Υπολογιστική Στατιστική

Κατερίνα Ορφανογιαννάκη

Τμήμα Μαθηματικών
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
korfanog@math.uoa.gr

2020-2021

Μέθοδοι Monte Carlo Εισαγωγή

Όταν αναφερόμαστε σε μεθόδους Monte Carlo ουσιαστικά αναφερόμαστε σε μεθόδους που χρησιμοιούν την προσωμοίωση.

Βασική ιδέα: Ακόμα και αν έχουμε απουσία γνώσης αναφορικά με την κατανομή μιας στατιστικής συνάρτησης αν μπορούμε να προσωμοιώσουμε τιμές αυτής της συνάρτησης τότε στην πραγματικότητα γνωρίζουμε πολλά πράγματα γιάυτή.

Έλεγχοι Monte Carlo

Πως λειτουργούν;

Προσωμοιώνουμε πολλά δείγματα από τη μηδενική υπόθεση και για κάθε δείγμα υπολογίζουμε την τιμή της ελεγχοσυνάρτησης. Οι τιμές αυτές αποτελούν μια καλή προσέγγιση της κατανομής της ελεγχοσυνάρτησης και επομένως συγκρινοντάς τες με την παρατηρούμενη τιμή της ελεγχοσυνάρτησης καταλήγουμε σε συμπερασμα.

Έλεγχοι Monte Carlo: Ποιά η διαφορά με τους ελέγχους τυχαιοποίησης;

Στους ελέγχους τυχαιοποίησης η μηδενική υπόθεση υποθέτει απουσία δομής. Χρησιμοποιόντας την απουσία δομής δημιουργούμε πολλά δείγματα και για να προσεγγίσουμε την κατανομή της ελεγχοσυνάρτησης.

ΣΤους ελέγχους Monte Carlo προσομοιώνουμε από την κατανομή που η μηδενική υπόθεση καθορίζει. Σε πολλές περιπτώσεις χρειάζεται να γνωρίζουμε την κατανομή του πληθυσμού.

Τα βήματα για τον έλεγχο Monte Carlo:

- Βήμα 1ο: Θέσε τη μηδενική και την εναλλακτική υπόθεση
- Βήμα 2ο: Διάλεξε την ελεγχοσυνάρτηση.
- Βήμα 3α: Προσομοίωσε k δείγματα μεγέθους n (το μέγεθος του δικού μας πραγματικού δείγματος) από τη μηδενική υπόθεση (αυτό συνήθως σημαίνει από την κατανομή και τις παραμέτρους που υπονοεί η μηδενική υπόθεση)
- Βήμα 3β: Για κάθε δείγμα που δημιούργησες υπολόγισε την τιμή της ελεγχοσυνάρτησης.
- Βήμα 4ο: Συγκρίνοντας την τιμή της ελεγχοσυνάρτησης για τα πραγματικά δεδομένα με αυτές τις τιμές από τα δείγματα που προσομοίωσες, εκτίμησε το $p - value$ ως

$$p - value = \frac{m+1}{k+1} \quad ,$$

όπου m μετράει πόσες φορές η τιμή της ελεγχοσυνάρτησης από τα προσομοιωμένα δείγματα είναι πιο ακραία από αυτή που παρατηρήσαμε.

Παράδειγμα I

Σε ένα εργοστάσιο, οι χρόνοι σε μέρες που μεσολάβησαν ανάμεσα σε δύο διαδοχικές βλάβες ενός μηχανήματος καταγράφηκαν με σκοπό να μελετηθεί η συχνότητα εμφάνισης βλαβών. Οι 12 παρατηρήσεις που πάρθηκαν ήταν οι εξής

0.18, 0.89, 0.61, 0.23, 0.16, 0.37, 0.29, 0.03, 0.05, 0.47, 1.17, 0.24

Να ελεγχθεί η μηδενική υπόθεση ότι ο μέσος χρόνος *m* ανάμεσα σε δύο διαδοχικές βλάβες είναι 1 μέρα ξέροντας πως η κατανομή του χρόνου είναι η εκθετική

Α. Αν η εναλλακτική είναι πως $\mu > 1$

Β. Αν η εναλλακτική είναι πως $\mu < 1$

Γ. Αν η εναλλακτική είναι πως $\mu \neq 1$

Ελεγχοσυνάρτηση για το παράδειγμα Ι:

$$T = \bar{x} - 1$$

H_0: $T = 0$

Απορρίπτουμε τη μηδενική υπόθεση αν:

- Για την περίπτωση Α: οι τιμές της T είναι μεγάλες (δεξιά ουρά).
- Για την περίπτωση Β: οι τιμές της T είναι μικρές (αριστερή ουρά).
- Για την περίπτωση Γ: οι τιμές της T είναι και στις δύο ουρές. Δηλαδή για αυτή την περίπτωση η ελεγχοσυνάρτηση είναι: $|T|$.

