1. Το κεντρικο οριακο θεωρημα

Θεώρημα: Έστω $(X_n)_{n\geq 1}$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με πεπερασμένη μέση τιμή $\mu = EX_1$ και πεπερασμένη και θετική διασπορά[1] $\sigma^2 = V(X_1)$. Θέτουμε $S_n := X_1 + \cdots + X_n$. Τότε για κάθε διάστημα $I \subset \mathbb{R}$ ισχύει

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \in I\right) = P(Z \in I)$$

όπου η Z ακολουθεί την τυπική κανονική κατανομή $N(0,1)$.

Δηλαδή, πρακτικά, για n μεγάλο, η τυχαία μεταβλητή

$$W_n := \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \tag{1}$$

ακολουθεί την κατανομή $N(0,1)$.

Παρατήρηση. 1) Για μιά τυχαία μεταβλητή X με πεπερασμένη μέση τιμή $\mu := E(X)$ και πεπερασμένη και θετική διασπορά $\sigma^2 = V(X)$, η τυχαία μεταβλητή

$$Y := \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X-\mu}{\sigma}$$

λέγεται **κανονικοποίηση** της X, και έχει $E(Y) = 0, V(Y) = 1$. Γιατί

$$\begin{aligned} E(Y) &= \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}\{E(X) - \mu\} = 0, \\ V(Y) &= V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(X-\mu) = \frac{1}{\sigma^2}V(X) = 1. \end{aligned}$$

2) Η W_n που ορίσαμε στην (1) είναι απλώς η κανονικοποίση της S_n. Γιατί

$$\begin{aligned} E(S_n) &= E(X_1 + \cdots + X_n) = n\mu, \\ V(S_n) &= V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n) = n\sigma^2. \end{aligned}$$

Στην τελευταία γραμμή χρησιμοποιήσαμε την ανεξαρτησία των $\{X_i : i \geq 1\}$.

Προσέξτε ότι η τυχαία μεταβλητή W_n της (1) ακολουθεί *ακριβώς* την $N(0,1)$ μόνο αν οι X_i ακολουθούν την κατανομή $N(\mu,\sigma^2)$, για κάποια μ, σ^2, και όχι σε άλλες περιπτώσεις. Είναι κάτι που δεν αποδείξαμε.

Για παράδειγμα, αν οι X_i είναι Bernoulli(1/2), τότε η W_n είναι διακριτή τυχαία μεταβλητή, ενώ η $N(0,1)$ είναι συνεχής κατανομή. Παρατηρούμε επιπλέον ότι η W_n παίρνει με θετική πιθανότητα μόνο τιμές στο σύνολο

$$\left\{\frac{k - n/2}{\sqrt{n/4}} : k = 0, 1, \ldots, n\right\}$$

το οποίο απέχει πάρα πολύ απο το να είναι το $\mathbb{R}$ (το οποίο είναι όλες οι δυνατές τιμές μιας μεταβλητής που ακολουθεί την $N(0,1)$). Είναι όμως ένα σύνολο με σημεία που εκτείνονται απο το $-\sqrt{n}$ ως το $\sqrt{n}$, και η απόσταση μεταξύ διαδοχικών σημείων είναι $2/\sqrt{n}$ (δηλαδή μικρή). Κατά μία έννοια, αυτό το σύνολο *προσεγγίζει* το $\mathbb{R}$ όταν $n \to \infty$.

[1] Μηδενική διασπορά σημαίνει ότι η X_1 είναι μια σταθερά έστω c. Τότε η ακολουθία $(X_n)_{n\geq 1}$ είναι σταθερή, και δεν έχουμε να κάνουμε τίποτα.

Άσκηση 1. Τα αιτήματα που φτάνουν σε έναν server έχουν το καθένα τυχαίο χρόνο εξυπηρέτησης που ακολουθεί την εκθετική κατανομή με παράμετρο $\theta = 1/2$ (δηλαδή πυκνότητα $(1/2)e^{-x/2}\mathbf{1}_{x>0}$, το x σε λεπτά). Ο server μπορεί να απασχολείται με μόνο ένα αίτημα σε κάθε δεδομένη στιγμή. Ποιά είναι προσεγγιστικά η πιθανότητα να εξυπηρετήσει τα πρώτα 100 αιτήματα μιας δεδομένης μέρας σε συνολικό χρόνο το πολύ 220 λεπτά; Δίνονται Φ(1)=0.8413, Φ(1.5)=0.9332, Φ(2)=0.9773.

