} \frac{\beta_1}{\alpha_1} \right\} = \\ \text{όπου} \quad &= \frac{1 - \rho_2 (\cos \mu_2 \cos 2\mu + \sin \mu_2 \sin 2\mu)}{2\rho^2 \cos^4 \mu}. \end{aligned}$$

3.4 Πολλαπλά δείγματα

Οι παραπάνω κατανομές που αναλύσαμε αντιστοιχούν σε δεδομένα προερχόμενα από ένα μοναδικό δείγμα. Σε αυτή την παράγραφο θα εξετάσουμε την περίπτωση που έχουμε παραπάνω από έναν πληθυσμούς για να επιλέξουμε το δείγμα μας.

Πιο συγκεκριμένα, έστω q ($q \geq 2$) πληθυσμοί υπό την κυκλική κανονική κατανομή με ίδια παράμετρο συγκέντρωσης, κ και διαφορετικές πολικές κατευθύνσεις $\mu_1, \mu_2, \dots, \mu_q$.

Αν συμβολίσουμε $\{\alpha_{i,j} : j = 1, \dots, n_i, i = 1, \dots, q\}$ τα δεδομένα από τους πληθυσμούς, τότε επαγωγικά ορίζουμε ως R_i το μήκος του συνιστάμενου διανύσματος για τον i -οστό πληθυσμό, βασισμένο σε δείγμα μεγέθους n_i και ως R το μήκος του συνολικού συνιστάμενου διανύσματος βάσει του συνολικού δείγματος μεγέθους $N = \sum_{i=1}^q n_i$.

Σκοπός μας είναι να αποδείξουμε ότι η δεσμευμένη κατανομή $f_{\kappa}(R_1, R_2, \dots, R_q | R)$ των ανεξάρτητων δειγματικών συνισταμένων, δεδομένου του μήκους του συνολικού συνιστάμενου διανύσματος είναι ανεξάρτητη της παραμέτρου συγκέντρωσης. Αυτό είναι ένα πολύ χρήσιμο συμπέρασμα, καθώς θα μας επιτρέψει να κατασκευάσουμε ακριβείς ελέγχους σημαντικότητας για τις πολικές κατευθύνσεις των κυκλικών κανονικών κατανομών, καθώς και εν γένει να μας δώσει πληροφορίες για τις μέσες

κατευθύνσεις των δειγμάτων, ακόμα και όταν η παράμετρος συγκέντρωσης είναι άγνωστη.

Ας ξεκινήσουμε από τη ομοιόμορφη κυκλική κατανομή, επισημαίνοντας ότι στην περίπτωση μας το συνολικό συνιστάμενο διάνυσμα συντίθεται από τα επί μέρους δειγματικά συνιστάμενα διανύσματα με μήκη R_1, R_2, \dots, R_q και τυχαίες γωνίες μεταξύ τους. Με αφετηρία αυτό μπορούμε να χρησιμοποιήσουμε το συμπέρασμα του γενικευμένου τυχαίου περιπάτου και να λάβουμε το εξής:

$$f_0(R_1, R_2, \dots, R_q | R) = R \int_0^\infty J_0(Rt) \left[\prod_{i=1}^q J_0(R_i t) \right] t dt. \quad (3.12)$$

Υποθέτοντας την ύπαρξη κοινής μέσης κατεύθυνσης μ (σε περίπτωση στατιστικού ελέγχου αυτή θα αποτελεί και την μηδενική υπόθεση) και υιοθετώντας τον συμβολισμό c ως το συνημίτονο της γωνίας του συνολικού συνιστάμενου διανύσματος με την κατεύθυνση μ , μπορούμε να συνδυάσουμε τα ήδη γνωστά αποτελέσματα $f_0(c)$, $f_0(R_i)$, $f_\kappa(R)$ και βάσει της ανεξαρτησίας τους υπό την ομοιόμορφη κατανομή να λάβουμε την εξής από κοινού συνάρτηση πυκνότητας πιθανότητας:

$$\begin{aligned} f_0(R_1, R_2, \dots, R_q, R, c) &= f_0(R | R_1, R_2, \dots, R_q) \prod_{i=1}^q f_0(c) f_0(R_i) = \\ &= \frac{1}{\pi \sqrt{1-c^2}} R \int_0^\infty J_0(Rt) \left[\prod_{i=1}^q J_0(R_i t) \right] t dt \prod_{i=1}^q R_i \psi_{n_i}(R_i). \end{aligned} \quad (3.13)$$

Τότε για την κανονική κατανομή:

$$\begin{aligned} f_\kappa(R_1, R_2, \dots, R_q, R) &= \int_{-1}^1 f_\kappa(R_1, R_2, \dots, R_q, R, c) dc = \\ &= \int_{-1}^1 \frac{e^{\kappa R c}}{I_0^N(\kappa)} f_0(R_1, R_2, \dots, R_q, R, c) dc = \frac{I_0(\kappa R)}{I_0^N(\kappa)} R \int_0^\infty J_0(Rt) \left[\prod_{i=1}^q J_0(R_i t) \right] t dt. \end{aligned} \quad (3.14)$$

Προκύπτει λοιπόν η επιθυμητή δεσμευμένη κατανομή:

$$\begin{aligned} f_\kappa(R_1, \dots, R_q | R) &= \frac{f_\kappa(R_1, \dots, R_q, R)}{f_\kappa(R)} = \\ &= \frac{1}{\psi_N(R)} \int_0^\infty J_0(Rt) \prod_{i=1}^q J_0(R_i t) t dt \prod_{i=1}^q R_i \psi_{n_i}(R_i), \end{aligned} \quad (3.15)$$

που είναι όντως ανεξάρτητη του συντελεστή συγκέντρωσης.

3.5 Προσεγγιστικές κατανομές για μεγάλο κ

Όπως έχουμε αποδείξει μέσω της σχέσης (2.13), η κυκλική κανονική κατανομή για μεγάλες τιμές της παραμέτρου συγκέντρωσης προσεγγίζεται ικανοποιητικά από την «γραμμική» κανονική κατανομή. Συγκεκριμένα, εάν $a \sim CN(\mu, \kappa)$ τότε

$$\beta = \sqrt{\kappa}(a - \mu) = \sqrt{\kappa}\theta \xrightarrow{d} N(0,1) \text{ ή διαφορετικά } a \sim N\left(\mu, \frac{1}{\kappa}\right) \text{ για μεγάλο } \kappa.$$

Θα χρησιμοποιήσουμε αυτή την προσέγγιση για να δείξουμε ότι για κ ικανοποιητικά μεγάλο, το σύνολο δεδομένων εμφανίζει συγκέντρωση σε ένα τόξο αρκετά μικρού μήκους κοντά στη πολική κατεύθυνση.

Έστω $(\alpha_1, \alpha_2, \dots, \alpha_n)$ ανεξάρτητες και ισόνομες παρατηρήσεις από την $CN(\mu, \kappa)$.

Τότε $\theta_i = (\alpha_i - \mu) \sim N\left(0, \frac{1}{\kappa}\right) \forall i \in [1, n]$ για μεγάλες τιμές του κ .

Γνωρίζουμε ότι για μικρά θ_i ισχύει η προσέγγιση $\cos \theta_i \approx \left(1 - \frac{\theta_i^2}{2}\right)$ και κατά συνέπεια από την θεωρητική προσέγγιση της κανονικής κατανομής λαμβάνουμε ότι: $(\sqrt{\kappa}\theta_i)^2 = \kappa\theta_i^2 \approx 2\kappa(1 - \cos \theta_i) \sim X_1^2$. Αθροίζοντας για όλα τα i παίρνουμε την σχέση $2\kappa(n - V) \doteq X_n^2$, όπου V η δειγματική διασπορά όπως ορίστηκε στην σχέση (3.3).

Βάσει υπόθεσης τα θ_i είναι ανεξάρτητα και ισχύει η προσέγγιση $\theta_i \approx \sin \theta_i$. Συνεπώς

για μεγάλες τιμές της παραμέτρου κ έχουμε: $\frac{\kappa}{n} \left(\sum_{i=1}^n \sin \theta_i \right)^2 \sim X_1^2$.

Όπως αποδείξαμε στο πρώτο κεφάλαιο το συνιστάμενο διάνυσμα R είναι αναλλοίωτο σε στροφές, δηλαδή λαμβάνει τις ίδιες τιμές για τις παρατηρήσεις α_i και για τις παρατηρήσεις $\theta_i = \alpha_i - \mu$.

Από την σχέση (1.5) υψώνοντας στο τετράγωνο λαμβάνουμε: $R^2 = \left(\sum_{i=1}^n \sin \theta_i \right)^2 + \left(\sum_{i=1}^n \cos \theta_i \right)^2$. Έτσι $\frac{\kappa}{n} \left(\sum_{i=1}^n \sin \theta_i \right)^2 = \frac{\kappa}{n} (R^2 - V^2) \approx \boxed{2\kappa(R - V) \sim X_1^2}$.

Η τελευταία ισότητα προκύπτει από το γεγονός ότι για αρκετά μεγάλες τιμές της παραμέτρου συγκέντρωσης, ισχύουν οι προσεγγίσεις $R \approx V \approx n$ και έτσι $R^2 - V^2 = (R - V)(R + V) \approx 2n(R - V)$.

Σύμφωνα με την παραπάνω ανάλυση η τυχαία μεταβλητή $2\kappa(n - V)$ που ακολουθεί την κατανομή X_n^2 μπορεί να γραφεί στην μορφή $2\kappa(n - R) + 2\kappa(R - V)$, δηλαδή ως άθροισμα δύο τυχαίων μεταβλητών που ακολουθούν την ίδια κατανομή αλλά με $n - 1$ και 1 βαθμό ελευθερίας αντίστοιχα.

Κεφάλαιο 4

Εκτίμηση παραμέτρων

4.1 Εισαγωγή

Στα προηγούμενα κεφάλαια αναλύσαμε τις βασικές κατανομές που χρησιμοποιούνται στην στατιστική ανάλυση των κυκλικών μεταβλητών, καθώς και τις μεθόδους εύρεσης των κατανομών των στατιστικών συναρτήσεων που συνήθως χρησιμοποιούμε. Όπως και στη ανάλυση των γραμμικών μεταβλητών, έτσι και σε αυτή την περίπτωση, το επόμενο και ιδιαίτερα σημαντικό βήμα είναι η εύρεση μεθόδων για την ορθή εκτίμηση των άγνωστων παραμέτρων της εκάστοτε κατανομής που ακολουθεί το δείγμα μας.

Ιδιαίτερη σημασία θα δοθεί στην κυκλική κανονική κατανομή για την οποία οι μέθοδοι που θα παρουσιάσουμε δίνουν ταυτόσημα αποτελέσματα, σε αντίθεση με τις οικογένειες των περιελιγμένων σταθερών κατανομών όπου συγκεκριμένες μέθοδοι οδηγούν σε δυσμεταχειρίστες επαναληπτικές μεθόδους.

4.2 Κυκλική κανονική

4.2.1 Εκτιμήτριες μέγιστης πιθανοφάνειας

Έστω $(\alpha_1, \alpha_2, \dots, \alpha_n)$ ένα σύνολο δεδομένων από την κυκλική κανονική κατανομή με παραμέτρους μ, κ . Η συνάρτηση πυκνότητας πιθανότητας των α_i είναι:

$$f(\alpha) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\alpha - \mu)], \quad 0 \leq \alpha < 2\pi. \quad (4.1)$$

Η συνάρτηση πιθανοφάνειας τότε είναι:

$$L(\mu, \kappa | \alpha_1, \alpha_2, \dots, \alpha_n) = \frac{1}{[2\pi I_0(\kappa)]^n} \exp \left[\sum_{i=1}^n \kappa \cos(\alpha_i - \mu) \right]. \quad (4.2)$$

Λογαριθμίζοντας παίρνουμε:

$$l = \log L = -n \log(2\pi I_0(\kappa)) + \kappa \sum_{i=1}^n \cos(\alpha_i - \mu). \quad (4.3)$$

Παραγωγίζοντας μερικώς ως προς μ και κ και εξισώνοντας με το μηδέν παίρνουμε το παρακάτω σύστημα εξισώσεων πιθανοφάνειας:

$$\begin{cases} \frac{\partial l}{\partial \mu} = \sum_{i=1}^n \sin(\alpha_i - \mu) = 0 \\ \frac{\partial l}{\partial \kappa} = -n \frac{I_1(\kappa)}{I_0(\kappa)} + \sum_{i=1}^n \cos(\alpha_i - \mu) = 0 \end{cases}. \quad (4.4)$$

Από την πρώτη εξίσωση λαμβάνουμε:

$$\begin{aligned} \sum_{i=1}^n (\sin \alpha_i \cos \mu - \cos \alpha_i \sin \mu) = 0 &\Rightarrow S \cos \mu - C \sin \mu = 0 \Rightarrow \\ &\Rightarrow \hat{\mu} = \arctan^* \left(\frac{S}{C} \right) = \overline{\alpha_0}. \end{aligned} \quad (4.5)$$

Χρησιμοποιώντας την παραπάνω εκτίμηση του μ στην δεύτερη εξίσωση του συστήματος και χρησιμοποιώντας το πόρισμα του πρώτου κεφαλαίου

$$\sum_{i=1}^n \cos(\alpha_i - \overline{\alpha_0}) = R,$$

παίρνουμε την εξής εκτίμηση για την παράμετρο συγκέντρωσης:

$$-n \frac{I_1(\kappa)}{I_0(\kappa)} + R = 0 \Rightarrow \frac{I_1(\kappa)}{I_0(\kappa)} = \frac{R}{n} \equiv A(\kappa). \quad (4.6)$$

Ο λόγος των συναρτήσεων Bessel που ορίσαμε ως $A(\kappa)$ έχει μια σειρά από ιδιότητες που μας δίνουν χρήσιμα συμπεράσματα για την εκτίμηση του κ .

- $\boxed{0 \leq A(\kappa) \leq 1}$
- $\begin{cases} A(\kappa) \rightarrow 0 \text{ για } \kappa \rightarrow 0 \\ A(\kappa) \rightarrow 1 \text{ για } \kappa \rightarrow \infty \end{cases}$
- $A'(\kappa) \equiv \frac{\partial A(\kappa)}{\partial \kappa} = \left(1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa) \right) \geq 0.$

Από τα παραπάνω βλέπουμε φυσικά ότι ο παραπάνω λόγος είναι μια συνάρτηση της παραμέτρου συγκέντρωσης κ , η οποία είναι γνησίως αύξουσα και συνεπώς η εκτίμηση του κ είναι μοναδική. Επίσης προκύπτει ότι ο Εσσιανός πίνακας (δηλαδή ο πίνακας δευτέρων παραγώγων του $\log L$) εκτιμώμενος στο $(\hat{\mu}, \hat{\kappa})$ είναι ο διαγώνιος πίνακας $Diag(-\hat{\kappa}R, -nA'(\hat{\kappa}))$, ο οποίος είναι αρνητικά ορισμένος. Αυτό σημαίνει ότι τα $(\hat{\mu}, \hat{\kappa})$ είναι όντως οι εκτιμήτριες μέγιστης πιθανοφάνειας των (μ, κ) αντίστοιχα.

4.2.2 Εκτιμήτριες με την μέθοδο των ροπών

Στη μέθοδο των ροπών εξισώνουμε τις θεωρητικές τριγωνομετρικές ροπές, όπως ορίζονται στην εξίσωση (2.1) με τις δειγματικές τριγωνομετρικές ροπές, όπως ορίζονται στην ακόλουθη εξίσωση: $\bar{C}_n(p) + i\bar{S}_n(p) = \frac{1}{n} \sum_{j=1}^n (e^{i\alpha_j})^p$, $p = 0, 1, 2$.

Στην περίπτωση μας εξισώνουμε τις $\begin{cases} E(\cos \alpha) = A(\kappa) \cos \mu \\ E(\sin \alpha) = A(\kappa) \sin \mu \end{cases}$ με τις δειγματικές ροπές

και προκύπτουν οι ίδιες εξισώσεις με αυτές τις μεθόδου μέγιστης πιθανοφάνειας. Έτσι οι εκτιμήτριες των δύο μεθόδων συμπίπτουν στην περίπτωση του δείγματος από την κυκλική κανονική κατανομή.

4.3 Μίξεις κυκλικής κατανομής

Οι πεπερασμένες μίξεις $CN(\mu_i, \kappa_i)$ με παράμετρο μίξης $p_i, i = 1, \dots, k$ μας δίνουν κατανομές με συνάρτηση πυκνότητας πιθανότητας

$$g(\alpha) = \sum_{i=1}^k p_i \frac{\exp[k_i \cos(\alpha - \mu_i)]}{2\pi I_0(k_i)}, \quad 0 \leq \alpha < 2\pi. \quad (4.7)$$

Εφόσον αυτές οι μίξεις προκύπτουν, οι εκτιμήτριες μέγιστης πιθανοφάνειας μπορούν να βρεθούν από τον κλασικό αλγόριθμο με τη μόνη διαφορά ότι εάν οι παράμετροι συγκέντρωσης λαμβάνουν παρόμοιες και πολύ κοντινές τιμές προτιμάται η μέθοδος των ροπών. Για παράδειγμα, στην απλή περίπτωση αξονικών δεδομένων από δύο κυκλικές κανονικές κατανομές με ίσες παραμέτρους συγκέντρωσης η μέθοδος των ροπών δίνει αρκετά απλούστερα αποτελέσματα και παράλληλα συγκρίσιμα με αυτά της μεθόδου μέγιστης πιθανοφάνειας.

