Προσομοίωση 2-6-2022

1. eclass – γραφτείτε!

2. Υλικό

   a) Robert – Casella: Introduction to Monte Carlo methods with R.
   (free from Springer με διεύθυνση IP uoa.gr).

   b) σημειώσεις – λυμένη σειρά ασκήσεων.

3. Γλώσσα R

4. Project (ασκήσεις) 30% προαιρετικό

5. Τελική εξέταση στο εργαστήριο.

Ενότητες

1. Εισαγωγή Monte Carlo Simulation
   Παραδείγματα
   Γεννήτριες Τυχαίων Αριθμών

2. Μέθοδοι Μείωσης Διασποράς

3. Μέθοδοι Markov Chain Monte Carlo (MCMC)
   Gibbs Sampling

Παραδείγματα – Εφαρμογές παντού

Εισαγωγή: Προσομοίωση Monte Carlo

Προσομοίωση: Αναπαράσταση ("μίμηση") λειτουργίας συστήματος με χρήση υπολογιστή.

π.χ.

① Ρίχνω ζάρι 16 φορές

X = αρ. των 6.

$X \sim Bin(16, 1/6)$

$P(X \geq 2)$ - - - - υπολογίζεται αναλυτικά

② Ρίχνω ζάρι 100 φορές

$P($άθροισμα $= 226)$

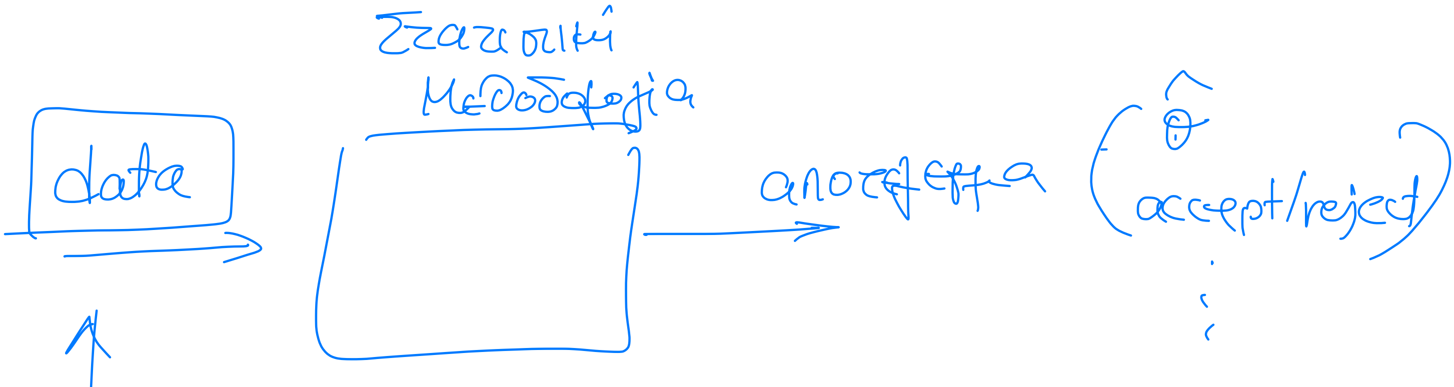
$P($άθροισμα $< 347)$ [προσεγγίζεται μέσω ΚΟΘ]

③ Στατιστική Μέθοδος
(π.χ. εκτιμήτρια,
έλεγχος υποθέσεων κτλ).

π.χ. $\hat{\theta}$ εκτιμήτρια παραμέτρου θ

Είναι αμερόληπτη;

$$\boxed{E(\hat{\theta}) = \theta \quad \forall \theta.}$$



Προσομοιωμένα δεδομένα

Πολλά προσομοιωμένα τυχαία δείγματα από υπολογιστή $\Rightarrow$ εκτίμηση ιδιοτήτων της μεθόδου

τεκμηρίωση

Παράδειγμα

Έστω $X \sim \mathcal{N}(\mu, \sigma^2)$, μ, σ^2 άγνωστα

$$\theta = E(\max(X-100, 0))$$

Δείγμα $X_1, X_2 \ldots X_n \Rightarrow \boxed{\hat{\theta} = ?}$

$$\boxed{\hat{\theta} = f(X_1, X_2, \ldots, X_n)}$$ εκτιμήτρια

Μεροληψία $E(\hat{\theta}) = ?$ εξαρτάται από μ, σ

Εφαρμογές

1) Τεκμηρίωση (Εκτίμηση Συμπεριφοράς) Στατιστικών μεθόδων όταν η θεωρητική ανάλυση δύσκολη ή αδύνατη.