Λύση

Για $i \geq 1$, έστω X_i ο χρόνος εξυπηρέτησης του i αιτήματος. Απο τις ιδιότητες της εκθετικής κατανομής έχουμε $E(X_1) = 1/\theta = 2, V(X_1) = 1/\theta^2 = 4$. Έστω $S_n = X_1 + X_2 + \cdots + X_n$ για κάθε $n \geq 1$. Το κεντρικό οριακό θεώρημα λέει ότι η

$$\frac{S_n - 2n}{\sqrt{4n}}$$

προσεγγιστικά ακολουθεί την κατανομή $N(0,1)$.

Ζητάμε την πιθανότητα

$$P\left(S_{100} \leq 220\right) = P\left(\frac{S_{100} - 200}{20} \leq 1\right) \approx P(Z \leq 1) = \Phi(1) \approx 0.8413.$$

Στις πιο πάνω ισότητες, η Z είναι μια τυχαία μεταβλητή που ακολουθεί την τυπική κανονική κατανομή $N(0,1)$, και χρησιμοποιήσαμε το κεντρικό οριακό θεώρημα.

Άσκηση 2. Έστω $X_1, X_2, \ldots, X_{80}$ ανεξάρτητες τυχαίες μεταβλητές που η καθεμία ακολουθεί τη διακριτή ομοιόμορφη κατανομή στο $\{1,2,3,4\}$. Να βρεθεί προσεγγιστικά η πιθανότητα το άθροισμά τους να είναι στο $[190, 220]$ (δίνονται Φ(1)=0.8413, Φ(1.5)=0.9332, Φ(2)=0.9773).

Λύση

Έστω $(X_i)_{i\geq 1}$ μια ακολουθία ανεξάρτητων ισόνομων τυχαίων μεταβλητών με κατανομή ομοιόμορφη στο $\{1,2,3,4\}$, και $S_n := X_1 + \cdots + X_n$ για κάθε $n \geq 1$. Οι δύο πρώτες ροπές της X_1 είναι

$$E(X_1) = \frac{1}{4}(1+2+3+4) = \frac{5}{2},$$
$$E(X_1^2) = \frac{1}{4}(1^2+2^2+3^2+4^2) = \frac{15}{2}.$$

Άρα η X_1 έχει μέση τιμή 5/2 και πεπερασμένη διασπορά $\sigma^2 = E(X_1^2) - (E(X_1))^2 = 5/4$. Το κεντρικό οριακό θεώρημα λέει οτι για n μεγάλο, η τυχαία μεταβλητή

$$\frac{S_n - 5n/2}{\sqrt{5/4}\sqrt{n}}$$

ακολουθεί προσεγιστικά την τυπική κανονική κατανομή, $N(0,1)$.

Για $n = 80$, έχουμε $5n/2 = 200$ και $\sqrt{(5/4)\,n} = 10$, οπότε

$$\begin{aligned} P(190 \leq S_{80} \leq 220) &= P\left(-1 \leq \frac{S_{80} - 200}{10} \leq 2\right) \approx P(-1 < Z < 2) \\ &= \Phi(2) - \Phi(-1) = \Phi(2) + \Phi(1) - 1 \approx 0.8186, \end{aligned}$$

όπου $Z \sim N(0,1)$. Η προσέγγιση στην δεύτερη ισότητα προκύπτει απο το κεντρικό οριακό θεώρημα.

Άσκηση 3. Πραγματοποιούμε μια ακολουθία ρίψεων ενός αμερόληπτου ζαριού.

(α) Θέτουμε $X_i = 1$ αν στην i ρίψη το αποτέλεσμα ήταν 5 ή 6, ενώ διαφορετικά θέτουμε $X_i = 0$. Ποιά η μέση τιμή και ποιά η διασπορά της X_i;

(β) Έστω W το σύνολο των φορών στις πρώτες 1800 ρίψεις που έρχεται 5 ή 6. Να υπολογιστεί προσεγγιστικά η πιθανότητα $P(580 < W < 640)$.

Δίνεται ότι $\Phi(1) = 0.8413, \Phi(2) = 0.9773$.

Λύση

(α) Έχουμε $P(X_i = 1) = 1/3, P(X_i = 0) = 2/3$. Δηλαδή, κάθε X_i έχει κατανομή Bernoulli με $p = 1/3$. Άρα $E(X_i) = 1/3, V(X_i) = p(1-p) = 2/9$.

(β) Θέτουμε $S_n := X_1 + \cdots + X_n$ για κάθε $n \geq 1$. Τότε $W = S_{1800}$, και το κεντρικό οριακό θεώρημα δίνει ότι

$$\begin{aligned} P(580 < S_{1800} < 640) &= P\left(-1 < \frac{S_{1800} - 1800\,(1/3)}{\sqrt{1800\,(2/9)}} < 2\right) \approx P(-1 < Z < 2) = \Phi(2) - \Phi(-1) \\ &= \Phi(2) + \Phi(1) - 1 \approx 0.8186 \end{aligned}$$

όπου $Z \sim N(0,1)$.