Μια ακόμη ιδιαίζουσα περίπτωση μίξης των VonMises κατανομών είναι αυτή των συμμετρικών γύρω από μηδενική μέση κατεύθυνση. Συγκεκριμένα, ας λάβουμε την ακόλουθη μίξη με $k = 2$, $\rho_1 = \rho_2 = \frac{1}{2}$, $\mu_1 = -\mu_2 = \mu$, $\kappa_1 = \kappa_2 = \kappa$, και ας προχωρήσουμε

σε εκτίμηση μέγιστης πιθανοφάνειας. Έστω δείγμα $\alpha_1, \alpha_2, \dots, \alpha_n$ με πιθανοφάνεια:

$$L(\mu, \kappa) = C - n \log \{I_0(\kappa)\} + \sum_{i=1}^n \log [\exp[\kappa \cos(\alpha_i - \mu)] + \exp[\kappa \cos(\alpha_i + \mu)]] \quad (4.8)$$

Το σημαντικό σε αυτή την συνάρτηση, ανεξαρτήτως των τιμών του δείγματος, είναι όταν $\mu = 0$ ισχύει $\frac{\partial L}{\partial \mu} = 0$ ανεξαρτήτως των τιμών του δείγματος. Αυτό με την σειρά

του σημαίνει ότι όταν $\mu = 0$ οι ποσότητες $E \left\{ \left(\frac{\partial L}{\partial \mu} \right)^2 \right\}$ και $E \left\{ \left(\frac{\partial L}{\partial \mu} \right) \left(\frac{\partial L}{\partial \kappa} \right) \right\}$ είναι

μηδενικές. Λόγω των παραπάνω, οι προϋποθέσεις που τέθηκαν από τους Kendall και Stuart (1979) για την ορθότητα της εκτίμησης της παραμέτρου μ δεν επηρεάζονται αλλά παύουν να ισχύουν αυτές τις ασυμπτωτικής αποτελεσματικότητας και κανονικότητας. Αυτό δεν καθιστά κατ'ανάγκη λανθασμένη την εκτίμηση αλλά μας αποδεικνύει ότι δεν μπορούμε να υπολογίσουμε για την παράμετρο μας τυπικό σφάλμα και διάστημα εμπιστοσύνης βάσει της ασυμπτωτικής συμπεριφοράς της, όταν η τιμή της είναι κοντά στο μηδέν.

Μια άλλη σημαντική συνέπεια του μηδενισμού της μερικής παραγώγου στο σημείο μηδέν είναι ότι στο σημείο αυτό η συνάρτηση L θα έχει πάντα τοπικό μέγιστο ή ελάχιστο. Συνακόλουθα πρέπει στην αριθμητική εκτίμηση της συνάρτησης να δοθεί ιδιαίτερη προσοχή στον χαρακτηρισμό αυτού του σημείου ακρότατου σε συνολικό μέγιστο. Η πιθανότητα να ταυτίζονται αυξάνει όσο η πραγματική τιμή του μ πλησιάζει το μηδέν. Αυτό έχει ως αποτέλεσμα η δειγματική κατανομή της εκτιμήτριας να εμφανίζει στο σημείο $\hat{\mu} = 0$ ένα μέγιστο, με την υπόλοιπη πιθανότητα να κατανέμεται στις υπόλοιπες θετικές εκτιμήσεις του μ .

4.4 Εκτίμηση μέγιστης πιθανοφάνειας της περιελιγμένης Cauchy

Σε αυτή την παράγραφο θα δούμε πως με έναν επαναληπτικό αλγόριθμο μπορούμε να εκτιμήσουμε με τη μέθοδο μέγιστης πιθανοφάνειας την παράμετρο μ της περιελιγμένης Cauchy κατανομής.

Αρχικά αναπαραμετροποιούμε την συνάρτηση πυκνότητας πιθανότητας ως εξής:

$$\mu_1 = \frac{2\rho \cos \mu}{(1 + \rho^2)} \quad \mu_2 = \frac{2\rho \sin \mu}{(1 + \rho^2)} \quad c = c(\mu_1, \mu_2) = \frac{1}{\sqrt{1 - \mu_1^2 - \mu_2^2}}$$

Έρχεται λοιπόν σε αυτή την μορφή: $f(\theta; \mu_1, \mu_2) = [2\pi c(1 - \mu_1 \cos \theta - \mu_2 \sin \theta)]^{-1}$.

Ας υποθέσουμε ότι διαθέτουμε ένα τυχαίο δείγμα $(\theta_1, \theta_2, \dots, \theta_n)$ από την WC κατανομή. Για να εκτελέσουμε την μέθοδο μέγιστης πιθανοφάνειας κάνουμε μια ακόμη αναπαραμετροποίηση ως εξής:

$$\begin{cases} n_1 = c\mu_1 \\ n_2 = c\mu_2 \end{cases} \Rightarrow c = \sqrt{1 + n_1^2 + n_2^2}, w_i = (1 - \mu_1 \cos \theta_i - \mu_2 \sin \theta_i)^{-1}.$$

Παραγωγίζοντας ως προς n_1, n_2 αντίστοιχα την συνάρτηση πιθανοφάνειας της συνάρτησης πυκνότητας πιθανότητας προκύπτει το ακόλουθο σύστημα:

$$\begin{cases} \frac{1}{c} \sum_{i=1}^n w_i [\cos \theta_i - \mu_1] = 0 \\ \frac{1}{c} \sum_{i=1}^n w_i [\sin \theta_i - \mu_2] = 0 \end{cases} \Rightarrow \begin{cases} \mu_1 = \frac{\left\{ \sum_{i=1}^n w_i \cos \theta_i \right\}}{\left\{ \sum_{i=1}^n w_i \right\}} \\ \mu_2 = \frac{\left\{ \sum_{i=1}^n w_i \sin \theta_i \right\}}{\left\{ \sum_{i=1}^n w_i \right\}} \end{cases}.$$

Για την επίλυση του συστήματος αυτού λοιπόν εκτελούμε τον ακόλουθο αλγόριθμο:

- Ξεκινάμε με δύο αυθαίρετες αρχικές τιμές $\mu_{1,0}^2 + \mu_{2,0}^2$ υπό τον περιορισμό $\mu_{1,0}^2 + \mu_{2,0}^2$.
- Βάσει των τιμών $\mu_{1,v}, \mu_{2,v}$ στην v -οστή επανάληψη επανα-υπολογίζουμε τις τιμές βάσει του παρακάτω:

$$\begin{cases} \mu_{1,v+1} = \frac{\left\{ \sum_{i=1}^n w_{i,v} \cos \theta_i \right\}}{\left\{ \sum_{i=1}^n w_{i,v} \right\}} \\ \mu_{2,v+1} = \frac{\left\{ \sum_{i=1}^n w_{i,v} \sin \theta_i \right\}}{\left\{ \sum_{i=1}^n w_{i,v} \right\}} \end{cases}$$

όπου $w_{i,v} = (1 - \mu_{1,v} \cos \theta_i - \mu_{2,v} \sin \theta_i)^{-1}, \forall i = 1, 2, \dots, n$.

- Επαναλαμβάνουμε το παραπάνω βήμα μέχρι ο αλγόριθμος να συγκλίνει.

Η ύπαρξη και η μοναδικότητα των εκτιμητριών μέγιστης πιθανοφάνειας και η σύγκλιση του αλγορίθμου επιτυγχάνεται μέσω της σχέσης της κατανομής WC με το γωνιακή κεντρική Gaussian κατανομή (angular central Gaussian distribution) και γνωστές αποδείξεις της ανάλυσης της όπως παρουσιάζονται στη δημοσίευση του Tyler D.E. (1987), Statistical analysis for the angular central Gaussian distributions on the sphere, Biometrika, 74, p. 579–589.

Πειραματικά έχει αποδειχθεί ότι μόνο για μικρές τιμές του ρ οι εκτιμήσεις του παραπάνω αλγόριθμου είναι αναξιόπιστες!

4.5 Κυκλική βήτα

Η κυκλική Βήτα κατανομή έχει πάρει το όνομά της από την συνάρτηση βήτα και έχει την συνάρτηση πυκνότητας πιθανότητας που ορίστηκε στην σχέση (2.20).

Για την εκτίμηση των παραμέτρων α, β της συνάρτησης μπορούν να χρησιμοποιηθεί η μέθοδος της μέγιστης πιθανοφάνειας. Συγκεκριμένα:

➤ Μέθοδος Μέγιστης Πιθανοφάνειας

Έστω $\{\theta_1, \theta_2, \dots, \theta_n\}$ τυχαίο δείγμα μεγέθους n από την κυκλική βήτα κατανομή με παραμέτρους α, β . Η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας της συνάρτησης πυκνότητας πιθανότητας σε αυτή την περίπτωση είναι:

$$\log L = n \log \Gamma(\alpha + \beta) - n(\alpha + \beta) \log 2 - \log \Gamma(\alpha) - \log \Gamma(\beta) +$$

$$+ \left(\alpha - \frac{1}{2}\right) \sum_{i=1}^n \log(1 + \bar{\mu}'\theta_i) + \left(\beta - \frac{1}{2}\right) \sum_{i=1}^n \log(1 - \bar{\mu}'\theta_i).$$

.

.

Θέτοντας $\cos w_i = \bar{\mu}'\theta_i$ και παραγωγίζοντας ως προς α και β κατά αντιστοιχία ισχύει:

$$\begin{cases} \varphi(\alpha) - \varphi(\alpha + \beta) = \frac{1}{n} \sum_{i=1}^n \log\left(\frac{1 + \cos w_i}{2}\right) \\ \varphi(\beta) - \varphi(\alpha + \beta) = \frac{1}{n} \sum_{i=1}^n \log\left(\frac{1 - \cos w_i}{2}\right) \end{cases}, \text{ όπου } \varphi(x) = \frac{d \log \Gamma(x)}{dx}.$$

Το σύστημα αυτό μέσω ενός επαναληπτικού αλγόριθμου μας δίνει τις επιθυμητές εκτιμήσεις για τις παραμέτρους (α, β) .

Κεφάλαιο 5

Έλεγχοι υποθέσεων και ομοιομορφίας

5.1 Εισαγωγή

Σε αυτό το κεφάλαιο, αναλύεται ίσως το πιο σύνθετο και ανεξερευνήτο κομμάτι της στατιστικής ανάλυσης των κυκλικών μεταβλητών, αυτό του ελέγχου υποθέσεων. Παρότι χρησιμοποιούμε δεδομένα που ακολουθούν τη συνηθέστερη κατανομή κυκλικών μεταβλητών την VonMises, οι δυσκολίες της πραγματοποίησης των ελέγχων είναι πάρα πολλές. Όπως θα διαπιστώσουμε στη πορεία του κεφαλαίου, το γεγονός ότι η παράμετρος συγκέντρωσης δεν είναι παράμετρος κλίμακας δυσχεραίνει την μελέτη των ελέγχων. Στη περίπτωση του ελέγχου την μέσης κατεύθυνσης αποκλείονται οι απαραμετρικοί έλεγχοι, ενώ στους ελέγχους της παραμέτρου συγκέντρωσης κ , οι κρίσιμες τιμές καθορίζονται από την υποθετική τιμή του. Τέλος παρουσιάζονται δύο βασικοί έλεγχοι ομοιομορφίας.

5.2 Μονός Πληθυσμός

A. Έλεγχος για τη μέση κατεύθυνση

Έστω δεδομένα $\alpha_i, \forall i = 1, \dots, n$ κυκλικής τυχαίας μεταβλητής που ακολουθεί την κυκλική κανονική κατανομή με μέση κατεύθυνση μ και παράμετρο συγκέντρωσης κ , και έστω $\bar{\alpha}_0$ η μέση κατεύθυνση του δείγματος μας και R το μήκος του συνιστάμενου διανύσματος των παρατηρήσεων.

- Έλεγχος λόγου πιθανοφάνειας (LRT)

➤ Η παράμετρος συγκέντρωσης κ είναι γνωστή.

Έστω ότι θέλουμε να πραγματοποιήσουμε τον έλεγχο με μηδενική υπόθεση $H_0 : \mu = \mu_0$ και εναλλακτική υπόθεση $H_1 : \mu \neq \mu_0$. Ορίζουμε τους ακόλουθους μέσους όρους για το δείγμα μας:

$$\begin{cases} \bar{C} = \frac{1}{n} \sum_{i=1}^n \cos(\alpha_i - \mu_0) \\ \bar{S} = \frac{1}{n} \sum_{i=1}^n \sin(\alpha_i - \mu_0) \end{cases}, \mu \varepsilon \quad \bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}.$$

Ο έλεγχος λόγου πιθανοφάνειας απορρίπτει τη μηδενική υπόθεση για μεγάλες τιμές της στατιστικής συνάρτησης:

$$w = 2n\kappa(\bar{R} - \bar{C}). \quad (5.1)$$

Υπό την μηδενική υπόθεση και για μεγάλες τιμές της παραμέτρου συγκέντρωσης η δειγματοσυνάρτηση $w \sim \chi_1^2$.

➤ Η παράμετρος συγκέντρωσης κ είναι άγνωστη.

Έστω ξανά ότι θέλουμε να πραγματοποιήσουμε τον έλεγχο με μηδενική υπόθεση $H_0: \mu = \mu_0$ και εναλλακτική υπόθεση $H_1: \mu \neq \mu_0$.

Ο έλεγχος λόγου πιθανοφάνειας απορρίπτει την μηδενική υπόθεση για μεγάλες τιμές της στατιστικής συνάρτησης:

$$w = 2n\hat{\kappa}(\bar{R} - \bar{C}), \quad (5.2)$$

όπου $\hat{\kappa} = A^{-1}\left(\frac{R}{n}\right)$, όπως προκύπτει από τη σχέση (4.6).

B. Έλεγχος για την παράμετρο συγκέντρωσης

- Η μέση κατεύθυνση μ είναι γνωστή
 - i. Έλεγχος με μηδενική υπόθεση $H_0: \kappa = \kappa_0 (\neq 0)$

Στην περίπτωση που η τιμή της μέσης κατεύθυνσης είναι γνωστή η συνάρτηση $V = \sum_{i=1}^n \cos(\alpha_i - \mu_0)$ όπως ορίστηκε στην σχέση (3.3) είναι επαρκής για την εκτίμηση της παραμέτρου συγκέντρωσης κ , και μάλιστα γνωρίζουμε και την κατανομή της υπό την μηδενική υπόθεση ($f_{\kappa_0}(v)$) χρησιμοποιώντας την σχέση (3.11).

Συγκεκριμένα, για τον έλεγχο με εναλλακτική υπόθεση την $H_1: \kappa \neq \kappa_0$ απορρίπτουμε την H_0 όταν $V \leq v_1$ ή $V \geq v_2$ με $P_0(V \leq v_1) + P_0(V \geq v_2) = \alpha$, όπου α το επίπεδο σημαντικότητας του ελέγχου.

- ii. Έλεγχος με μηδενική υπόθεση $H_0: \kappa = 0$ και εναλλακτική $H_1: \kappa > 0$

Στην περίπτωση αυτή, ο έλεγχος μας βασίζεται στην στατιστική συνάρτηση V και απορρίπτει την H_0 , όταν $V \geq v_0$ όπου το v_0 λαμβάνεται βάσει του επιθυμητού επιπέδου σημαντικότητας. Η κατανομή της στατιστικής συνάρτησης V σε αυτή την περίπτωση είναι η $f_0(v)$ όπως ορίστηκε στη σχέση (3.4).

- Η μέση κατεύθυνση μ είναι άγνωστη
 - i. Έλεγχος με μηδενική υπόθεση $H_0: \kappa = \kappa_0 (\neq 0)$

Στην περίπτωση που η μέση κατεύθυνση του δείγματος μας είναι άγνωστη η κατάλληλη στατιστική συνάρτηση για τον έλεγχο με εναλλακτική υπόθεση την $H_1: \kappa \neq \kappa_0$ είναι το μήκος του συνιστάμενου διανύσματος R . Συγκεκριμένα ο έλεγχος απορρίπτει την μηδενική υπόθεση για $R > r_a$, όπου το r_a καθορίζεται από το επίπεδο σημαντικότητας α . Τέλος γνωρίζουμε την ακριβή κατανομή της R υπό την μηδενική υπόθεση ($f_{\kappa_0}(r)$) μέσω της σχέσης (3.10) και κατά συνέπεια μπορούμε να προσδιορίσουμε και τα ακριβή κρίσιμα σημεία.

- ii. Έλεγχος με μηδενική υπόθεση $H_0: \kappa = 0$ και εναλλακτική $H_1: \kappa > 0$

Ακολουθώντας παρόμοιο συλλογισμό με την περίπτωση που η τιμή της μέσης κατεύθυνσης είναι γνωστή, ο έλεγχος βασίζεται στο μήκος R , και απορρίπτει την μηδενική υπόθεση όταν $R > r_0$. Τέλος, η ακριβής κατανομή της R υπό την ομοιόμορφη δίνεται ως γνωστόν από την $f_0(r)$, όπως ορίστηκε στην σχέση (3.1).

5.3 Δύο ή περισσότεροι πληθυσμοί

A. Έλεγχος μέσης κατεύθυνσης

- Ακριβής Έλεγχος

Θεωρούμε δύο κυκλικούς ανεξάρτητους κανονικούς πληθυσμούς με παραμέτρους (μ_1, κ_1) , (μ_2, κ_2) αντίστοιχα. Θα εκτελέσουμε τον έλεγχο με μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ και εναλλακτική υπόθεση την $H_1: \mu_1 \neq \mu_2$, υπό την προϋπόθεση ύπαρξης κοινής αλλά άγνωστης παραμέτρου συγκέντρωσης $\kappa = \kappa_1 = \kappa_2$.

