

Δειγματοληψία με άνισες πιθανότητες επιλογής

Η ιδέα της δειγματοληψίας με άνισες πιθανότητες επιλογής:

Πολλές φορές οδηγούμαστε σε δειγματοληπτικά σχήματα όπου κάθε μονάδα (ή και συστάδα) του πληθυσμού έχει διαφορετική πιθανότητα επιλογής, π.χ. ανάλογη του μεγέθους της (probability proportional to size)

Πρόβλημα:

Εκτίμηση

- της μέσης τιμής μ
- της ολικής τιμής τ

Δειγματοληψία με επανάθεση και άνισες πιθανότητες επιλογής: Συμβολισμοί

Πληθυσμός:

1. N : μέγεθος πληθυσμού
2. U_j : η τιμή του χαρακτηριστικού U για τη j μονάδα του πληθυσμού, $j = 1, \dots, N$
3. p_j : η πιθανότητα επιλογής σε μία δοκιμή (γνωστή ή προσδιορίσιμη) της j μονάδας του πληθυσμού, $j = 1, \dots, N$, είναι $\sum_{j=1}^N p_j = 1$

Δείγμα:

1. n : μέγεθος δείγματος
2. X_i : η τιμή του χαρακτηριστικού U για τη i μονάδα του δείγματος, $i = 1, \dots, n$
3. p_i^* : η πιθανότητα επιλογής σε μία δοκιμή που αντιστοιχεί στην i μονάδα του δείγματος, π.χ.

$$X_1 = U_2, \quad p_1^* = p_2$$

Δειγματοληψία με επανάθεση και άνισες πιθανότητες επιλογής

Πρόταση 1 (εκτιμητές Hansen-Hurwitz)

Έστω

$$\hat{\tau}_{HH} = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{p_i^*} \quad \text{και} \quad \hat{\mu}_{HH} = \frac{1}{N} \hat{\tau}_{HH}$$

1. Οι εκτιμητές $\hat{\tau}_{HH}$ και $\hat{\mu}_{HH}$ είναι ΑΕ.

$$2. \text{Var}(\hat{\tau}_{HH}) = \frac{1}{n} \sum_{j=1}^N \left(\frac{U_j}{p_j} - \tau \right)^2 p_j \quad \text{και} \quad \text{Var}(\hat{\mu}_{HH}) = \frac{1}{N^2} \text{Var}(\hat{\tau}_{HH})$$

$$3. \hat{\text{Var}}(\hat{\tau}_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{X_i}{p_i^*} - \hat{\tau}_{HH} \right)^2 \quad \text{και}$$

$\hat{\text{Var}}(\hat{\mu}_{HH}) = \frac{1}{N^2} \hat{\text{Var}}(\hat{\tau}_{HH})$ είναι ΑΕ των $\text{Var}(\hat{\tau}_{HH})$ και $\text{Var}(\hat{\mu}_{HH})$, αντίστοιχα.

Απόδειξη.

Δειγματοληψία με επανάθεση και άνισες πιθανότητες επιλογής

Σχόλιο:

Αν $p_1 = \dots = p_N$ τότε $p_j = 1/N$, για κάθε $j = 1, \dots, N$ και συνεπώς $p_i^* = 1/N$ για κάθε $i = 1, \dots, n$, δηλαδή έχουμε **ΤΔΕ**. Προφανώς,

$$\hat{\tau}_{HH} = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{1/N} = N \frac{1}{n} \sum_{i=1}^n X_i = N \bar{X}$$

Δειγματοληψία με ή χωρίς επανάθεση και άνισες πιθανότητες επιλογής: Συμβολισμοί

Πληθυσμός:

1. N : μέγεθος πληθυσμού
2. U_j : η τιμή του χαρακτηριστικού U για τη j μονάδα του πληθυσμού, $j = 1, \dots, N$
3. π_j : η πιθανότητα συμμετοχής στο δείγμα (γνωστή ή προσδιορίσιμη) της j μονάδας του πληθυσμού, $j = 1, \dots, N$, π.χ.

$$\text{ΤΔΕ} \quad p_j = 1/N \text{ και } \pi_j = 1 - \left(1 - \frac{1}{N}\right)^n$$

$$\text{ΑΤΔ} \quad p_j = 1/N \text{ και } \pi_j = \frac{n}{N}$$

4. π_{j_1, j_2} : η πιθανότητα συμμετοχής στο δείγμα (γνωστή ή προσδιορίσιμη) των j_1, j_2 μονάδων του πληθυσμού

Δειγματοληψία με ή χωρίς επανάθεση και άνισες πιθανότητες επιλογής: Συμβολισμοί

Δείγμα:

1. n : μέγεθος δείγματος
2. X_i : η τιμή του χαρακτηριστικού U για τη i μονάδα του δείγματος, $i = 1, \dots, n$
3. ν : ο αριθμός των διαφορετικών μονάδων στο δείγμα μεγέθους n , π.χ.

