

Προβλήματα Εκτίμησης Και Ελέγχων Υποθέσεων Σε Πεπερασμένα Μείγματα Της Κατανομής Poisson

Δημήτρης Καρλής

Περίληψη

Η σημασία των μοντέλων μειγμάτων κατανομών στη Στατιστική είναι μεγάλη και το πεδίο εφαρμογών τους συνεχώς αυξάνεται. Η διατριβή αυτή ασχολείται με μείγματα της κατανομής Poisson τα οποία χρησιμοποιούνται ως εναλλακτικά μοντέλα στις περιπτώσεις που η απλή κατανομή Poisson αποτυγχάνει να περιγράψει τα δεδομένα. Για παράδειγμα, είναι γνωστό ότι η κατανομή Poisson έχει τη χαρακτηριστική ιδιότητα ότι η διακύμανση της είναι ίση με την αναμενόμενη τιμή της. Αυτό πολλές φορές δεν συμβαίνει στην πράξη. Τα μείγματα της κατανομής Poisson αποτελούν ευέλικτα εναλλακτικά μοντέλα που μπορούν να περιγράψουν την ανομοιογένεια του πληθυσμού. Η λογική τους στηρίζεται στο γεγονός ότι εξαιτίας της ανομοιογένειας του πληθυσμού, τα άτομα που συνιστούν τον πληθυσμό δεν έχουν την ίδια συχνότητα εμφάνισης του υπο εξέταση γεγονότος. Η συχνότητα αυτή περιγράφεται από την παράμετρο της κατανομής Poisson, και συνεπώς ο καθένας μπορεί να έχει μια διαφορετική τιμή για τη συχνότητα αυτή. Επομένως η συχνότητα εμφάνισης του γεγονότος που περιγράφεται από την παράμετρο της κατανομής Poisson είναι μια τυχαία μεταβλητή που ακολουθεί κάποια κατανομή, η οποία αποκαλείται κατανομή μίξης. Τότε από το θεώρημα ολικής πιθανότητας τα μείγματα της κατανομής Poisson προκύπτουν. Οι ιδιότητες τους εξετάζονται σε αυτή τη διατριβή. Λόγω της πολυπλοκότητας τους, μόνο λίγες τέτοιες κατανομές έχουν ερευνηθεί. Σε αυτή τη διατριβή ένας μεγάλος αριθμός μελών της οικογένειας αυτής παρουσιάζεται και δίνεται έμφαση στις μεταξύ τους σχέσεις.

Μεταξύ των μειγμάτων της κατανομής Poisson τα πεπερασμένα μείγματα αποτελούν μια ενδιαφέρουσα κατηγορία. Είναι ιδιαίτερα διαδεδομένα κυρίως λόγω της απλής φυσικής ερμηνείας τους ως μοντέλα που περιγράφουν έναν πληθυσμό με πεπερασμένο αριθμό υποπληθυσμών. Επιπλέον, ακόμα και στις περιπτώσεις στις οποίες η κατανομή μίξης είναι συνεχής, η εκτίμησή της συνίσταται στην εκτίμηση μιας κατανομής με πεπερασμένο αριθμό σημείων με μη αρνητική πιθανότητα. Στη διατριβή αυτή εξετάζονται μέθοδοι εκτίμησης για πεπερασμένα μείγματα της κατανομής Poisson. Στην πράξη εμφανίζονται δυο διαφορετικές περιπτώσεις. Στην πρώτη ο αριθμός των μελών του μείγματος είναι δεδομένος οπότε απαιτείται η βελτιστοποίηση ενός κατάλληλου κριτηρίου για όλες τις κατανομές μίξης με το δεδομένο αριθμό σημείων με μη αρνητική πιθανότητα. Στη δεύτερη περίπτωση, ο αριθμός των μελών του μείγματος είναι άγνωστος και πρέπει επομένως να εκτιμηθεί από τα δεδομένα. Η περίπτωση αυτή είναι γνωστή ως ημιπαραμετρική περίπτωση. Για την πρώτη περίπτωση, ο αλγόριθμος EM προσφέρεται για εκτίμηση με τη μέθοδο μεγίστης πιθανοφάνειας και έχει χρησιμοποιηθεί ευρέως σε πολλές εφαρμογές. Στη διατριβή αυτή γίνεται μια κριτική επισκόπηση των χρήσεων του αλγορίθμου. Το πρόβλημα της επιλογής αρχικών τιμών που μπορούν να βελτιώσουν την ταχύτητα του αλγορίθμου εξετάζεται επίσης μέσω