Εάν συμβολίσουμε ως R_1, R_2 τα μήκη των επιμέρους συνιστάμενων διανυσμάτων και ως R το αντίστοιχο του συνδιασμένου δείγματος τότε προφανώς ισχύει:

$$|R_1 - R_2| \leq R \leq R_1 + R_2.$$

Όμως, εάν $\mu_1 = \mu_2$ τότε $S \equiv R_1 + R_2 \approx R$ και η κατανομή του S λαμβάνεται από την σχέση (3.10). Συνεπώς, ένας προσεγγιστικός έλεγχος για την απόρριψη της H_0 είναι εάν $S > s_0$ με το s_0 επιλεγμένο, έτσι ώστε $P_{H_0}^\kappa(S > s_0 | R) = \alpha$, όπου το α είναι σταθερά επιλεγμένη βάσει του επιθυμητού επιπέδου σημαντικότητας του ελέγχου.

- Προσεγγιστικός έλεγχος και ANOVA

Στην περίπτωση που θέλουμε να εξετάσουμε την ισότητα δύο ή περισσότερων μέσων κατευθύνσεων μεταξύ ανεξάρτητων κυκλικών κανονικών κατανομών με την ίδια παράμετρο συγκέντρωσης, ακολουθούμε μια διαφορετική προσέγγιση. Υποθέτουμε, καταρχάς, ότι η παράμετρος συγκέντρωσης είναι αρκετά μεγάλη για να ισχύουν οι προσεγγίσεις που παρατέθηκαν στην παράγραφο (3.5).

Εκτελούμε τον έλεγχο με μηδενική υπόθεση $H_0 : \mu_1 = \mu_2 = \dots = \mu_p$, όπου μ_i οι πολικές κατευθύνσεις των p πληθυσμών με την ίδια άγνωστη παράμετρο συγκέντρωσης κ .

Θέτουμε ως n_i, R_i το μέγεθος και το μήκος του συνιστάμενου διανύσματος, των επιμέρους δειγμάτων από κάθε πληθυσμό, ενώ ως n, R τα αντίστοιχα μεγέθη για το σύνολο των δειγμάτων. Τέλος ορίζουμε ως μ τη κοινή κατεύθυνση υπό την H_0 .

Όπως έχουμε αναλύσει στην πορεία της εργασίας, ένα μέτρο διασποράς γύρω από την μέση κατεύθυνση είναι το $(n - R)$ για το σύνολο των δειγμάτων και το $(n_i - R_i)$ για την εσωτερική διασπορά στα επιμέρους δείγματα. Έτσι, προκύπτει η ακόλουθη σχέση:

$$n - R = \left(\sum_{i=1}^p R_i \right) + \left(\sum_{i=1}^p R_i - R \right) = \left(\sum_{i=1}^p (n_i - R_i) \right) + \left(\sum_{i=1}^p R_i - R \right).$$

Όμως έχουμε ήδη αποδείξει στην παράγραφο (3.5) ότι για μεγάλες τιμές του κ ισχύει:

$$2\kappa(n - R) \sim \chi_{n-1}^2,$$

$$2\kappa(n - R) = 2\kappa \sum_{i=1}^p (n_i - R_i) + 2\kappa \left(\sum_{i=1}^p R_i - R \right) \sim \chi_{n-p}^2 + \chi_{p-1}^2.$$

Βάσει αυτού, μπορούμε να κατασκευάσουμε την ακόλουθη προσεγγιστική ανάλυση διασποράς (ANOVA):

Πηγή	Βαθμοί ελευθερίας	Άθροισμα Τετραγώνων
Μεταξύ ομάδων	$p - 1$	$\sum_{i=1}^p R_i - R$
Εντός ομάδων	$n - p$	$\sum_{i=1}^p (n_i - R_i)$
Συνολική	$n - 1$	$n - R$

Από τον πίνακα λαμβάνουμε το στατιστικό ελέγχου:

$$F = \frac{\left(\sum_{i=1}^p R_i - R \right)}{(p-1)} \cdot \frac{\sum_{i=1}^p (n_i - R_i)}{(n-p)} \quad (5.3)$$

Παρατηρούμε, ότι ισχύει πάντα $R \leq \sum_i^p R_i$, αλλά στην περίπτωση που το άθροισμα των επιμέρους μηκών είναι σημαντικά μεγαλύτερο από αυτό του συνολικού, έχουμε βάσιμους λόγους να απορρίψουμε την μηδενική υπόθεση.

Εν γένει, βάσει του ότι $F \sim F(p-1, n-p)$ για μεγάλες τιμές της παραμέτρου συγκέντρωσης, ο έλεγχος απορρίπτει την μηδενική υπόθεση εάν

$$F \equiv \frac{(n-p)(V-R)}{(p-1)(n-V)} > F_0.$$

Στην περίπτωση που έχουμε $1 < \kappa < 2$, χρησιμοποιούμε την στατιστική $F' = F \cdot \left(1 + \frac{3}{8\hat{\kappa}} \right)$ με την F' να ακολουθεί την κατανομή $F(p-1, n-p)$ και $\hat{\kappa}$ να είναι η εκτίμηση της συνολικής παραμέτρου συγκέντρωσης.

Β. Έλεγχοι παραμέτρου συγκέντρωσης

Σε αυτό το στάδιο επιθυμούμε να εξετάσουμε το κατά πόσο δύο ανεξάρτητες κυκλικές κανονικές κατανομές εμφανίζουν την ίδια παράμετρο συγκέντρωσης. Με άλλα λόγια, πραγματοποιούμε έναν έλεγχο για τις $CN(\mu_i, \kappa_i)$, $i=1,2$ με μηδενική υπόθεση $H_0 : \kappa_1 = \kappa_2 = \kappa$ με άγνωστη τιμή για το κ και εναλλακτική υπόθεση την $H_1 : \kappa_1 \neq \kappa_2$. Εάν συμβολίσουμε ως (n_1, R_1, n_2, R_2) τα μεγέθη και τα μήκη των συνιστάμενων διανυσμάτων για τα δύο δείγματα, τότε μπορούμε να χρησιμοποιήσουμε τις παρακάτω μεθόδους.

- Ακριβής Έλεγχος

Γνωρίζουμε ήδη ότι για την κυκλική κανονική κατανομή ισχύει:
 $\hat{\kappa}_i = A^{-1}\left(\frac{R_i}{n_i}\right)$, $i = 1, 2$ και η $A^{-1}(\cdot)$ είναι μονότονη συνάρτηση. Συνεπώς, μπορούμε να πραγματοποιήσουμε έναν ακριβή στατιστικό έλεγχο βάσει της ποσότητας:

$$D = \left| \frac{R_1}{n_1} - \frac{R_2}{n_2} \right|,$$

και να απορρίψουμε την μηδενική υπόθεση, εάν $D > d_0$ με d_0 επιλεγμένο ανάλογα με το επιθυμητό επίπεδο σημαντικότητας.

Ο παραπάνω έλεγχος εμφανίζει μια αδυναμία εν γένει. Αυτό συμβαίνει, γιατί οι ποσότητες (R_1, R_2) και συνακόλουθα και η D έχουν κατανομή που εμπεριέχει την παράμετρο συγκέντρωσης για την οποία πραγματοποιούμε τον έλεγχο. Για αυτό εκμεταλλευόμενοι το γεγονός της ανεξαρτησίας της $f_{\kappa}(R_1, R_2 | R)$ (η οποία προκύπτει από την εξίσωση (3.15) για $q = 2$) από την ποσότητα κ , χρησιμοποιούμε το στατιστικό ελέγχου δεδομένου R υπολογίζοντας την πιθανότητα $P(D \geq d_0 | R) = \alpha$.

Τέλος, στην ειδική περίπτωση που τα μεγέθη των δειγμάτων είναι ίσα το στατιστικό ελέγχου, απλοποιείται στην απόλυτη τιμή της διαφοράς των δύο μηκών των επιμέρους συνιστάμενου διανύσματος, δεδομένου του μήκους του αντίστοιχου συνολικού.

- Διάστημα εμπιστοσύνης για μεγάλο κ

Όπως έχουμε δει, για μεγάλες τιμές του κ ισχύει $2\kappa(n - R) \overset{\sim}{\sim} X_{n-1}^2$. Βάσει αυτού, μπορούμε να κατασκευάσουμε ένα διάστημα εμπιστοσύνης για το κ της μορφής:

$$P[\kappa_L < \kappa < \kappa_U] = 1 - \alpha.$$

Από τον πίνακα της κατανομής X^2 με $(n - 1)$ βαθμούς ελευθερίας, λαμβάνουμε τις τιμές X_L^2, X_U^2 ώστε $P[X_L^2 < X_{(n-1)}^2 < X_U^2] = 1 - \alpha$.

Συνεπώς προκύπτει:

$$P[X_L^2 < 2\kappa(n - R) < X_U^2] = 1 - \alpha \Rightarrow \begin{cases} \kappa_L = \frac{X_L^2}{2(n - R)} \\ \kappa_U = \frac{X_U^2}{2(n - R)} \end{cases},$$

και δημιουργούμε το $100(1 - \alpha)\%$ διάστημα εμπιστοσύνης που ζητάμε.

- Προσεγγιστικός έλεγχος ισότητας παραμέτρου συγκέντρωσης

Χρησιμοποιώντας μία προσέγγιση των συναρτήσεων Bessel, έχουμε για μεγάλες τιμές του κ

$$\frac{I_0(\kappa)}{I_1(\kappa)} \approx 1 - \frac{1}{2\kappa}.$$

Για την εκτίμηση λοιπόν του κ έχουμε:

$$1 - \frac{1}{2\kappa} \approx \frac{R}{n} \quad \text{ή} \quad \hat{\kappa} \approx \frac{n}{2(n-R)}.$$

Στην πραγματικότητα έχει αποδειχθεί ότι η ποσότητα $\hat{\kappa} = \frac{n-1}{2(n-R)}$, μας δίνει μια ελαφρώς καλύτερη προσέγγιση του κ . Για μια ακόμη φορά χρησιμοποιούμε την προσέγγιση $\frac{2\kappa(n-R)}{n-1} \sim \frac{\chi^2_{n-1}}{n-1}$ και συνδυάζοντας την με την παραπάνω εκτίμηση προκύπτει:

$$\frac{\kappa}{\hat{\kappa}} \sim \frac{\chi^2_{n-1}}{n-1}.$$

Με αφετηρία αυτό το συμπέρασμα, για να εξετάσουμε την μηδενική υπόθεση $H_0 : \kappa_1 = \kappa_2$ με εναλλακτική την $H_1 : \kappa_1 \neq \kappa_2$ χρησιμοποιούμε το στατιστικό ελέγχου:

$$F = \frac{\hat{\kappa}_1}{\hat{\kappa}_2} = \frac{(n_1 - R_1)(n_2 - 1)}{(n_2 - R_2)(n_1 - 1)},$$

το οποίο υπό την μηδενική υπόθεση ακολουθεί την κατανομή F_{n_1-1, n_2-1} .

5.4 Έλεγχοι ομοιομορφίας

Μια πολύ σημαντική πτυχή της στατιστικής ανάλυσης κυκλικών δεδομένων είναι αυτή του ελέγχου της ομοιομορφίας. Δηλαδή, το κατά πόσο οι παρατηρήσεις μας που προέρχονται από ένα ή περισσότερα δείγματα, ακολουθούν ή όχι την ομοιόμορφη κυκλική κατανομή ή εναλλακτικά εάν εμφανίζουν συνηθέστερα κάποια συγκεκριμένη κατεύθυνση. Στην παρούσα διπλωματική θα αναλύσουμε δύο από τους σημαντικότερους ελέγχους σε αυτό τον τομέα, τον έλεγχο Rayleigh και τον έλεγχο Rao.

- **Έλεγχος Rayleigh**

Έστω σύνολο παρατηρήσεων $(\alpha_1, \alpha_2, \dots, \alpha_n)$ κυκλικής τυχαίας μεταβλητής με το μήκος του συνιστάμενου διανύσματος τους να είναι ίσο με R . Τότε μπορούμε να πραγματοποιήσουμε τον έλεγχο Rayleigh με μηδενική υπόθεση την H_0 : Η τυχαία μεταβλητή ακολουθεί την κυκλική ομοιόμορφη κατανομή και εναλλακτική την H_1 : Η τυχαία μεταβλητή ακολουθεί κάποια κατανομή με μία ή δύο κορυφές. Η στατιστική συνάρτηση που χρησιμοποιούμε στον συγκεκριμένο έλεγχο είναι η:

$$z = nr^2, \quad (5.4)$$

όπου $r = \left(\frac{R}{n}\right)$, και n το μέγεθος του δείγματος. Ο έλεγχος απορρίπτει την H_0 για $z < a$, όπου a επιλέγεται βάσει του επιπέδου σημαντικότητας. Η κατανομή της στατιστικής συνάρτησης για μέγεθος δείγματος $n < 15$ θεωρείται ότι προσεγγίζει την κανονική κατανομή ενώ για μεγαλύτερα δείγματα χρησιμοποιείται η συνάρτηση

$$z' = 2nr^2, \quad (5.5)$$

η οποία σύμφωνα με την έρευνα των Mardia K. and Jupp P. (2000), Directional Statistics, p. 94-95, υπό την μηδενική υπόθεση προσεγγίζει ασυμπτωτικά την χ^2_2 .

- **Έλεγχος Rao**

Ο συγκεκριμένος απαραμετρικός έλεγχος βασίζεται στην ιδέα ότι εάν η κατανομή της τυχαίας μεταβλητής από την οποία προέρχεται το δείγμα είναι η ομοιόμορφη κυκλική, τότε οι παρατηρήσεις μας $\alpha_i, \forall i = 1, \dots, n$ θα πρέπει προσεγγιστικά να είναι ομοιόμορφα κατανεμημένες στην περιφέρεια του κύκλου. Δηλαδή το κάθε τόξο ανάμεσα στις παρατηρήσεις μας να είναι περίπου ίσο με $\left(\frac{360}{n}\right)^\circ$.

Πραγματοποιούμε λοιπόν τον έλεγχο με μηδενική υπόθεση H_0 : Η τυχαία μεταβλητή ακολουθεί την κυκλική ομοιόμορφη κατανομή και εναλλακτική την H_1 : Η τυχαία μεταβλητή ακολουθεί κάποια κατανομή με περισσότερες από μία κορυφές.

Η κατάλληλη στατιστική συνάρτηση για τον συγκεκριμένο έλεγχο είναι η

$$U = \frac{1}{2} \sum_{i=1}^n |T_i - \lambda|. \quad (5.6)$$

Η ποσότητα T_i ορίζεται από τον ακόλουθο αναδρομικό τύπο

$$T_i = \alpha_{i+1} - \alpha_i \quad \forall i \in [1, n-1], \quad \lambda = \frac{2\pi}{n}, \quad \text{και} \quad T_n = (2\pi - \alpha_n) + \alpha_1,$$

όπου α_1, α_n είναι η πρώτη και η τελευταία παρατήρηση διατεταγμένες κατά αύξουσα σειρά.

Όμως υπό την μηδενική υπόθεση το άθροισμα των θετικών αποκλίσεων ισούται με αυτό των αρνητικών αποκλίσεων και συνεπώς η στατιστική μας συνάρτηση παίρνει την μορφή:

$$U = \sum_{i=1}^n [T_i - \lambda].$$

Ο συγκεκριμένος έλεγχος απορρίπτει την μηδενική υπόθεση για μικρές τιμές της ποσότητας U , της οποίας η συνάρτηση κατανομής πιθανότητας δίνεται από τον τύπο:

$$f(u) = (n-1)! \sum_{j=1}^n \left[\binom{n}{j} \left(\frac{u}{2\pi} \right)^{n-j-1} \frac{g_j(nu)}{(n-j-1)! n^{j-1}} \right], \quad u \in \left(0, 2\pi \left[1 - \frac{1}{n} \right] \right),$$

$$\text{με } g_j(nu) = \left(\frac{1}{(j-1)! 2\pi} \right) \sum_{w=0}^{\infty} \left\{ (-1)^w \binom{j}{w} \left[\left(\frac{nu}{2\pi} \right) - w \right]^{j-1} \right\}.$$

Κεφάλαιο 6

Συντελεστής συσχέτισης-παλινδρόμησης

6.1 Εισαγωγή

Σε πολλές περιπτώσεις μελετών τα δεδομένα που προκύπτουν, είτε διμετάβλητα είτε πολλαπλών μεταβλητών, αποτελούνται από μείγμα κυκλικών και γραμμικών μεταβλητών. Σε αντιστοιχία με την γραμμική περίπτωση, μας απασχολεί φυσικά το κατά πόσο σχετίζονται αυτές οι μεταβλητές μεταξύ τους και κατά πόσο οι τιμές της μίας επηρεάζουν τις τιμές των ετέρων. Σκοπός μας, λοιπόν, είναι να προσδιορίσουμε τις κατάλληλες ποσότητες που θα καθορίζουν τη συσχέτιση των μεταβλητών, πράγμα που θα μας επιτρέψει να προσδιορίσουμε την τιμή μίας μεταβλητής δεδομένης κάποιας ή κάποιων άλλων.

6.2 Συντελεστής συσχέτισης κυκλικών μεταβλητών

Στο συγκεκριμένο κομμάτι θα επικεντρωθούμε στα δεδομένα δύο κυκλικών μεταβλητών. Έστω τυχαίο δείγμα παρατηρήσεων $[(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)]$, δύο κυκλικών μεταβλητών A, B , με $0 \leq \alpha_i, \beta_j \leq 2\pi \forall i, j = 1, \dots, n$, και την ίδια μηδενική κατεύθυνση και σύμβαση θετικής φοράς. Ας συμβολίσουμε επίσης ως $f(\alpha, \beta)$ την από κοινού συνάρτηση πυκνότητας πιθανότητας και ως μ, ν τις μέσες κατευθύνσεις των δύο μεταβλητών αντίστοιχα.