χωρίς επανάθεση $\nu = n$, όλες οι μονάδες του δείγματος διαφορετικές
με επανάθεση ν είναι τ.μ. με τιμές $\nu = 1, 2, \dots, n$

4. Z_k : η τιμή του χαρακτηριστικού U για την k διαφορετική μονάδα του δείγματος, $k = 1, \dots, \nu$
5. π_k^* : η πιθανότητα συμμετοχής στο δείγμα που αντιστοιχεί στην k διαφορετική μονάδα του δείγματος, $k = 1, \dots, \nu$,
6. π_{i_1, i_2}^* : η πιθανότητα συμμετοχής στο δείγμα που αντιστοιχεί στις i_1, i_2 διαφορετικές μονάδες του δείγματος

Δειγματοληψία με ή χωρίς επανάθεση και άνισες πιθανότητες επιλογής: Συμβολισμοί

Παράδειγμα:

χωρίς επανάθεση έστω $X_1 = U_2, X_2 = U_5, X_3 = U_{10}$, τότε $\nu = 3$ και

$$\begin{array}{lll} Z_1 = X_1 & Z_2 = X_2 & Z_3 = X_3 \\ \pi_1^* = \pi_2 & \pi_2^* = \pi_5 & \pi_3^* = \pi_{10} \end{array}$$

με επανάθεση έστω $X_1 = U_2, X_2 = U_5, X_3 = U_2$ τότε $\nu = 2$ και

$$\begin{array}{ll} Z_1 = X_1 & Z_2 = X_2 \\ \pi_1^* = \pi_2 & \pi_2^* = \pi_5 \end{array}$$

Δειγματοληψία με ή χωρίς επανάθεση και άνισες πιθανότητες επιλογής

Πρόταση 2 (εκτιμητές Horvitz-Thompson)

Έστω

$$\hat{\tau}_{HT} = \sum_{i=1}^{\nu} \frac{Z_i}{\pi_i^*} \quad \text{και} \quad \hat{\mu}_{HT} = \frac{1}{N} \hat{\tau}_{HT}$$

1. Οι εκτιμητές $\hat{\tau}_{HT}$ και $\hat{\mu}_{HT}$ είναι ΑΕ.

$$2. \text{Var}(\hat{\tau}_{HT}) = \sum_{j=1}^N \frac{U_j^2}{\pi_j} (1 - \pi_j) + 2 \sum_{j_1 < j_2} \frac{U_{j_1} U_{j_2}}{\pi_{j_1} \pi_{j_2}} (\pi_{j_1, j_2} - \pi_{j_1} \pi_{j_2})$$

όπου π_{j_1, j_2} η πιθανότητα συμμετοχής συγχρόνως των j_1, j_2 στο δείγμα,

$$\text{και} \quad \text{Var}(\hat{\mu}_{HT}) = \frac{1}{N^2} \text{Var}(\hat{\tau}_{HT})$$

$$3. \hat{\text{Var}}(\hat{\tau}_{HT}) = \sum_{i=1}^{\nu} \frac{Z_i^2}{(\pi_i^*)^2} (1 - \pi_i^*) + 2 \sum_{i_1 < i_2} \frac{Z_{i_1} Z_{i_2}}{\pi_{i_1}^* \pi_{i_2}^*} \left(\frac{\pi_{i_1, i_2}^* - \pi_{i_1}^* \pi_{i_2}^*}{\pi_{i_1, i_2}^*} \right)$$

και $\hat{\text{Var}}(\hat{\mu}_{HT}) = \frac{1}{N^2} \hat{\text{Var}}(\hat{\tau}_{HT})$ είναι ΑΕ των $\text{Var}(\hat{\tau}_{HT})$ και $\text{Var}(\hat{\mu}_{HT})$, αντίστοιχα.

Απόδειξη. (στην περίπτωση χωρίς επανάθεση όπου $\nu = n$)

Δειγματοληψία με ή χωρίς επανάθεση και άνισες πιθανότητες επιλογής: Σύνοψη

Εκτιμητές Hansen-Hurwitz (με επανάθεση):

$$\hat{\tau}_{HH} = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{p_i^*} \quad \text{και} \quad \hat{\mu}_{HH} = \frac{1}{N} \hat{\tau}_{HH}$$

- p_i^* : η πιθανότητα επιλογής σε μία δοκιμή που αντιστοιχεί στην i μονάδα του δείγματος

Εκτιμητές Horvitz-Thompson (με ή χωρίς επανάθεση):

$$\hat{\tau}_{HT} = \sum_{i=1}^{\nu} \frac{Z_i}{\pi_i^*} \quad \text{και} \quad \hat{\mu}_{HT} = \frac{1}{N} \hat{\tau}_{HT}$$

- ν : ο αριθμός των διαφορετικών μονάδων στο δείγμα μεγέθους n
- Z_i : η τιμή του χαρακτηριστικού U για την i διαφορετική μονάδα του δείγματος, $i = 1, \dots, \nu$
- π_i^* : η πιθανότητα συμμετοχής στο δείγμα που αντιστοιχεί στην i διαφορετική μονάδα του δείγματος