2) Προσομοίωση Βιολογικών Πληθυσμών.

Παράδειγμα

Πληθυσμός $N = 1000$ ατόμων

Κάθε άτομο έχει επαφή με m άλλα άτομα

Σε κάθε επαφή πιθ. μετάδοσης ασθένειας $= p$

P(μετά από 25 μέρες υπάρχουν >300 ασθενείς)

Monte Carlo

Παραδείγματα στο R.

Προσομοίωση Monte Carlo

Έστω τυχαία μεταβλητή X
με συνάρτηση κατανομής $F(x)$.
και $Y = h(X)$

Θέλουμε να υπολογίσουμε

$$\theta = E(h(X))$$

Αν X διακριτή $\quad \theta = \sum_{x} h(x) P(X=x)$

Αν X συνεχής με σππ $f(x)$ $\Rightarrow \theta = \int_{-\infty}^{\infty} h(x) f(x) dx$.

Υπολογιστικά μπορεί να είναι δύσκολο
ή αδύνατο

Προσομοίωση

Έστω $X_1, X_2, \dots, X_n$ δείγμα από την κατανομή της X

Εκτίμηση του θ : $\hat{\theta}_n = \dfrac{\sum_{i=1}^{n} h(x_i)}{n}$

Συνεπής : $\lim_{n\to\infty} \hat{\theta}_n = \theta$

Πώς βρίσκω το δείγμα;

Στον υπολογιστή μέσω κατάλληλου λογισμικού

Πώς είναι αυτό εφικτό;

Τέτοιοι αλγόριθμοι λέγονται γεννήτορες τυχαίων αριθμών

Δημιουργούν ακολουθία ψευδοτυχαίων αριθμών

$x_0, x_1, x_2, \ldots$

$x_1 = g(x_0)$
$x_2 = g(x_1)$
$x_3 = g(x_2)$
$\vdots$

στατιστικά οι αριθμοί περνούν όλους τους ελέγχους καλής προσαρμογής από την επιθυμητή κατανομή

π.χ. $x_{n+1} = (1232 + 37\, x_n) \bmod (127)$

Αν εφαρμοστεί ο ίδιος αλγόριθμος
με την ίδια αρχική τιμή x_0
θα δώσει πάλι τα ίδια <u>αποτελέσματα</u>

$$\theta = E(h(X))$$

π.χ. $\theta = P(X > 20)$

$$Z = h(X) = 1(X > 20) = \begin{cases} 1 & \text{αν } X > 20 \\ 0 & \text{αν } X < 20 \end{cases}$$

$$E(Z) = 0 \cdot P(Z=0) + 1 \cdot P(Z=1)$$

$$= P(Z=1) = P(X > 20) = \theta$$

$$X_1, X_2, \dots, X_n$$

$$Z_1, Z_2, \dots, Z_n$$

$$\hat{\theta} = \frac{Z_1 + \dots + Z_n}{n} = \frac{\text{αρ. } Z \text{ όπου } X > 20}{n}$$

$= \%$ παρατη. όπου $X > 20$

Παράδειγμα

Έστω $X \sim \mathcal{N}(\mu = 50, \sigma = 10)$

$$\theta = E(\max(X - 60, 0))$$

$$h(X) = \max(X - 60, 0)$$

① Δημ. $X_1, \dots, X_N \sim \mathcal{N}(50, 10^2)$

②
$$\left.\begin{aligned} Y_1 &= \max(X_1 - 60, 0) \\ Y_2 &= \max(X_2 - 60, 0) \\ &\vdots \\ Y_N &= \dots \end{aligned}\right\} \quad \hat{\theta} = \overline{Y}_N$$

Άσκηση Υλοποιήστε το παραμετρικά

$$X \sim \mathcal{N}(\mu, \sigma^2), \quad \theta = E(\max(X - d, 0))$$

Προσ. δείγμα μεγέθους N

thest(μ, σ, d, N)