Μια αποδεκτή ποσότητα ως συντελεστής κυκλικής συσχέτισης, που διατηρεί κατά αντιστοιχία αρκετές από τις ιδιότητες του συντελεστή στην γραμμική περίπτωση είναι:

$$\rho_c(\alpha, \beta) = \frac{E_{A,B} \{ \sin(\alpha - \mu) \sin(\beta - \nu) \}}{\sqrt{V_{A,B}(\sin(\alpha - \mu)) V_{A,B}(\sin(\beta - \nu))}}, \quad \alpha, \beta \in [0, 2\pi]. \quad (6.1)$$

Αυτή είναι μια φυσιολογικά παραγόμενη σχέση, καθώς οι όροι $\sin(\alpha - \mu)$, $\sin(\beta - \nu)$ αναπαριστούν την απόκλιση των α, β από τις μέσες κατευθύνσεις τους αντίστοιχα και συνεπώς είναι ένας δείκτης του κατά πόσο υπάρχει συσχέτιση μεταξύ τους. Συνακόλουθα μπορούμε να δούμε, ότι $E_{A,B}(\sin(\alpha - \mu)) = E_{A,B}(\sin(\beta - \nu)) = 0$ που είναι συμβατό κατά αναλογία με την γραμμική περίπτωση, όπου η κεντρική ροπή είναι μηδενική.

Στο σημείο αυτό, οφείλουμε να κάνουμε κάποιες παρατηρήσεις που αφορούν στην καλύτερη κατανόηση του συντελεστή συσχέτισης:

1. Ο συντελεστής συσχέτισης μπορεί να γραφεί στην μορφή:

$$\rho_c(\alpha, \beta) = \frac{E_{A,B}[\cos(\alpha - \beta - \mu + \nu) - \cos(\alpha + \beta - \mu - \nu)]}{2\sqrt{E_A(\sin^2(\alpha - \mu))E_B(\sin^2(\beta - \nu))}}.$$

Εφόσον η ποσότητα $E_A(\cos(\alpha - \mu))$ αποτελεί μια εκτίμηση της συγκέντρωσης του α γύρω από την μέση κατεύθυνση, παρατηρούμε ότι ο πρώτος όρος του αριθμητή αναπαριστά το κατά πόσο συγκεντρωμένη είναι η κατανομή της τυχαίας μεταβλητής $(A - \mu) - (B - \nu)$ και συνδράμει στο θετικό κομμάτι της συσχέτισης. Αντίστοιχα, ο δεύτερος όρος στον αριθμητή ποσοτικοποιεί το αρνητικό κομμάτι της συσχέτισης.

2. Στην περίπτωση που οι μέσες κατευθύνσεις μ, ν δεν μας είναι γνωστές είτε δεν ορίζονται όπως στην περίπτωση που είτε η μία είτε η άλλη τυχαία μεταβλητή μας προέρχεται από την ομοιόμορφη κατανομή, επιλέγονται έτσι, ώστε να μεγιστοποιούν τους όρους του αριθμητή: $|E_{A,B} \cos(\alpha + \beta - \mu - \nu)|$ και $|E_{A,B} \cos(\alpha + \beta - \mu - \nu)|$ ξεχωριστά. Έτσι, οποτεδήποτε υπάρχει αμφιβολία για την επιλογή της μέσης κατεύθυνσης των δύο ποσοτήτων, επιλέγουμε τέτοιες τιμές, ώστε να επιτύχουμε την μεγαλύτερη δυνατή σχέση στις θετικές και αρνητικές κατευθύνσεις. Αυτό οδηγεί στην επιλογή των μ, ν , ώστε η ποσότητα $(\mu - \nu)$ να αναπαριστά την μέση κατεύθυνση της τυχαίας μεταβλητής $(A - B)$ ενώ η ποσότητα $(\mu + \nu)$ την μέση κατεύθυνση της τυχαίας μεταβλητής $(A + B)$.

Σε αυτή την περίπτωση ο αριθμητής στη σχέση του συντελεστή συσχέτισης αναπαριστά τη διαφορά στα μήκη των μέσων κατευθύνσεων των παραπάνω ποσοτήτων. Εάν συγκεκριμένα συμβολίσουμε $R_{\alpha \pm \beta} = |E_{A,B} e^{i(\alpha \pm \beta)}|$, τότε:

$$\rho_c(\alpha, \beta) = \frac{(R_{\alpha - \beta} - R_{\alpha + \beta})}{2\sqrt{E_A(\sin^2(\alpha - \mu))E_B(\sin^2(\beta - \nu))}}. \quad (6.2)$$

Ο συντελεστής κυκλικής συσχέτισης ικανοποιεί μια σειρά από ιδιότητες οι βασικότερες εκ των οποίων είναι οι ακόλουθες:

- $\rho_c(\alpha, \beta)$ δεν εξαρτάται από την μηδενική κατεύθυνση για καμία από τις μεταβλητές.
- $\rho_c(\alpha, \beta) = \rho_c(\beta, \alpha)$.
- $|\rho_c(\alpha, \beta)| \leq 1$.
- $\rho_c(\alpha, \beta) = 0$ εάν οι τυχαίες μεταβλητές (A, B) είναι ανεξάρτητες μεταξύ τους.

Τέλος, παραθέτουμε τον δειγματικό συντελεστή συσχέτισης στην περίπτωση που $[(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)]$ είναι ένα τυχαίο δείγμα παρατηρήσεων:

$$r_{c,n} = \frac{\sum_{i=0}^n \sin(\alpha_i - \bar{\alpha}) \sin(\beta_i - \bar{\beta})}{\sqrt{\sum_{i=0}^n \sin^2(\alpha_i - \bar{\alpha}) \sin^2(\beta_i - \bar{\beta})}}, \quad (6.3)$$

όπου $\bar{\alpha}, \bar{\beta}$ είναι δειγματικές μέσες διευθύνσεις. Η ποσότητα αυτή είναι μία εκτίμηση του συντελεστή κυκλικής συσχέτισης που αναλύσαμε παραπάνω, χρησιμεύει δε στην περίπτωση που η από κοινού κατανομή των α, β δεν είναι πλήρως γνωστή για να πραγματοποιήσουμε προσεγγιστικούς ελέγχους υποθέσεων για το ρ_c , όταν το μέγεθος του δείγματος είναι αρκετά μεγάλο.

6.2.1 Συντελεστής ρ_c σε μοντέλο με ομοιόμορφες περιθώριες

Σε αυτό το σημείο, θα δούμε μια εφαρμογή του συντελεστή κυκλικής συσχέτισης σε ένα μοντέλο με συγκεκριμένα χαρακτηριστικά.

Έστω, λοιπόν, η τυχαία μεταβλητή A να προέρχεται από την ομοιόμορφη κατανομή και η δεσμευμένη κατανομή της τυχαίας μεταβλητής B δεδομένης της A είναι μια κυκλική κανονική κατανομή με συνάρτηση πυκνότητας πιθανότητας:

$$f(\beta|\alpha; \theta, \varepsilon, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \{\cos(\beta - \varepsilon\alpha - \theta)\}] \quad \kappa \geq 0, 0 \leq \theta < 2\pi, \varepsilon = \pm 1. \quad (6.4)$$

Η περιθώρια κατανομή της B θα είναι επίσης η ομοιόμορφη κυκλική, ενώ η από κοινού κατανομή των A, B θα έχει την συνάρτηση πυκνότητας πιθανότητας:

$$f(\alpha, \beta; \theta, \varepsilon, \kappa) = \frac{1}{4\pi^2 I_0(\kappa)} \exp[\kappa \{\cos(\beta - \varepsilon\alpha - \theta)\}]. \quad (6.5)$$

Επειδή οι τυχαίες μεταβλητές A, B προέρχονται από την ομοιόμορφη κυκλική κατανομή, θα επιλέξουμε τις μέσες κατευθύνσεις μ, ν σύμφωνα με την δεύτερη παρατήρηση που αφορά τον συντελεστή συσχέτισης και παρουσιάστηκε στην παράγραφο (6.2).

Στην περίπτωση που η άγνωστη παράμετρος ε παίρνει την τιμή ένα ο όρος του αριθμητή της εξίσωσης (6.1) είναι:

$$\begin{aligned}
E_{A,B} \{ \cos(\alpha + \beta - \mu - \nu) \} &= E_A E_{B|A} \{ \cos(\beta - \alpha - \theta + (2\alpha + \theta - \mu - \nu)) \} \\
&= \frac{I_1(\kappa)}{I_0(\kappa)} E_A \{ \cos(2\alpha + \theta - \mu - \nu) \} = 0.
\end{aligned} \tag{6.6}$$

Όπως ξέρουμε, πρέπει να μεγιστοποιήσουμε την ποσότητα $E_{A,B} \{ \cos(\alpha - \beta - \mu + \nu) \}$ και αυτό συμβαίνει όταν η ποσότητα $(\mu - \nu)$ είναι η μέση κατεύθυνση της τυχαίας μεταβλητής $A - B$, η οποία είναι ίση με $-\theta$.

Τότε προκύπτει:

$$\begin{cases} E_{A,B} \{ \cos(\beta - \alpha - \theta) \} = A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)} \\ E_A \{ \sin^2(\alpha - \mu) \} = E_B \{ \sin^2(\beta - \nu) \} = \frac{1}{2} \end{cases},$$

και φυσικά $\rho_c = A(\kappa)$.

Αντίστοιχα, στην περίπτωση που $\varepsilon = -1$ οδηγούμαστε πανομοιότυπα στον συντελεστή

$$\rho_c = -A(\kappa).$$

6.3 Συντελεστής συσχέτισης κυκλικών-γραμμικών μεταβλητών

Σε πολλές μελέτες χρειάζεται να προσδιοριστεί η συσχέτιση μεταξύ κυκλικών και γραμμικών μεταβλητών. Ένα παράδειγμα τέτοιου είδους προβλήματος, προκύπτει εάν θέλουμε να μελετήσουμε την κατεύθυνση του ανέμου σε σχέση με την ποσότητα της βροχής ή την συγκέντρωση της ατμοσφαιρικής ρύπανσης.

Έστω λοιπόν κυκλική τυχαία μεταβλητή A και «γραμμική» τυχαία μεταβλητή X , καθώς και δείγμα παρατηρήσεών τους $(\alpha_1, x_1), \dots, (\alpha_n, x_n)$. Η μέθοδος η οποία ακολουθείται για τον προσδιορισμό του συντελεστή συσχέτισης μεταξύ τους βασίζεται στην μετατροπή των παρατηρήσεων της κυκλικής μεταβλητής σε γραμμική μορφή, για να μπορέσουμε να χρησιμοποιήσουμε τον δειγματικό συντελεστή γραμμικής συσχέτισης

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \cdot \sqrt{\sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}}. \tag{6.7}$$

Για να επιτύχουμε την επιθυμητή μορφή των δεδομένων μετατρέπουμε τις παρατηρήσεις της κυκλικής μεταβλητής μας στο ζεύγος «γραμμικών» μεταβλητών $(c_i, s_i) \equiv (\cos \alpha_i, \sin \alpha_i), \forall i = [1, n]$ και υπολογίζουμε τον συντελεστή συσχέτισης των τυχαίων μεταβλητών X, A βάσει του ακόλουθου τύπου:

$$r^2 = [r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs} / 1 - r_{cs}^2],$$

όπου r_{xc}, r_{xs}, r_{cs} υπολογίζονται σύμφωνα με τον τύπο (6.7).

Μία εναλλακτική μετατροπή των δεδομένων της κυκλικής μεταβλητής μας είναι εφικτή υπό προϋποθέσεις, στην περίπτωση που έχουμε προσαρμόσει ένα γραμμικό μοντέλο παλινδρόμησης και θα αναφερθεί στην επόμενη παράγραφο.

Τέλος στην περίπτωση που οι τυχαίες μεταβλητές είναι ανεξάρτητες μεταξύ τους, με την «γραμμική» X να ακολουθεί την κανονική κατανομή, σύμφωνα με τους Liddel and Ord (1978) ισχύει η ακόλουθη προσέγγιση για την κατανομή του r :

$$\frac{(n-3)r^2}{1-r^2} \sim F_{2,n-3}.$$

6.4 Παλινδρόμηση κυκλικών-γραμμικών μεταβλητών

Όπως έχουμε ήδη αναφέρει, σε πολλές μελέτες που αναλύουμε προκύπτουν προβλήματα με γραμμική μεταβλητή απόκρισης και κυκλική επεξηγηματική μεταβλητή. Ένα απλό παράδειγμα, στο οποίο προκύπτουν δεδομένα της μορφής (y, α) , είναι η μέτρηση του ποσοστού βροχόπτωσης (Y) σε σχέση με την κατεύθυνση του ανέμου (A).

Στην περίπτωση που η κυκλική μεταβλητή A έχει μοναδική περίοδο, προτείνεται το εξής απλό μοντέλο παλινδρόμησης για το περιοδικό φαινόμενο:

$$E_Y(Y | A = \alpha) = D_0 + D_1 \cos w(\alpha - \alpha_0), \quad (6.8)$$

όπου D_0, D_1 είναι άγνωστες σταθερές, α_0 η γωνία ακρόφασης και w η γωνιακή συχνότητα $\left(w = \frac{2\pi}{T} \alpha \nu \text{ T η περίοδος του } \alpha \right)$. Ως γωνία ακρόφασης καλούμε την τιμή α_0 για την οποία το μοντέλο μας παίρνει την μέγιστη τιμή.

Το μοντέλο μας μπορεί να γραφεί εναλλακτικά και στην μορφή:

$$E_Y (Y | A = \alpha) = D_0 + D_1 \cos(w\alpha - \varphi), \varphi = w\alpha_0,$$

όπου φ το σημείο που η μεταβλητή Y φτάνει στην μέγιστη τιμή της.

Το γραμμικό μοντέλο λοιπόν που προκύπτει είναι το

$$y_i = D_0 + D_1 \cos(wa_i - a_0) + \varepsilon_i,$$

όπου ε_i είναι το τυχαίο σφάλμα του μοντέλου μας και ακολουθεί την $N(0, \sigma^2)$.

Το παραπάνω γραμμικό μοντέλο μπορεί να γραφεί επίσης στην μορφή

$$y_i = D_0 + D_2 \cos(wa_i) + D_3 \sin(wa_i) + \varepsilon_i,$$

όπου $D_2 = D_1 \cos(w\alpha_0)$, $D_3 = D_1 \sin(w\alpha_0)$. Οι σταθερές του μοντέλου (D_0, D_2, D_3) μπορούν να βρεθούν με την μέθοδο των ελαχίστων τετραγώνων και σύμφωνα με αυτές να υπολογίσουμε και τις ποσότητες D_1, α_0 μέσω των τύπων:

$$D_1 = \sqrt{D_2^2 + D_3^2}, w\alpha_0 = \arctan^* \left(\frac{D_3}{D_2} \right).$$

Έχοντας υπολογίσει την γωνία ακρόφασης μπορούμε να προχωρήσουμε στον προσδιορισμό του συντελεστή συσχέτισης των δύο μεταβλητών (Y, A) ορίζοντας την ποσότητα $x_i = \cos(wa_i - \varphi)$ και χρησιμοποιώντας την σχέση (6.7) για τα ζεύγη παρατηρήσεων (x_i, y_i).

Παρατηρούμε ότι το μέγιστο και το ελάχιστο της ταλάντωσης που περιγράφεται στην σχέση (6.8) παρατηρούνται σε ίσους μεταξύ τους χρόνους μέσα σε μία περίοδο. Στην περίπτωση που η ταλάντωση αλλοιώνεται από εξωτερική επίδραση, ώστε να μην εναλλάσσονται πια σε ίσα διαστήματα έχουμε το φαινόμενο της λοξής ταλάντωσης. Η μοντελοποίηση τότε έχει ως εξής:

$$E_Y (Y | A = \alpha) = D_0 + D_1 \cos(\psi + \nu \cos \psi) \text{ όπου } \psi = (w\alpha - \varphi).$$

Η παράμετρος ν καλείτε παράμετρος λοξότητας και όσο πιο πολύ απομακρύνεται από την τιμή μηδέν τόσο μεγαλύτερη είναι η λοξότητα. Κινείται δε συνήθως στο διάστημα $-30^\circ \leq \nu \leq 30^\circ$. Η γωνία $\psi = \psi_0$ καλείται γωνία ακρόφασης και είναι η γωνία για την οποία η μεταβλητή απόκρισης λαμβάνει το μέγιστό της ($D_0 + D_1$). Για τον καθορισμό της εν λόγω γωνίας ζητάμε φυσικά την τιμή που μηδενίζει την παράγωγο:

$$\frac{dy}{d\psi} = -\sin(\psi + \nu \cos \psi)(1 - \nu \sin \psi).$$

Προκύπτει, λοιπόν, ότι αφού ο πρώτος όρος μηδενίζεται για $\psi_0 + \nu \cos \psi_0 = 0$, λαμβάνουμε από εκεί την τιμή της γωνίας ακρόφασης. Ισχύει δε, ότι η απόσταση μεταξύ μέγιστου και του επόμενου ελάχιστου στο μοντέλο μας είναι $\delta = \pi - 2\psi_0$.

6.5 Παλινδρόμηση γραμμικών-κυκλικών μεταβλητών

Θα αναλύσουμε σε αυτό το κομμάτι μερικά μοντέλα για την περίπτωση δειγμάτων με γραμμική επεξηγηματική μεταβλητή X και κυκλική μεταβλητή απόκρισης Θ . Η μοντελοποίηση αυτής της εξάρτησης μπορεί να προκύψει εφόσον εξασφαλισθεί ότι είτε η μέση κατεύθυνση, είτε η συγκέντρωση, είτε και τα δύο μεγέθη της κυκλικής μεταβλητής εξαρτώνται από την X .