Ο έλεγχος για το παράδειγμα Ι:

- Για την περίπτωση Α μετράμε πόσες παρατηρήσεις είναι μεγαλύτερες από -0.61 που πήραμε για το δείγμα μας.
- Για την περίπτωση Β μετράμε πόσες παρατηρήσεις είναι μικρότερες από -0.61 που πήραμε για το δείγμα μας.
-
- Για την περίπτωση Γ μετράμε πόσες παρατηρήσεις είναι μικρότερες από -0.61 και πόσες μεγαλύτερες από 0.61.

Παράδειγμα II

Σε μια αγωνιστική του πρωταθλήματος ποδοσφαίρου σημειώθηκαν συνολικά σε κάθε παιχνίδι τα γκολ που φαίνονται στον παρακάτω πίνακα. Θέλουμε να απαντήσουμε στο ερώτημα κατά πόσο τα δεδομένα ακολουθούν την κατανμή *Poisson*.

Αριθμός Γκολ	Συχνότητα (Αριθμός ομάδων)
0	6
1	3
2	4
3	2
4	1

Ο έλεγχος

Ο έλεγχος που θέλουμε να κάνουμε είναι ένας έλεγχος καλής προσαρμογής. Επειδή δεν έχουμε καμιά πληροφορία σχετικά με την παράμετρο της κατανομής Ποισσον θα πρέπει να την εκτιμήσουμε.

Επομένως η μηδενική υπόθεση είναι:
H_0: η κατανομή είναι *Poisson* έναντι της
H_1: η κατανομή δεν είναι *Poisson*

Άλλη εναλλακτική

Επειδή η εναλλακτική είναι πολύ γενική θα υποθέσουμε πως η εναλλακτική είναι η:

H_1: η κατανομή είναι Αρνητική Διωνυμική.

Πρέπει: Να βρούμε κάποια ποσότητα που χαρακτηρίζει την κατανομή σε σχέση με την εναλλακτική.

Ποια ιδιότητα χαρακτηρίζει την κατανομή *Poisson*;

Για την περίπτωση της *Poisson* ξέρουμε πως έχει την ιδιότητα να έχει μέση τιμή και διακύμανση ίσες κάτι που δεν ισχύει για την αρνητική διωνυμική. Άρα, μια ελεγχοσυνάρτηση που μπορούμε να χρησιμοποιήσουμε είναι η:

$$T = \frac{s^2}{\bar{x}}$$

για την οποία περιμένουμε τιμές κοντά στο 1 αν τα δεδομένα είναι από την κατανομή *Poisson* ενώ αν προέρχονται από την αρνητική διωνυμική οι τιμές πρέπει να είναι μεγαλύτερες από τη μονάδα.

Ο έλεγχος για το παράδειγμα II είναι:

- Προσομοιώνουμε k, έστω 99, δείγματα από την κατανομή *Poisson* με παράμετρο 1.31 (ο μέσος των τιμών του δείγματος που έχουμε). Κάθε τέτοιο δείγμα έχει 16 παρατηρήσεις. Η εκτιμήτρια μεγίστης πιθανοφάνειας για την παράμετρο της κατανομής Ποισσον είναι η μέση τιμή του δείγματος.
- Για κάθε δείγμα υπολογίζουμε την ελεγχοσυνάρτηση
- Στη συνέχεια βρίσκουμε πόσες από τις 99 τιμές της ελεγχοσυνάρτησης είναι πιο μεγάλες από αυτή του δείγματος μας.

Συμπεράσματα

- Σχετικά με την επιλογή της ελεγχοσυνάρτησης, τον αριθμό των επαναλήψεων και την εκτίμηση του $p - value$, ισχύει ότι και στους ελέγχους τυχαιοποίησης.
- Στους ελέγχους *Monte Carlo* χρησιμοποιούμε επιπλέον την πληροφορία που έχουμε σχετικά με την κατανομή του πληθυσμού.
- Οι έλεγχοι αυτοί είναι ιδιαίτερα χρήσιμοι για ελέγχους καλής προσαρμογής όπου η μηδενική υπόθεση μας κατευθύνει για την κατανομή από την οποία θα προσομοιώσουμε τα δείγματα μας.