προσομοίωσης. Επιπλέον χρησιμοποιώντας ιδιότητες των μειγμάτων κατανομών της εκθετικής οικογένειας κατανομών προτείνεται μια νέα μέθοδος που βελτιώνει την ταχύτητα του αλγόριθμου. Στην ημιπαραμετρική περίπτωση παρουσιάζονται οι αλγόριθμοι που υπάρχουν στη βιβλιογραφία. Οι αλγόριθμοι αυτοί δεν είναι κατάλληλοι για την περίπτωση των μειγμάτων Poisson γιατί στα μείγματα αυτά ο αριθμός των μελών του μείγματος είναι συνήθως μικρός με αποτέλεσμα να αποτυγχάνουν οι αλγόριθμοι.

Το πρόβλημα της εκτίμησης του αριθμού των μελών του μείγματος εξετάζεται επίσης. Μια νέα μέθοδος προτείνεται η οποία βασίζεται στην διαδοχική χρήση του ελέγχου λόγου πιθανοφανειών με μεθόδους bootstrap. Η χρήση των μεθόδων bootstrap είναι απαραίτητη γιατί η κατανομή της ελεγχοσυνάρτησης είναι άγνωστη. Οι ιδιότητες της μεθόδου ερευνώνται και προσδιορίζονται συνθήκες που διευκολύνουν τη χρήση του ελέγχου.

Εναλλακτικές μέθοδοι εκτίμησης ερευνώνται επίσης. Για τη μέθοδο των ροπών αποδεικνύεται ότι πολύ συχνά οι εκτιμήτριες δεν υπάρχουν επειδή το σύστημα των εξισώσεων δεν έχει λύση. Με βάση τις συγκρίσεις με τη μέθοδο της μεγίστης πιθανοφάνειας για μικρά δείγματα και αποδεικνύεται ότι η μέθοδος των ροπών έχει χαμηλότερη απόδοση. Μια εναλλακτική μέθοδος που εξετάζεται χρησιμοποιεί την παρατηρούμενη συχνότητα της τιμής 0 αντί της τρίτης ροπής για τις περιπτώσεις όπου η παρατηρούμενη σχετική συχνότητα της τιμής 0 είναι σχετικά μεγάλη.

Μια νέα μέθοδος, η οποία είναι ταυτόχρονα αποτελεσματική και εύρωστη μελετήθηκε. Η μέθοδος βασίζεται στην ελαχιστοποίηση της απόστασης Hellinger. Οι εκτιμήτριες που λαμβάνονται εξετάζονται και αποδεικνύεται ότι είναι εύρωστες σε σχέση με τις εκτιμήτριες μεγίστης πιθανοφάνειας. Η ιδιότητα αυτή μπορεί να χρησιμοποιηθεί για την ανάπτυξη μεθοδολογιών στατιστικής συμπερασματολογίας βασισμένων στην απόσταση Hellinger, όπως ημιπαραμετρική εκτίμηση, διαγνωστικά γραφήματα τα οποία έχουν σκοπό να διαγνώσουν αν η απλή κατανομή Poisson είναι κατάλληλη και ελέγχους υποθέσεων εναλλακτικούς των ελεγχών υποθέσεων βασισμένων στο λόγο πιθανοφανειών. Τέτοιοι έλεγχοι έχουν μεγάλη ισχύ και είναι ανθεκτικοί στην παρουσία ακραίων τιμών.