Στην πιο γενική περίπτωση όχι μόνο η δεσμευμένη μέση κατεύθυνση αλλά και η δεσμευμένη διασπορά της κυκλικής μεταβλητής Θ δεδομένου $X = x$ είναι γραμμική συνάρτηση του x . Έτσι υποθέτουμε τυχαίο δείγμα $(\Theta_1, X_1), \dots, (\Theta_n, X_n)$ και $(\theta_1, x_1), \dots, (\theta_n, x_n)$ τις αντίστοιχες παρατηρήσεις με τα θ_i ανεξάρτητα μεταξύ τους ώστε να ισχύει: $\Theta_i | X = x \sim CN(\mu_i, \kappa) \quad i = 1, \dots, n$.

Έστω ότι η δεσμευμένη μέση κατεύθυνση έχει την μορφή :

$$\mu_i = \mu + g(\beta x_i),$$

όπου $g(\cdot)$ καλείται συνάρτηση σύνδεσης και αντιστοιχεί την ευθεία των πραγματικών αριθμών στο διάστημα $(-\pi, \pi)$. Για την ορθή ερμηνεία της συνάρτησης σύνδεσης φυσικά πρέπει $g(0) = 0$, ενώ ένα παράδειγμα μια κατάλληλης συνάρτησης σε αυτή την κατηγορία είναι η:

$$g(x) = 2 \tan^{-1}(x).$$

Στην περίπτωση που έχουμε προσδιορίσει την κατάλληλη συνάρτηση σύνδεσης, τότε με την μέθοδο μέγιστης πιθανοφάνειας μπορούμε να λάβουμε εκτιμήσεις για τις σταθερές μ, β .

6.6 Παλινδρόμηση κυκλικών-κυκλικών μεταβλητών

Θα αναλύσουμε, τέλος, ένα γραμμικό μοντέλο παλινδρόμησης δύο κυκλικών μεταβλητών A, B . Έστω λοιπόν ότι το ζεύγος αυτών των μεταβλητών έχει από κοινού συνάρτηση κατανομής πιθανότητας την $f(\alpha, \beta) \forall \alpha \in (0, 2\pi]$ και $\beta \in (0, 2\pi]$.

Για να εκτιμήσουμε τη μέση τιμή της τυχαίας μεταβλητής B δεδομένου της A θα χρησιμοποιήσουμε την ακόλουθη εξίσωση που προκύπτει από την εξίσωση (2.1):

$$E_{B|A} (e^{i\beta} | A = \alpha) = \rho(\alpha)e^{i\mu(\alpha)} = g_1(\alpha) + ig_2(\alpha). \quad (6.9)$$

Εδώ η ποσότητα $\mu(\alpha)$ αναπαριστά τη δεσμευμένη μέση κατεύθυνση της $B | A = \alpha$ και αντίστοιχα η ποσότητα $\rho(\alpha)$ τη δεσμευμένη συγκέντρωση, με $\rho(\alpha) \in [0,1]$.

Ισοδύναμα από την εξίσωση (6.8) λαμβάνουμε:

$$\begin{cases} E_{B|A=\alpha} (\cos \beta | A = \alpha) = g_1(\alpha) \\ E_{B|A=\alpha} (\sin \beta | A = \alpha) = g_2(\alpha) \end{cases}, \quad (6.10)$$

και από την εξίσωση (6.9) την εκτιμήτρια για το $\mu(\alpha)$

$$\hat{\mu}(\alpha) = \arctan^* \left(\frac{g_1(\alpha)}{g_2(\alpha)} \right). \quad (6.11)$$

Επειδή η δομή των ποσοτήτων που ορίσαμε ως $g_1(\alpha)$, $g_2(\alpha)$ δεν είναι ξεκάθαρη η εκτίμηση που περιγράψαμε είναι δύσκολο να γίνει απευθείας από το δείγμα μας. Εκμεταλλευόμενοι το γεγονός ότι και οι δύο αυτές ποσότητες όμως είναι περιοδικές με περίοδο 2π μπορούμε να τις εκφράσουμε ως αναπτύγματα της σειράς Fourier, δηλαδή σε τριγωνομετρικά πολυώνυμα έστω m βαθμού. Έτσι λαμβάνουν τη μορφή:

$$\begin{aligned} g_1(\alpha) &\approx \sum_{k=0}^m (A_k \cos k\alpha + B_k \sin k\alpha) \\ g_2(\alpha) &\approx \sum_{k=0}^m (C_k \cos k\alpha + D_k \sin k\alpha) \end{aligned}, \quad (6.12)$$

όπου A_k, B_k, C_k, D_k σταθερές που για μικρές τιμές του βαθμού του πολυωνύμου μπορούν να προσδιοριστούν με την μέθοδο των ελαχίστων τετραγώνων.

Βάσει των παραπάνω εξισώσεων λοιπόν το γενικό γραμμικό μοντέλο που περιγράφει τις δύο κυκλικές μεταβλητές μας είναι της μορφής:

$$\begin{cases} \cos \beta \approx \sum_{k=0}^m (A_k \cos k\alpha + B_k \sin k\alpha) + \varepsilon_1 \\ \sin \beta \approx \sum_{k=0}^m (C_k \cos k\alpha + D_k \sin k\alpha) + \varepsilon_2 \end{cases},$$

όπου $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ είναι το διάνυσμα σφάλματος με τα $(\varepsilon_1, \varepsilon_2)$ να ακολουθούν κατανομή με μέσο διάνυσμα $\vec{0}$ και πίνακα διασποράς (Σ) τον πίνακα συνδιακύμανσης

$$Cov(\vec{\varepsilon}) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Κεφάλαιο 7

Στατιστικά πακέτα κυκλικών μεταβλητών στην R

7.1 Εισαγωγή

Σκοπός αυτού του κεφαλαίου είναι να παρατεθούν οι βασικές εντολές και επιλεγμένα παραδείγματα, βάσει των οποίων υλοποιείται η στατιστική ανάλυση κυκλικών μεταβλητών με την χρήση του στατιστικού πακέτου *R*.

7.2 «CircStats» βιβλιοθήκη

Το βασικό πακέτο εντολών που χρησιμοποιήθηκε στην παρούσα εργασία είναι το «CircStats». Θα αναλύσουμε λοιπόν τις βασικές εντολές που χρησιμοποιήσαμε για την πρακτική ανάλυση και την εποπτική παρουσίαση των κυκλικών μεταβλητών. Στις περισσότερες περιπτώσεις χρησιμοποιήθηκε τυχαίο δείγμα από τις διάφορες κυκλικές κατανομές, που καλείται βάσει εντολών του συγκεκριμένου πακέτου.

I. Μορφή, βασικά μέτρα θέσης-διασποράς και απεικόνιση κυκλικών μεταβλητών

- **rad(x)**

Εάν x διάνυσμα δεδομένων σε δεκαδικό σύστημα, η παραπάνω εντολή τα μετατρέπει σε ακτίνια.

- **deg(x)**

Εάν x διάνυσμα δεδομένων σε ακτίνια, η μετατροπή τους σε δεκαδικό σύστημα γίνεται με την συγκεκριμένη εντολή.

- **circ.mean(x)**

Επιστρέφει την μέση κατεύθυνση του διανύσματος κυκλικών δεδομένων.

- **circ.summary(x)**

Επιστρέφει τον αριθμό των παρατηρήσεων, το μήκος του δειγματικού συνιστάμενου διανύσματος $\left(\frac{R}{n}\right)$, καθώς και τη μέση κατεύθυνση του διανύσματος κυκλικών δεδομένων.

- **circ.range(x, test = «FALSE»)**

Υπολογίζει το κυκλικό εύρος, δηλαδή το μικρότερο τόξο που περιλαμβάνει όλα τα δεδομένα μας, και εκτελεί εάν ζητηθεί έναν έλεγχο ομοιομορφίας Rao .

- **circ.disp(x)**

Υπολογίζει το μέγεθος του δείγματος, το μήκος του συνιστάμενου διανύσματος, το μήκος του δειγματικού συνιστάμενου διανύσματος καθώς και την διασπορά $\left(1 - \left(\frac{R}{n}\right)\right)$ των τιμών ενός διανύσματος κατευθυνόμενων δεδομένων.

- **circ.plot(x, stack="boolean", bins= «#»)**

Απεικονίζει τις παρατηρήσεις που περιέχει το δείγμα μας ως σημεία στην περιφέρεια κύκλου. Στην περίπτωση που οι παρατηρήσεις μας είναι επαναλαμβανόμενες ή πολύ κοντά, δίνουμε στην παράμετρο stack την λογική τιμή true και ο κύκλος χωρίζεται σε τμήματα που ορίζουμε με την τελευταία παράμετρο εισόδου. Οι παρατηρήσεις μας πλέον «στιβάζονται» στο αντίστοιχο τμήμα που ανήκουν δίνοντάς μας μια πιο πλήρη εικόνα.

- **rose.diag(x, bins= «#»)**

Απεικονίζει τις παρατηρήσεις του δείγματος μας σε ένα ροδόγραμμα. Με την δεύτερη παράμετρο εισόδου καθορίζουμε τον αριθμό των τμημάτων στα οποία θα χωριστεί ο κύκλος ενώ αυτόματα το κάθε κομμάτι έχει ακτίνα ίση με την τετραγωνική ρίζα της σχετικής συχνότητας των παρατηρήσεων στο εν λόγω τμήμα. Αυτό μας δίνει μια πιο άμεση εικόνα για το πώς οι παρατηρήσεις μας είναι κατανομημένες.

- **trig.moment(x,p)**

Υπολογίζει τις τριγωνομετρικές ροπές, p τάξης, το $\left(\frac{R}{n}\right)$ καθώς και το πραγματικό και φανταστικό κομμάτι των ροπών.

Παράδειγμα 7.1:

#Ορίζουμε ένα διάνυσμα τιμών x

```
x <- c(58,79,34,120,21,357)
```

#Μετατρέπουμε τις τιμές σε ακτίνια

```
x <- -rad(x)
```

```
> x = [1.0122910, 1.3788101, 0.5934119, 2.0943951, 0.3665191, 6.2308254]
```

#Καλούμε τις εντολές για τα βασικά μέτρα θέσης-διασποράς όπως αναλύσαμε παραπάνω.

```
> circ.mean(x)
[1] 0.8736535
> circ.summary(x)
  n   mean.dir   rho
16  0.8736535  0.7741011
> list(trig.moment(x, 1), trig.moment(x, 2), trig.moment(x, 3))
[[1]]
  mu.p   rho.p   cos.p   sin.p
0.8736535 0.7741011 0.496996 0.5934876

[[2]]
  mu.p   rho.p   cos.p   sin.p
1.441611 0.3191866 0.04111972 0.3165269

[[3]]
  mu.p   rho.p   cos.p   sin.p
0.9535154 0.2000064 0.1157677 0.1630963
> circ.disp(x)
  n     r     rbar     var
16  4.644607 0.7741011 0.2258989
> circ.range(x, test = TRUE)
  range   p.value
2.146755 0.02793601
```

II. Συναρτήσεις πυκνότητας πιθανότητας και παραγωγή τυχαίων δειγμάτων από κατανομές κυκλικών μεταβλητών

- **dcard(theta,mu,r)**

Επιστρέφει τη συνάρτηση πυκνότητας πιθανότητας της καρδιοειδούς κατανομής εκτιμώμενη για το δοθέν διάνυσμα κυκλικών δεδομένων, με συγκεκριμένη μέση κατεύθυνση και παράμετρο συγκέντρωσης.

- **rcard(n,mu,r)**

Επιστρέφει τυχαίο δείγμα επιθυμητού μεγέθους από την καρδιοειδή κατανομή με δεδομένες μέση κατεύθυνση και παράμετρο συγκέντρωσης.

- **dtri(theta,r)**

Υπολογίζει τις τιμές της συνάρτησης πυκνότητας πιθανότητας της τριγωνικής κατανομής για συγκεκριμένο διάνυσμα παρατηρήσεων και δεδομένη τιμή συγκέντρωσης.

- **rtri(n,r)**

Επιστρέφει τυχαίο δείγμα που ακολουθεί την τριγωνική κατανομή με συγκεκριμένο μέγεθος και παράμετρο συγκέντρωσης.

- **dvm(theta,mu,k)**

Επιστρέφει τις τιμές της συνάρτησης κατανομής πιθανότητας για την VonMises κατανομή με τις συγκεκριμένες παραμέτρους για το διάνυσμα που ορίζουμε.

- **rvm(n,mu,k)**

Παράγει τυχαίο δείγμα μεγέθους n , που ακολουθεί την κυκλική κανονική κατανομή με συγκεκριμένη μέση κατεύθυνση και παράμετρο συγκέντρωσης.

- **dmixedvm(theta,mu1,mu2,k1,k2,rho)**

Υπολογίζει τις τιμές της συνάρτησης κατανομής πιθανότητας της μίξης δύο κυκλικών κανονικών κατανομών για συγκεκριμένο διάνυσμα παρατηρήσεων και δεδομένες παραμέτρους των δύο κατανομών και μίξης.

- **rmixedvm(n,mu1,mu2,k1,k2,rho)**

Δημιουργεί ένα διάνυσμα τιμών μεγέθους n , που προέρχονται από την μίξη δύο κυκλικών κανονικών κατανομών με συγκεκριμένες επιμέρους παραμέτρους και δεδομένη παράμετρο μίξης.

- **dwrpcauchy(theta,mu,k)**

Επιστρέφει τις τιμές της συνάρτησης κατανομής πιθανότητας για την περιελιγμένη Cauchy κατανομή με δεδομένες μέση κατεύθυνση και παράμετρο συγκέντρωσης στο συγκεκριμένο δείγμα.

- **rwrpcauchy(n,location=0,rho=exp(-1))**

Δημιουργείται ένα διάνυσμα συγκεκριμένου μεγέθους που ακολουθεί την περιελιγμένη Cauchy κατανομή με τις δεδομένες παραμέτρους. Τα αποτελέσματα επιστρέφονται ως υπόλοιπα της Ευκλείδειας διαίρεσης με το $2\pi \cdot (\text{mod}(2\pi))$

- **dwrpnorm(theta,mu,k, tol=1e-5)**

Επιστρέφει τις τιμές της συνάρτησης κατανομής πιθανότητας για την περιελιγμένη κανονική κατανομή στο εν λόγω διάνυσμα παρατηρήσεων, με δεδομένες παραμέτρους και επίπεδο ακρίβειας.

- **rwrpnorm(n,mu,rho)**

Παράγονται κατά τυχαίο τρόπο n τιμές που ακολουθούν την περιελιγμένη κανονική κατανομή με τις συγκεκριμένες παραμέτρους συγκέντρωσης και μέση κατεύθυνση.

- **plot.edf(theta)**

Αποτυπώνει σε διάγραμμα την εμπειρική συνάρτηση κατανομής του συγκεκριμένου συνόλου κυκλικών παρατηρήσεων.

Παράδειγμα 7.2:

#Επιστρέφει τις τιμές 10 τυχαίων αριθμών που ακολουθούν την κυκλική κανονική κατανομή με τις παραμέτρους που ορίσαμε.

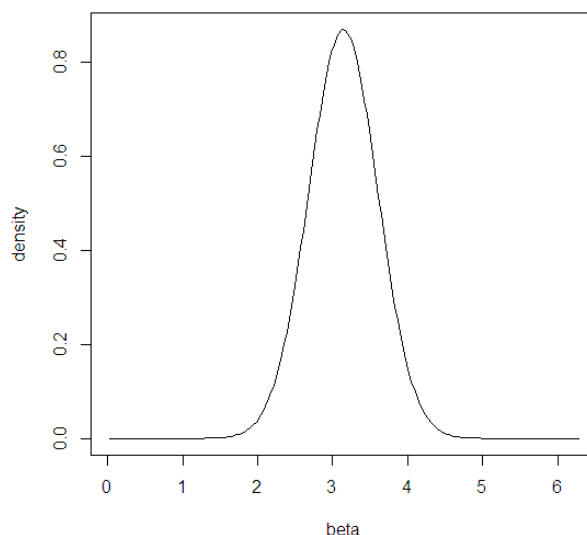
```
> rvm(10,rad(45),8)
[1] 1.4905725 1.1716173 0.7420650 1.1620762 0.2915993
[6] 1.2557957 0.9942196 0.8703553 1.0681816 1.6147593
```

#Δημιουργούμε ένα διάνυσμα στο οποίο θα προσδιορισθεί η συνάρτηση κατανομής πιθανότητας.

```
> beta <- -c(1:200) * 2 * pi/200
```

#Δημιουργούμε ένα κενό διάνυσμα ίδιου μήκους με το beta, και συμπληρώνουμε τις θέσεις του με την υπολογιζόμενη τιμή της συνάρτησης πυκνότητας πιθανότητας της περιελιγμένης Cauchy κατανομής για το διάνυσμα beta, με δεδομένες παραμέτρους.

```
> density <- -c(1:200)
> for(i in 1:200) density[i] <- -dwrpnorm(beta[i],pi,0.9)
> plot(beta,density,type = l)
```



Διάγραμμα 7-1:Γραφική απεικόνιση συνάρτησης πυκνότητας πιθανότητας της περιελιγμένης Cauchy κατανομής για διάνυσμα παρατηρήσεων

III. Εκτιμήσεις παραμέτρων και εκτιμήτριες μέγιστης πιθανοφάνειας

- **pvm(theta,mu,kappa,acc=1e-20)**

Εκτιμά τις τιμές της αθροιστικής πιθανότητας της κυκλικής κανονικής κατανομής με δεδομένη μέση κατεύθυνση, και παράμετρο συγκέντρωσης. Επιστρέφει δηλαδή την πιθανότητα η τυχαία μεταβλητή που ακολουθεί την VonMises να βρίσκεται ανάμεσα στο μηδέν και την τιμή theta.

- **vm.ml(theta)**

Υπολογίζει τις εκτιμήτριες μέγιστης πιθανοφάνειας των παραμέτρων της κυκλικής κανονικής κατανομής, με την βοήθεια του συγκεκριμένου διανύσματος τιμών (theta).

- **wrapcauchy.ml(theta,mu0,rho0,acc=1e-15)**

Υπολογίζει τις εκτιμήτριες μέγιστης πιθανοφάνειας των παραμέτρων της περιελιγμένης Cauchy κατανομής, για το συγκεκριμένου διάνυσμα τιμών, με αρχικές εκτιμήσεις για τις δύο παραμέτρους τις δωθείσες, καθώς και συγκεκριμένο επίπεδο ακρίβειας.

- **pp.plot(theta,ref.line=TRUE)**

Εμφανίζει σε διάγραμμα την εμπειρική κατανομή του συνόλου δεδομένων σε σχέση την πραγματική συνάρτηση κατανομής VonMises.

- **est.rho(theta)**

Εκτιμά το μέσο μήκος του συνισταμένου διανύσματος του δεδομένου συνόλου τιμών. Συγκεκριμένα κάθε παρατήρηση του διανύσματος αντιμετωπίζεται ως μοναδιαίο διάνυσμα ή σημείο στην περιφέρεια κύκλου. Υπολογίζεται λοιπόν το συνιστάμενο διάνυσμα R όπως ορίστηκε στην σχέση (1.4) και μας επιστρέφεται η ποσότητα $\frac{R}{n}$.

- **I.0(x)**

Επιστρέφει την μηδενικής τάξης συνάρτηση Bessel υπολογισμένη στον δοθέντα μη μηδενικό, πραγματικό αριθμό x .

- **I.1(x)**

Επιστρέφει την πρώτης τάξης συνάρτηση Bessel υπολογισμένη στον δοθέντα μη μηδενικό, πραγματικό αριθμό x .

- **I.p(p,x)**

Υπολογίζει την p -τάξης συνάρτηση Bessel στον αριθμό x .

- **A1(kappa)**

Υπολογίζει την πρώτη και μηδενικής τάξης συνάρτηση Bessel πρώτου είδους σε συγκεκριμένο μη αρνητικό πραγματικό αριθμό kappa και επιστρέφει τον λόγο $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ όπου κ ο δοθείς αριθμός kappa.

- **A1inv(x)**

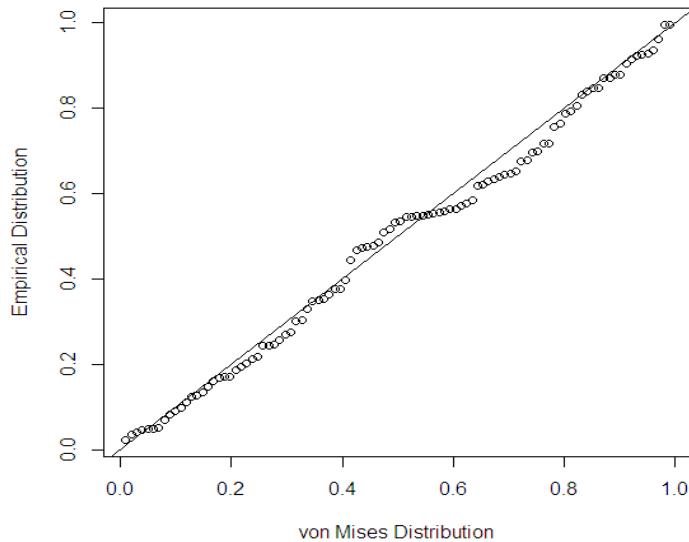
Υπολογίζει τον αντίστροφο λόγο της πρώτης και μηδενικής τάξης συνάρτησης Bessel πρώτου είδους για τιμές μεταξύ μηδέν και της μονάδας. Επιστρέφει δηλαδή την τιμή για την οποία ο παραπάνω λόγος ισούται με το δεδομένου εισόδου μας (x).

- **est.kappa(theta)**

Εκτελεί την εκτίμηση της παραμέτρου συγκέντρωσης με την μέθοδο μέγιστης πιθανοφάνειας μιας κυκλικής κανονικής κατανομής δεδομένου ενός συνόλου κυκλικών δεδομένων.

Παράδειγμα 7.3:

```
#Δημιουργεί εκατό ψευδο-τυχαία δεδομένα που ακολουθούν την VonMises
κατανομή με τις συγκεκριμένες παραμέτρους.
> theta <- rvm(100,4.5,0.75)
#Απεικονίζει σε διάγραμμα την εμπειρική κατανομή των δεδομένων μας, ενώ
με ευθεία γραμμή υποδηλώνεται η πραγματική συνάρτηση κυκλικής
κανονικής κατανομής της οποίας οι παράμετροι μας επιστρέφονται.
> pp.plot(theta, ref.line = TRUE)
      mu      kappa
5.031545  0.6079667
#Μας επιστρέφει την εκτίμηση του μήκους του συνισταμένου διανύσματος
του εν λόγω συνόλου παρατηρήσεων.
> est.rho(theta)
[1]0.2908186
#Οι δύο παρακάτω εντολές μας επιστρέφουν τις τιμές της εκτίμηση της
παραμέτρου συγκέντρωσης και της μέσης κατεύθυνσης για την κυκλική
κανονική κατανομή βάσει της μεθόδου της μέγιστης πιθανοφάνειας.
> est.kappa(theta)
[1] 0.6079667
> vm.ml(theta)
      mu      kappa
-1.25164  0.6079667
```



Διάγραμμα 7-2:Εμπειρική κατανομή δεδομένων σε σύγκριση με την πραγματική Von Mises Κατανομή

IV. Έλεγχοι υποθέσεων

- **r.test(x, degree=FALSE)**

Εκτελεί έναν έλεγχο ομοιογένειας Rayleigh, εκτιμώντας την σπουδαιότητα του μήκους του συνιστάμενου διανύσματος. Η εναλλακτική υπόθεση του ελέγχου είναι μια μονοκόρυφη κατανομή με άγνωστη μέση κατεύθυνση και μήκος R . Υπάρχει η δυνατότητα με το δεύτερο όρισμα να καθορίσουμε το σύστημα μονάδων στο οποίο δίδονται οι κυκλικές παρατηρήσεις (ακτίνια ή δεκαδικά).

- **v0.test(x,mu=0,degree=FALSE)**

Εκτελεί έναν έλεγχο ομοιογένειας Rayleigh εκτιμώντας την σημαντικότητα του μήκους του συνιστάμενου διανύσματος. Σε αυτή την περίπτωση, η εναλλακτική υπόθεση του ελέγχου είναι μια μονοκόρυφη κατανομή με γνωστή μέση κατεύθυνση που καθορίζεται από το όρισμα μ και άγνωστο μήκος του συνιστάμενου διανύσματος. Ομοίως με την προηγούμενη εντολή, καθορίζουμε το σύστημα μονάδων των παρατηρήσεων με την προεπιλεγμένη επιλογή να είναι μετρήσεις σε ακτίνια.

Παράδειγμα 7.4:

```
#Καλούμε ένα διάνυσμα τιμών x με μονάδα μέτρησης τα ακτίνια.  
> x  
[1] 1.0122910 1.3788101 0.5934119 2.0943951 0.3665191  
[6] 6.2308254  
# Εκτελούμε έναν έλεγχο ομοιογένειας Rayleigh και μας επιστρέφεται το  
μήκος του συνιστάμενου διανύσματος καθώς και η p-τιμή του ελέγχου.  
  
> r.test(x)  
  
$r.bar  
[1] 0.7741011  
  
$p.value  
[1] 0.01980831  
#Εκτελεί έναν έλεγχο ομοιογένειας Rayleigh με την μέση κατεύθυνση της  
εναλλακτικής υπόθεσης να καθορίζεται από το όρισμα mu.  
> v0.test(x, mu = 0)  
$r0.bar  
[1] 0.496996  
  
$p.value  
[1] 0.0426773
```

V. Κυκλική παλινδρόμηση-συσχέτιση

- **circ.cor(alpha,beta,test=FALSE)**

Υπολογίζει τον συντελεστή συσχέτισης όπως ορίστηκε στην σχέση (6.1) για τις δύο κυκλικές μεταβλητές που δίδονται στα δύο πρώτα ορίσματα. Βάσει του τελευταίου ορίσματος, υπάρχει η δυνατότητα εκτέλεσης και ελέγχου σημαντικότητας της συσχέτισης.

- **circ.reg(alpha,beta,order=1,level=0.05)**

Προσαρμόζει ένα μοντέλο παλινδρόμησης για μία ανεξάρτητη και μία εξαρτημένη κυκλική μεταβλητή. Ορίζεται η τάξη του πολυωνύμου που προσαρμόζεται και το επίπεδο σημαντικότητας.

Παράδειγμα 7.5:

Δημιουργεί δέκα ψευδο-τυχαίες κυκλικά παρατηρήσεις που ακολουθούν την Von Mises κατανομή με τις συγκεκριμένες παραμέτρους.

```
> alpha <- rvm(10,0,2)
```

#Δημιουργεί δέκα κυκλικά δεδομένα εξαρτημένα από το προηγούμενο διάνυσμα δεδομένων.

```
> beta <- -alpha + 2 * pi + rvm(10,0,5)
```

#Εκτελεί τον υπολογισμό του συντελεστή Pearson των δύο κυκλικών μεταβλητών και πραγματοποιεί έναν έλεγχο σημαντικότητας του.

```
> circ.cor(alpha, beta, test = TRUE)
```

<i>r</i>	<i>test.stat</i>	<i>p.value</i>
0.8014024	2.09801	0.03590423

#Προσομοιώνουμε ένα μοντέλο παλινδρόμησης της ανεξάρτητης και εξαρτημένης κυκλικής μεταβλητής που ορίσαμε παραπάνω, ορίζοντας την τάξη του πολυωνύμου που ζητάμε.

```
> circ.lm <- -circ.reg(alpha, beta, order = 1)
```

#Μας επιστρέφει την τετραγωνική ρίζα του μέσου όρου των τετραγώνων των εκτιμώμενων παραμέτρων συγκέντρωσης του *beta* δεδομένου του *alpha*.

```
> circ.lm$rho
```

0.9006875

#Μας επιστρέφει τις προσαρμοσμένες τιμές του μοντέλου μας. Αυτές υπολογίζονται βάσει της αντίστροφης εφαπτομένης του λόγου των προβλεπόμενων τιμών του ημιτόνου του *beta* προς το συνημιτόνου του *beta*.

```
> circ.lm$fitted
```

1	2	3	4	5
6.264284312	6.162292834	5.655813154	0.001523732	2.354622164
6	7	8	9	10
4.824226634	0.697795518	0.529576545	5.867894238	5.294819583

```

#Μας επιστρέφει τα υπόλοιπα του μοντέλου μας.
> circ.lm$residuals
      1          2          3          4          5
6.06754568  6.21824491  0.31318904  6.26244757  0.10287553
      6          7          8          9         10
5.92908471  0.28032211  0.05235598  5.36393810  0.92251114

#Μας επιστρέφει τον πίνακα με τις εκτιμώμενους συντελεστές του μοντέλου. Η πρώτη
στήλη αφορά τους συντελεστές του μοντέλου που προβλέπουν την τιμή του
συνημιτόνου του beta ενώ η δεύτερη αντίστοιχα του ημιτόνου του beta.
> circ.lm$coef
              [1]          [2]
(Intercept) - 0.18223618    0.0444936
cos.alpha    1.13564063   - 0.1695310
sin.alpha    0.06202078    0.7642204

#Μας επιστρέφει τις p-τιμές του ελέγχου κατά πόσο οι τριγωνομετρικοί όροι
διαφέρουν από το μηδέν.
> circ.lm$pvalues
      [1]      [2]
0.7028406  0.2226162

#Μας επιστρέφει τη μέση τιμή των συνημιτόνων των υπολοίπων.
> circ.lm$A.k
0.902838

#Εκτιμώντας ότι τα υπόλοιπα του μοντέλου μας ακολουθούν την κυκλική κανονική
κατανομή υπολογίζεται η εκτίμηση, βάσει της μεθόδου της μέγιστης πιθανοφάνειας,
του συντελεστή συγκέντρωσης.
> circ.lm$kappa
5.435777

```

Κεφάλαιο 8

Ανάλυση δείγματος κυκλικών δεδομένων και συμπεράσματα

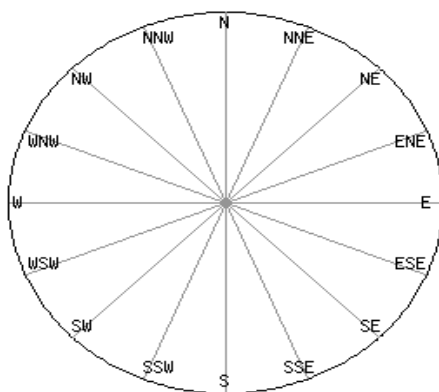
8.1 Εισαγωγή

Στο κεφάλαιο αυτό ακολουθεί μια πρακτική εφαρμογή της στατιστικής ανάλυσης κυκλικών δεδομένων, μέσω των εντολών της R που αναπτύχθηκαν στο έβδομο κεφάλαιο, σε δείγμα πραγματικών δεδομένων για την εξαγωγή συμπερασμάτων. Συγκεκριμένα θα χρησιμοποιήσουμε δεδομένα κατεύθυνσης του ανέμου, για το χρονικό διάστημα του μηνός Σεπτεμβρίου του έτους 2014. Τα δεδομένα ελήφθησαν από την ηλεκτρονική σελίδα <http://www.wunderground.com/>, και καταγράφηκαν από τον μετεωρολογικό σταθμό με κωδικό IU0391U028, στην τοποθεσία του Βύρωνα Αττικής.

8.2 Μορφή δεδομένων

Τα δεδομένα που θα χρησιμοποιήσουμε στην ανάλυση μας αφορούν την ημερήσια κατεύθυνση του ανέμου για ολόκληρο τον μήνα Σεπτέμβριο για δύο συγκεκριμένες χρονικές στιγμές, αυτές των 10:00 πμ και 18:00 μμ.

Τα συγκεκριμένα μετεωρολογικά δεδομένα εμφανίζουν την ιδιαιτερότητα της μέτρησης βάσει ενός ιδιαίτερου συστήματος συντεταγμένων. Συγκεκριμένα, όπως φαίνεται και στα διαγράμματα 8-1(α) και 8-1(β), λαμβάνουν τιμές στο διάστημα $[0^\circ, 360^\circ]$ με την αρχή και το τέλος του διαστήματος να συμπίπτουν και να ταυτίζονται με την κατεύθυνση του μαγνητικού βορρά.



Διάγραμμα 8-1(α): Κατευθύνσεις ανέμου στον τριγωνομετρικό κύκλο

Συνομογραφία	Κατεύθυνση Ανέμου	Μοίρες
N	Βόρεια	0°
NNE	Βόρεια-Βορειοανατολικά	22.5°
NE	Βορειοανατολικά	45°
ENE	Ανατολικά-Βορειοανατολικά	67.5°
E	Ανατολικά	90°
ESE	Ανατολικά-Νοτιοανατολικά	112.5°
SE	Νοτιοανατολικά	135°
SSE	Νότια-Νοτιοανατολικά	157.5°
S	Νότια	180°
SSW	Νότια-Νοτιοδυτικά	202.5°
SW	Νοτιοδυτικά	225°
WSW	Δυτικά-Νοτιοδυτικά	247.5°
W	Δυτικά	270°
WNW	Δυτικά-Βορειοδυτικά	292.5°
NW	Βορειοδυτικά	315°
NNW	Βόρεια-Βορειοδυτικά	337.5°

Διάγραμμα 8-1(β): Επεξηγηματικός πίνακας κατευθύνσεων ανέμων

Όπως παρατηρούμε όχι μόνο διαφέρει η αρχή μέτρησης των παρατηρήσεων μας σε σχέση με τις συνηθισμένες μεταβλητές που μετριοούνται σε μοίρες αλλά και η θετική φορά είναι ακριβώς η αντίθετη. Η συγκεκριμένη ιδιαιτερότητα μας επιτρέπει να επαληθεύσουμε μία από της αρχικές μας εκτιμήσεις για την στατιστική ανάλυση των κυκλικών δεδομένων, αυτή της μη εξάρτησης του συνιστάμενου διανύσματος και της μέσης κατεύθυνσης από την αρχή μέτρησης των παρατηρήσεων μας αλλά και από την παραδοχή της θετικής φοράς.

8.2 Ανάλυση του δείγματος και εξαγωγή συμπερασμάτων

Στον παρακάτω κώδικα ορίζουμε τα δύο διανύσματα με τις παρατηρήσεις μας στις δύο διαφορετικές χρονικές στιγμές ημερησίως.

```
> x_10<-
c(292,180,201,180,338,22,292,292,315,315,338,135,292,248,338,248,45,0,68,292,24
8,180,202,90,135,315,315,315,68,90)

> x_18<-
c(270,225,180,225,225,338,135,112,180,0,135,315,0,202,45,135,315,338,225,180,18
0,270,225,315,225,22,112,68,292,315)
```

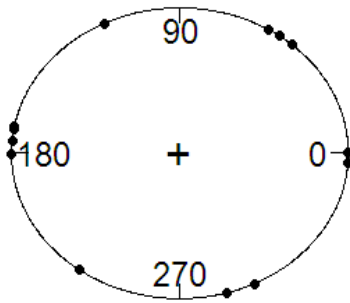
Εν συνεχεία δημιουργούμε ένα κυκλικό διάγραμμα (Διάγραμμα 8-2) για τα δύο σύνολα παρατηρήσεων μας ξεχωριστά χρησιμοποιώντας τις παρακάτω εντολές:

```

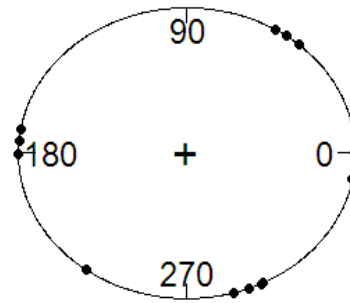
> par(mfrow=c(1,2))
> circ.plot(x_10,main="Παρατηρήσεις στις 10 πμ")
> circ.plot(x_18,main="Παρατηρήσεις στις 6 μμ")

```

Παρατηρήσεις στις 10 πμ



Παρατηρήσεις στις 6 μμ



Διάγραμμα 8-2:Κυκλικά διαγράμματα κατευθύνσεων του ανέμου

Με τις εξής εντολές μετατρέπουμε τις τιμές των δεδομένων μας σε ακτίνια.

```

> x_10<-rad(x_10)
> x_18<-rad(x_18)

```

ενώ για να υπολογίσουμε το μήκος του συνισταμένου διανύσματος και την μέση κατεύθυνση για κάθε ένα από τα δύο σύνολα παρατηρήσεων χρησιμοποιούμε των παρακάτω κώδικα:

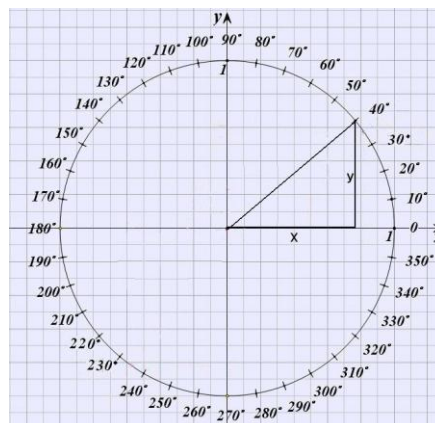
<pre> > circ.summary(x_10) n mean.dir rho 30 -0.9970214 0.2563729 > deg(circ.mean(x_10)) -57.12512 </pre>	<pre> > circ.summary(x_18) n mean.dir rho 30 -2.095552 0.1978742 > deg(circ.mean(x_18)) -120.0663 </pre>
--	---

Με βάση τα αποτελέσματα των εντολών που δώσαμε παρατηρούμε ότι φυσικά οι παρατηρήσεις στις δύο διαφορετικές χρονικές στιγμές είναι ισοπληθικές ($n = 30$). Μας επιστρέφεται επίσης για τα δύο διανύσματα που ορίσαμε η ποσότητα ρ που αντιστοιχεί στην ποσότητα του μήκους του συνισταμένου διανύσματος διαιρεμένου με το πλήθος των παρατηρήσεων καθώς και η μέση κατεύθυνση του δείγματος $mean.dir$.

Μετατρέποντας λοιπόν την μέση κατεύθυνση από ακτίνια σε βαθμούς διαπιστώνουμε ότι η μέση κατεύθυνση του ανέμου στις 10 π.μ., βάσει του διαγράμματος 8-1(β), είναι στο τμήμα Δυτική-Βορειοδυτική και Βορειοδυτική (302.87488°) ενώ η αντίστοιχη κατεύθυνση στις 18 μ.μ. είναι στο τμήμα Νοτιοδυτική και Δυτική-Νοτιοδυτική ($239,9337^\circ$).

Στην συνέχεια ορίζουμε δύο νέα διανύσματα με τιμές τις παρατηρήσεις μας ως είχαν, αλλά χρησιμοποιώντας το σύνθετες τριγωνομετρικό σύστημα συντεταγμένων και αφού τις μετατρέψουμε σε ακτίνια καλούμε τις εντολές για τον υπολογισμό του συνιστάμενου διανύσματος και της μέσης κατεύθυνσης.

<pre>> xn10 <- c(158,270,249,270,112,68,158,158,135,135,112,315,158,202,112,202,45,90,22,158,2 02,270,248,0,315,135,135,135,22,0) >xn10<-rad(xn10) > xn18<- c(180,225,270,225,225,112,315,338,270,90,315,135,90,248,45,315,135,112,225,270, 270,180,225,135,225,68,338,22,158,135) >xn18<-rad(xn18)</pre>	
<pre>> circ.summary(xn10) n mean.dir rho 30 2.567818 0.2563729</pre>	<pre>> deg(circ.mean(xn10)) 147.1251</pre>
<pre>> circ.summary(xn18) n mean.dir rho 30 -2.616837 0.1978742</pre>	<pre>> deg(circ.mean(xn18)) -149.9337</pre>



Διάγραμμα 8-3: Τριγωνομετρικός κύκλος με παρατηρήσεις στο «κλασικό» σύστημα συντεταγμένων

Λαμβάνουμε λοιπό τα αντίστοιχα αποτελέσματα για την μέση κατεύθυνση και το συνιστάμενο διάνυσμα για τις τροποποιημένες παρατηρήσεις. Μετατρέποντας την μονάδα μέτρησής τους σε μοίρες παρατηρούμε ότι η μέση κατεύθυνση του ανέμου στην πρωινή και στην απογευματινή μέτρηση είναι $147,1251^\circ$ και $210,0663^\circ$ αντίστοιχα. Κάνοντας χρήση του διαγράμματος 8-3 παρατηρούμε ότι οι κατευθύνσεις αυτές συμπίπτουν με αυτές που προέκυψαν για τις αρχικές μετρήσεις μας. Αυτό επιβεβαιώνει το γεγονός ότι για την σωστή στατιστική ερμηνεία των κυκλικών δεδομένων χρησιμοποιούμε την ποσότητα του συνιστάμενου διανύσματος, η οποία είναι ανεξάρτητη της επιλογής της μηδενικής κατεύθυνσης και της θετικής φοράς μέτρησης των δεδομένων.

Στο επόμενο βήμα της ανάλυσής μας υπολογίζουμε το κυκλικό εύρος για τα δύο διανύσματα παρατηρήσεων.

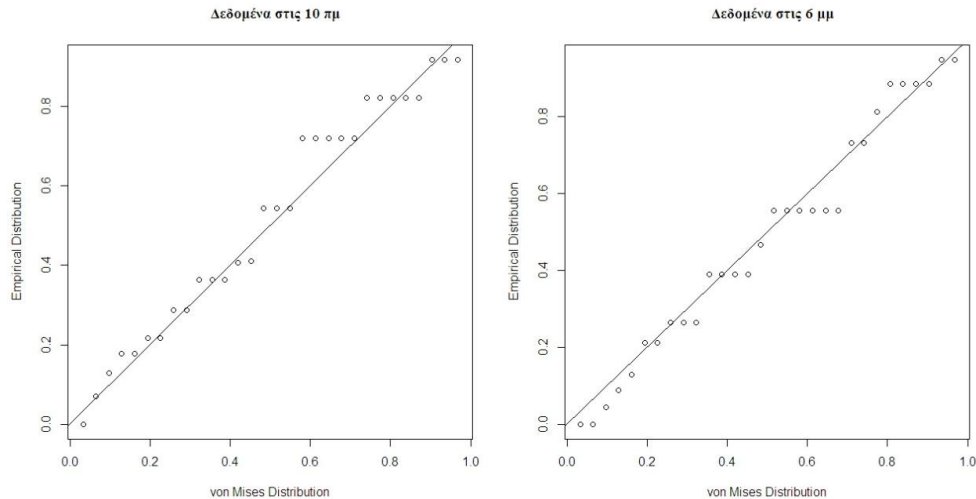
<code>> circ.range(x_10)</code>	<code>> deg(circ.range(x_10))</code>
range	range
5.480334	314
<code>> circ.range(x_18)</code>	<code>> deg(circ.range(x_18))</code>
range	range
5.497787	315

Από τα αποτελέσματα των εντολών που εκτελέσαμε παραπάνω παρατηρούμε ότι τα δύο διανύσματα έχουν πολύ μεγάλο εύρος παρατηρήσεων. Συγκεκριμένα το ελάχιστο κυκλικό τμήμα που περιλαμβάνει όλες τις παρατηρήσεις για κάθε διάνυσμα προσεγγίζει το σύνολο του κυκλικού δίσκου. Αυτό αποτελεί μια ένδειξη του ότι οι παρατηρήσεις μας ακολουθούν την κυκλική ομοιόμορφη κατανομή.

Για να διαπιστώσουμε εάν η υπόθεση αυτή ευσταθεί απεικονίζουμε στο Διάγραμμα 8-4 την εμπειρική κατανομή των δεδομένων μας σε σχέση με την πραγματική συνάρτηση κυκλικής κανονικής κατανομής με την χρήση των παρακάτω εντολών:

<code>> par(mfrow=c(1,2))</code>	
<code>> pp.plot(x_10,ref.line=TRUE)</code>	<code>> pp.plot(x_18,ref.line=TRUE)</code>
mu kappa	mu kappa
5.286164 0.5305194	4.187633 0.4037489

Στα αποτελέσματα μας επιστρέφονται για τα δύο διανύσματα παρατηρήσεων, οι παράμετροι της μέσης κατεύθυνσης μ και συγκέντρωσης κ της κυκλικής κανονικής κατανομής, οι οποίες υπολογίζονται με την μέθοδο της μέγιστης πιθανοφάνειας και χρησιμοποιούνται στην απεικόνιση της Von Mises.



Διάγραμμα 8-4: Εμπειρική κατανομή δεδομένων σε σύγκριση με την πραγματική Von Mises κατανομή

Από το Διάγραμμα 8.4 παρατηρούμε ότι τα δεδομένα μας δεν προσαρμόζονται ικανοποιητικά από την κυκλική κανονική κατανομή.

Για να επιβεβαιώσουμε λοιπόν υπόθεση ότι τα δεδομένα μας και για τις δύο χρονικές στιγμές προέρχονται από την κυκλική ομοιόμορφη κατανομή, εκτελούμε με τις παρακάτω εντολές τους ελέγχους $r.test$ και $v0.test$, με κοινή μηδενική υπόθεση H_0 : τα δεδομένα ακολουθούν την κυκλική ομοιόμορφη κατανομή. Η εναλλακτική υπόθεση του ελέγχου $r.test$ είναι η H_1 : Τα δεδομένα ακολουθούν μια μονοκόρυφη κατανομή με άγνωστες παραμέτρους, ενώ η αντίστοιχη υπόθεση του ελέγχου $v0.test$ είναι η H_1 : Τα δεδομένα ακολουθούν μια μονοκόρυφη κατανομή με συγκεκριμένη μέση κατεύθυνση.

<code>> r.test(x_10)</code>	<code>> r.test(x_18)</code>
<code>\$r.bar</code>	<code>\$r.bar</code>
<code>[1] 0.2563729</code>	<code>[1] 0.1978742</code>
<code>\$p.value</code>	<code>\$p.value</code>
<code>[1] 0.1392792</code>	<code>[1] 0.3114874</code>
<code>> v0.test(x_10,mu=0)</code>	<code>> v0.test(x_18,mu=0)</code>
<code>\$r.bar</code>	<code>\$r.bar</code>
<code>[1] 0.1391608</code>	<code>[1] -0.09913531</code>
<code>\$p.value</code>	<code>\$p.value</code>
<code>[1] 0.1414649</code>	<code>[1] 0.7775724</code>

Από τα αποτελέσματα των παραπάνω ελέγχων λαμβάνουμε τις ποσότητες $r.bar$ και $r0.bar$ που αντιστοιχούν στο μήκος του μέσου συνιστάμενου διανύσματος $\left(\frac{R}{n}\right)$.

Παρατηρούμε ότι η ποσότητα $r.bar$ συμπίπτει με την ποσότητα rho καθώς υπολογίζονται με βάση τον τύπο (3.10), ενώ για τον υπολογισμό της ποσότητας $r0.bar$ χρησιμοποιούνται οι προσεγγίσεις της παραγράφου (3.3.1). Τέλος προκύπτει ότι για όλους τους ελέγχους μας επιστρέφονται μεγάλες p - τιμές και κατά συνέπεια αποδεχόμαστε την μηδενική υπόθεση ότι τα δεδομένα μας προέρχονται από την κυκλική ομοιόμορφη κατανομή.

Στην συνέχεια της ανάλυσης των δεδομένων μέσω της παρακάτω εντολής υπολογίζουμε τον κυκλικό συντελεστή συσχέτισης των δύο συνόλων παρατηρήσεων όπως ορίστηκε στην σχέση (6.3) και εκτελούμε τον έλεγχο με μηδενική υπόθεση $H_0 : r = 0$ και εναλλακτική υπόθεση $H_1 : r \neq 0$.

> circ.cor(x_10,x_18,T)		
r	test.stat	p.value
0.4810172	2.651775	0.00800698

Παρατηρούμε λοιπόν ότι η συσχέτιση μεταξύ της κατεύθυνσης του ανέμου στις πρωινές μετρήσεις και αυτής στις απογευματινές μετρήσεις είναι αρκετά μεγάλη ($r = 0,48$). Από την p - τιμή του ελέγχου μάλιστα μπορούμε με βεβαιότητα να απορρίψουμε την υπόθεση ότι δεν υπάρχει συσχέτιση μεταξύ των δύο διανυσμάτων.

Τέλος προσαρμόζουμε ένα μοντέλο γραμμικής παλινδρόμησης της κατεύθυνσης του ανέμου στην απογευματινή μέτρηση σε σχέση με την κατεύθυνση του ανέμου της πρωινής μέτρησης και βλέπουμε κάποια βασικά χαρακτηριστικά του μέσω των παρακάτω εντολών:

> circ.lm<-circ.reg(x_18,x_10,order=1,level=0.05)		
> circ.lm\$coef		
	[,1]	[,2]
(Intercept)	0.2608106	-0.3011678
cos.alpha	0.4031944	0.2521104
sin.alpha	0.4769590	-0.6472663
> circ.lm\$pvalues		> circ.lm\$rho
[,1]	[,2]	[,1]
0.9076908	0.2497873	0.6711724.

Από τα αποτελέσματα της εντολής *circ.lm\$coef* παρατηρούμε τους συντελεστές $A_0, C_0, A_1, B_1, C_1, D_1$ του συστήματος εξισώσεων που δίνεται στην σχέση (6.12). Ακολουθώντας μας δίνονται οι p-τιμές του ελέγχου που εξετάζει ένα οι δύο τριγωνομετρικοί όροι είναι της τάξης του μηδενός. Όπως παρατηρούμε για το ημίτονο της κατεύθυνσης του ανέμου στις απογευματινές ώρες λαμβάνουμε αποτελέσματα που είναι της τάξης του μηδενός πράγμα που αποτελεί ένα μειονέκτημα για το μοντέλο που προσαρμόσαμε. Παρόλα αυτά δεν μπορούμε να βγάλουμε ασφαλή συμπεράσματα ως προς την καταλληλότητά του καθώς, όπως φαίνεται από το αποτέλεσμα της τελευταίας εντολής, έχει ικανοποιητική προσαρμογή σε περίπου 67% των παρατηρήσεων μας το οποίο είναι ένα πάρα πολύ καλό αποτέλεσμα ως προς την ισχύ του μοντέλου παλινδρόμησης.

Παράρτημα Α

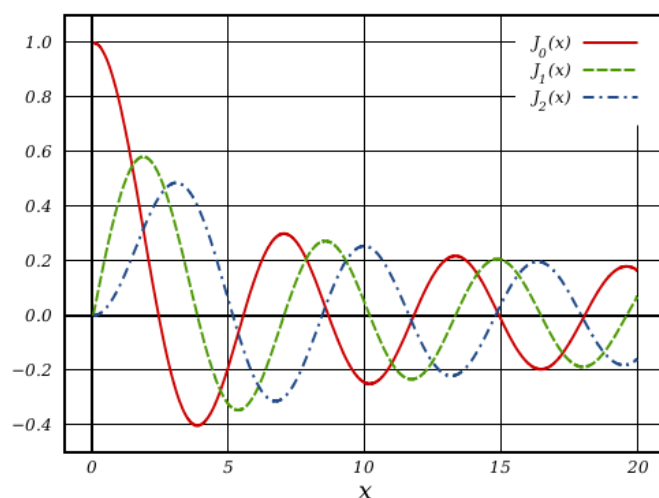
Συναρτήσεις Bessel

A.1 Ορισμός συναρτήσεων Bessel

Στην πορεία της παρούσας εργασίας κάνουμε εκτεταμένη χρήση των λεγόμενων συναρτήσεων Bessel και των ιδιοτήτων τους. Είναι, λοιπόν, απαραίτητο να παραθέσουμε μερικά στοιχεία για αυτές, καθώς και κάποιες επιλεγμένες αποδείξεις των ιδιοτήτων τους.

Οι συναρτήσεις Bessel ν -τάξης προκύπτουν ως λύση της ακόλουθης διαφορικής εξίσωσης: $\frac{d^2 x}{dz^2} + \frac{1}{z} \frac{dx}{dz} + \left(1 - \frac{\nu^2}{z^2}\right) x = 0$. Έχει τις εξής ισοδύναμες μορφές:

$$\left\{ \begin{array}{l} J_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\nu + \frac{1}{2}\right)} \int_0^\pi e^{iz \cos \theta} \sin^{2\nu} \theta d\theta, \nu > -\frac{1}{2} \\ J_\nu(z) = \sum_{r=0}^{\infty} (-1)^r \frac{\left(\frac{z}{2}\right)^{\nu+2r}}{r! \Gamma(\nu + r + 1)} \end{array} \right. \quad (\text{A.1.1})$$



Διάγραμμα A.1: Συναρτήση Bessel για διαφορετικές τιμές του ν

Για n θετικό ακέραιο ισχύει συγκεκριμένα $J_n(z) = \frac{1}{\pi} \int_0^\pi \cos(n\theta - z \sin \theta) d\theta$ και

$$\frac{d \{ z^\nu J_\nu(z) \}}{dz} = z^\nu J_{\nu+1}(z), \quad \nu > 0.$$

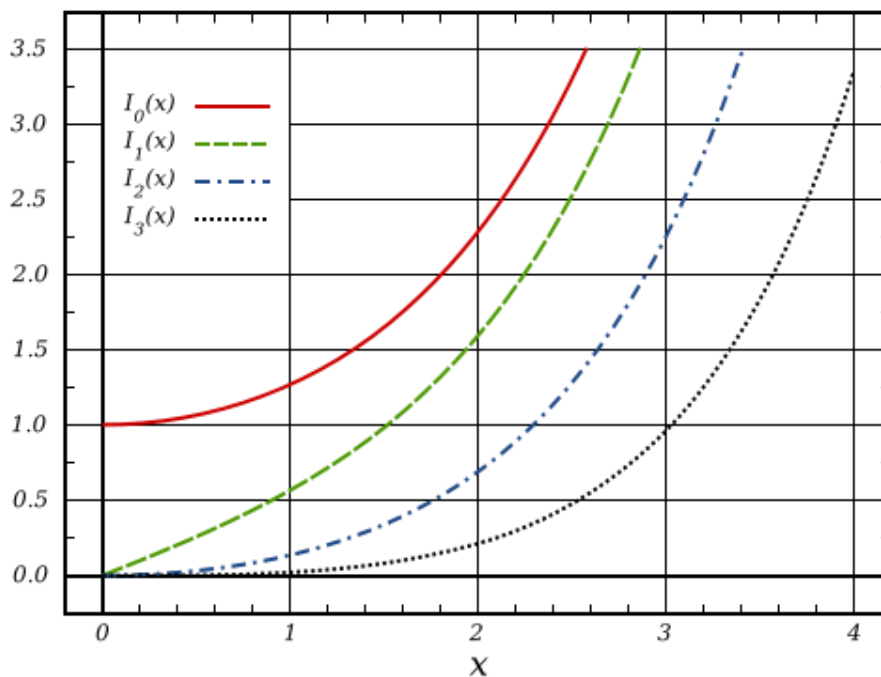
A.2 Τροποποιημένη συνάρτηση Bessel πρώτου είδους

Ως επί το πλείστον, χρησιμοποιούμε στην ανάλυση της κυκλικής κανονικής κατανομής την λεγόμενη τροποποιημένη συνάρτηση Bessel πρώτου είδους και τάξης ν , που σχετίζεται με την συνάρτηση Bessel που ορίζεται στην (A.1.1), ως εξής:

$$I_\nu(z) = i^{-\nu} J_\nu(iz) = \frac{1}{2\pi} \int_0^{2\pi} \cos \nu \theta \cdot e^{z \cos \theta} d\theta = \sum_{r=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{\nu+2r}}{r! \Gamma(\nu+r+1)}$$

Βάσει αυτού, βρίσκουμε την σταθερά της κυκλικής κανονικής κατανομής:

$$I_0(\rho) = \frac{1}{2\pi} \int_0^{2\pi} e^{\rho \cos \alpha} d\alpha.$$



Διάγραμμα A.2: Τροποποιημένη συνάρτηση Bessel πρώτου είδους για διαφορετικές τιμές του ν

Η εν λόγω συνάρτηση έχει μια σειρά από χρήσιμες ιδιότητες:

1. $I_{-ν}(z) = I_ν(z)$
2. $I_{ν-1}(z) + I_{ν+1}(z) = 2I'_ν(z)$
3. $I_{ν-1}(z) - I_{ν+1}(z) = \frac{2ν}{z}I_ν(z)$
4. $\frac{dI_0(z)}{dz} = I'_0(z) = I_1(z)$
5. $I''_0(z) = I'_1(z) = \frac{1}{2}(I_0(z) + I_2(z)) = I_2(z) + \frac{I_1(z)}{z} = I_0(z) - \frac{I_1(z)}{z}$

Θα αποδείξουμε βάσει των παραπάνω ιδιοτήτων, δύο βασικές σχέσεις που χρησιμοποιήσαμε και στην ροή της εργασίας:

$$\triangleright \boxed{\frac{d}{d\kappa}(\kappa I_1(\kappa)) = \kappa I_0(\kappa)} \quad (\text{A.2.1})$$

$$\begin{aligned} \frac{d}{d\kappa}(\kappa I_1(\kappa)) &= I_1(\kappa) + \kappa I'_1(\kappa) = I_1(\kappa) + \kappa \left(I_0(\kappa) - \frac{I_1(\kappa)}{\kappa} \right) = \\ &= I_1(\kappa) + \kappa I_0(\kappa) - I_1(\kappa) = \kappa I_0(\kappa) \end{aligned}$$

$$\triangleright \boxed{\frac{d}{d\kappa} \left(\frac{I_1(\kappa)}{\kappa} \right) = \frac{I_2(\kappa)}{\kappa}}$$

$$\begin{aligned} \frac{d}{d\kappa} \left(\frac{I_1(\kappa)}{\kappa} \right) &= \left(-\frac{I_1(\kappa)}{\kappa^2} \right) + \frac{I'_1(\kappa)}{\kappa} = \frac{1}{\kappa} \left(I'_1(\kappa) - \frac{I_1(\kappa)}{\kappa} \right) = \\ &= \frac{1}{\kappa} \left[\left\{ I_2(\kappa) + \frac{I_1(\kappa)}{\kappa} \right\} - \frac{I_1(\kappa)}{\kappa} \right] = \frac{I_2(\kappa)}{\kappa}. \end{aligned}$$

$$\triangleright \text{Αν } A(\kappa) \equiv \frac{I_1(\kappa)}{I_0(\kappa)} \text{ τότε } \boxed{A'(\kappa) = 1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)}.$$

$$\begin{aligned} A'(\kappa) &= \left(\frac{I_1(\kappa)}{I_0(\kappa)} \right)' = \frac{I'_1(\kappa)}{I_0(\kappa)} - \left(\frac{I_1(\kappa)}{I_0^2(\kappa)} \right) I'_0(\kappa) = \\ &= \frac{1}{I_0(\kappa)} \left(I_0(\kappa) - \frac{I_1(\kappa)}{\kappa} \right) - \left(\frac{I_1(\kappa)}{I_0^2(\kappa)} \right) I_1(\kappa) = \\ &= 1 - \frac{1}{\kappa} \left(\frac{I_1(\kappa)}{I_0(\kappa)} \right) - \left(\frac{I_1^2(\kappa)}{I_0^2(\kappa)} \right) = 1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa). \end{aligned}$$

Τέλος, θα παραθέσουμε μερικές προσεγγίσεις για μεγάλες και μικρές αντίστοιχα τιμές της παραμέτρου κ .

Για μεγάλες τιμές λοιπόν έχουμε:

$$I_p(\kappa) = \frac{e^\kappa}{\sqrt{2\pi\kappa}} \left\{ 1 - \frac{(4p^2 - 1)}{8\kappa} + \frac{(4p^2 - 1)(4p^2 - 9)}{2(8\kappa)^2} - \frac{(4p^2 - 1)(4p^2 - 9)(4p^2 - 25)}{6(8\kappa)^3} + \dots \right\},$$

$$\text{συνεπώς ισχύει: } \begin{cases} I_0(\kappa) \approx \frac{e^\kappa}{\sqrt{2\pi\kappa}} \\ A(\kappa) \approx \left(1 - \frac{1}{2\kappa} + \frac{1}{8\kappa^2} - \dots\right) \end{cases}.$$

Αντίστοιχα για μικρές τιμές του κ ισχύουν οι ακόλουθες προσεγγίσεις:

$$\begin{cases} I_0(\kappa) \approx 1 + \frac{\kappa^2}{4} + \frac{\kappa^4}{64} + \dots \\ A(\kappa) \approx \frac{\kappa}{2} \left(1 - \frac{\kappa^2}{8} + \frac{\kappa^4}{48} - \dots\right) \end{cases}.$$

A.3 Βασικά θεωρήματα συναρτήσεων Bessel

Στο σημείο αυτό κρίνεται σκόπιμο να παρατεθούν δύο πολύ βασικά θεωρήματα με συνεχή εφαρμογές, των τροποποιημένων συναρτήσεων Bessel πρώτου είδους.

- **Θεώρημα παράγοντα ασυνέχειας του Webber**

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} J_0\left(t\sqrt{a^2 + b^2 - 2ab \cos \theta}\right) d\theta = J_0(at) J_0(bt) \quad (\text{A.3.1})$$

$$\int_0^{\infty} J_0(at) J_0(bt) dt = \begin{cases} 0, & a > b \\ \frac{1}{2b}, & a = b \\ \frac{1}{b}, & a < b \end{cases} \quad (\text{A.3.2})$$

- **Προσθετική Φόρμουλα του Naumann**

$$J_0\left(\sqrt{a^2 + b^2 - 2ab \cos \theta}\right) = J_0(a) J_0(b) + 2 \sum_{k=1}^{\infty} J_k(a) J_k(b) \cos k\theta \quad (\text{A.3.3})$$

Τα παραπάνω θεωρήματα έχουν ιδιαίτερη σημασία, καθώς απλοποιούν ιδιαίτερα τους υπολογισμούς των κατανομών των βασικών στατιστικών ποσοτήτων που περιγράφουν τις κυκλικές μεταβλητές.

Θα παραθέσουμε λοιπόν την πλήρη απόδειξη της σχέσης (3.1) που περιγράφει την συνάρτηση πυκνότητας πιθανότητας του συνιστάμενου διανύσματος \vec{R} στην περίπτωση που το δείγμα των παρατηρήσεων μας προέρχεται από την ομοιόμορφη κυκλική κατανομή.

Απόδειξη:

Έστω $\left| \vec{R}_k \right| \equiv R_k$ το μήκος του συνιστάμενου διανύσματος των πρώτων k παρατηρήσεων $(\alpha_1, \alpha_2, \dots, \alpha_k)$, σε σύνολο n παρατηρήσεων. Ας ορίσουμε ως θ_k την γωνία που σχηματίζει το R_k και α_{k+1} . Σε αυτή την περίπτωση, είναι προφανές, ότι το μήκος R_k είναι συνάρτηση των αντίστοιχων γωνιών $\theta_1, \theta_2, \dots, \theta_{k-1}$, οι οποίες είναι τιμές της τυχαίας μεταβλητής Θ η οποία ακολουθεί ομοιόμορφη κατανομή στο διάστημα $(-\pi, \pi)$.

Η αθροιστική κατανομή πιθανότητας του $R_n \equiv R$ δίνεται από την σχέση

$$P(R \leq r | \alpha_1, \alpha_2, \dots, \alpha_n) = \frac{1}{(2\pi)^{n-1}} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} d\theta_{n-1} d\theta_{n-2} \dots d\theta_1, \quad (\text{A.3.4})$$

όπου τα $\theta_1, \theta_2, \dots, \theta_{n-2}$ λαμβάνουν τιμές στο διάστημα $(-\pi, \pi)$ ενώ το θ_{n-1} έχει εύρος τιμών τέτοιο ώστε $R \leq r$.

Στο σημείο αυτό θα χρησιμοποιήσουμε το θεώρημα Weber με την εξίσωση (A.3.2) στην ακόλουθη μορφή:

$$r \int_0^{\infty} J_0(Rt) J_1(rt) dt = \begin{cases} 1, & R < r \\ 0, & R > r \end{cases}. \quad (\text{A.3.5})$$

Βασίζόμενοι, λοιπόν, στην εξίσωση (A.3.5), ως δείκτρια συνάρτηση που περιορίζει το R στο σωστό εύρος, η εξίσωση (A.3.4) γίνεται:

$$\begin{aligned} P(R \leq r | \alpha_1, \alpha_2, \dots, \alpha_n) &= \\ &= \frac{1}{(2\pi)^{n-1}} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \int_0^{\infty} r J_0(Rt) J_1(rt) dt d\theta_{n-1} d\theta_{n-2} \dots d\theta_1. \end{aligned} \quad (\text{A.3.6})$$

Αντικαθιστούμε στην εξίσωση (A.3.6) τη σχέση $R^2 = R_n^2 = R_{n-1}^2 + \alpha_n^2 - 2\alpha_n R_{n-1} \cos\theta_{n-1}$, και χρησιμοποιούμε την ακόλουθη εξίσωση που προκύπτει από την αθροιστική ταυτότητα του Naumman:

$$\int_0^{2\pi} J_0 \left(t \sqrt{a^2 + b^2 - 2ab \cos \theta} \right) d\theta = 2\pi J_0 (ta) J_0 (tb). \quad (\text{A.3.7})$$

Βάσει των εξισώσεων (A.3.6) και (A.3.7) η εξίσωση (A.3.4), αλλάζοντας την σειρά ολοκλήρωσης και υπολογίζοντας το ολοκλήρωμα ως προς θ_{n-1} αρχικά, γίνεται:

$$\frac{2\pi}{(2\pi)^{n-1}} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} r J_0 (R_{n-1} t) J_0 (a_n t) J_1 (rt) dt d\theta_{n-2} \dots d\theta_1. \quad (\text{A.3.8})$$

Εκμεταλλευόμενοι όπως και προηγουμένως την ιδιότητα του συνημίτονου, εκτελούμε τη αντικατάσταση $R_{n-1}^2 = R_{n-2}^2 + \alpha_{n-1}^2 - 2\alpha_{n-1} R_{n-2} \cos \theta_{n-2}$. Ακολουθώντας την ίδια επαναληπτική διαδικασία για n βήματα, λαμβάνουμε την εξής συνάρτηση κατανομής για την τυχαία μεταβλητή R :

$$F(r) = P(R \leq r | \alpha_1, \alpha_2, \dots, \alpha_n) = r \int_0^{\infty} J_1(rt) \prod_{i=1}^n J_0(\alpha_i t) dt. \quad (\text{A.3.9})$$

Παραγωγίζοντας την σχέση (A.3.9) και εφαρμόζοντας την ιδιότητα (A.2.1) παίρνουμε την επιθυμητή συνάρτηση πυκνότητας πιθανότητας του συνιστάμενου διανύσματος για $0 \leq r \leq \sum_{i=1}^n \alpha_i$:

$$\begin{aligned} f(r | \alpha_1, \dots, \alpha_n) &= \frac{dF(r)}{dr} = \frac{d}{dr} \int_0^{\infty} \frac{rt}{t} J_1(rt) \prod_{i=1}^n J_0(\alpha_i t) dt = \\ &= \int_0^{\infty} \frac{drt J_1(rt)}{dr} \frac{1}{t} \prod_{i=1}^n J_0(\alpha_i t) dt = \int_0^{\infty} rt J_0(rt) \prod_{i=1}^n J_0(\alpha_i t) dt. \end{aligned} \quad (\text{A.3.10})$$

Αυτή είναι και η λύση που δόθηκε από τον *Kluyver* το 1906 για το γενικευμένο πρόβλημα του περίπατου του *Pearson*.

Στην συγκεκριμένη περίπτωση που εξετάζουμε εμείς, δηλαδή για παρατηρήσεις τυχαίας μεταβλητής που ακολουθεί την κυκλική ομοιόμορφη κατανομή και αναπαρίστανται ως μοναδιαία διανύσματα, τοποθετούμε στην σχέση (A.3.10) $\alpha_i = 1, \forall i = 1, \dots, n$ και λαμβάνουμε:

$$f_0(r) = r \int_0^{\infty} J_0(rt) J_0^n(t) dt, 0 \leq r \leq n. \quad (\text{A.3.11})$$

ΒΙΒΛΙΟΓΡΑΦΙΑ

• Ελληνική Βιβλιογραφία

Κοκολάκης Γ. και Σπηλιώτης Ι. (2002), Εισαγωγή στις Πιθανότητες, Εκδόσεις Συμεών, Αθήνα.

Κοκολάκης Γ. και Φουσκάκης Δ. (2009), Στατιστική- Θεωρία και Εφαρμογές, Εκδόσεις Συμεών, Αθήνα.

Φουσκάκης Δ. (2010), Σημειώσεις μαθήματος «Ανάλυση δεδομένων με Χρήση H/Y» Σχολή Ε.Μ.Φ.Ε, ΕΜΠ, Αθήνα.

Μαργαρίτης Ν. Λοτότζης (2011), Στατιστική Ανάλυση Κυκλικών Δεδομένων: Μια Γενική Επισκόπηση και μια Εφαρμογή σε Μετεωρολογικά Δεδομένα, Μεταπτυχιακή Διατριβή ΠΜΣ «Θετικές Επιστήμες στην Γεωπονία», Γεωπονικό Πανεπιστήμιο Αθηνών, Αθήνα.

• Ξενόγλωσση Βιβλιογραφία

Kluyver J.C. (1906), A local probability theorem, Ned.Akad.Wet.Proc. Ser. A, 8, p.341-350

Kent J.T. and Mardia K.V. and Rao J.S. (1979), A characterization of uniform distribution on the circle, Ann.Statis. ,15, p.247-254.

Tyler D.E. (1987), Statistical analysis for the angular central Gaussian distributions on the sphere, Biometrika ,74, p. 579–589

Mardia K. and Jupp P. (2000), Directional Statistics, John Wiley & Sons.

Jammalamadaka S.R. and Gupta A.S. (2001), Topics in Circular Statistics, World Scientific Publishing, Singapore.

Cox J. Nicholas (2001), Analyzing circular data in Stata, University of Durham, Durham City

Arnold C.B. and Gupta A.S. (2004), Probability distributions and statistical inference for axial data, University of California

Zar J. (2010), Biostatistical Analysis, Prentice–Hall.