

HY-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες - Εαρινό Εξάμηνο 2025
Διδάσκων: Π. Τσακαλίδης

Φροντιστήριο 2

Άσκηση 1

Ψηφοφόροι κάποιας περιφέρειας ψήφισαν για έναν από τους υποψηφίους Α και Β. Η αρχική καταμέτρηση έδειξε ότι ο υποψήφιος Β προηγείται του Α, κατά 537 ψήφους. Όμως, υπάρχουν 25.600 ψηφοδέλτια τα οποία εκκρεμούν στο εκλοδικείο και δεν έχουν καταχωρηθεί σε κανέναν από τους δύο υποψηφίους. Υποθέστε ότι τελικά και τα 25.600 ψηφοδέλτια, θα βρεθούν έγκυρα και θα αποδοθούν ισοπίθानα στους δύο υποψηφίους, κάθε ένα ανεξάρτητα από τα υπόλοιπα. Έστω X και Y το πλήθος των ψηφοδελτίων (από τα 25.600) που θα αποδοθούν στους Α και Β αντίστοιχα.

- (α) Τι είδους μεταβλητή είναι η X ;
- (β) Να βρεθούν η μέση τιμή και η διασπορά της τ.μ X .
- (γ) Είναι οι τ.μ X και Y ανεξάρτητες;
- (δ) Από τη στιγμή που και οι 25.600 ψήφοι προσμετρηθούν στο τελικό αποτέλεσμα, ποια είναι η μέση τιμή και η διασπορά της τ.μ $Z = X - Y$, του καθαρού κέρδους του υποψηφίου Α στην ψηφοφορία;
- (ε) Από τη στιγμή που και οι 25.600 ψήφοι προσμετρηθούν στο τελικό αποτέλεσμα, χρησιμοποιήστε το Κεντρικό Οριακό Θεώρημα για να βρείτε την πιθανότητα: "Να κερδίσει ο Β την εκλογή".

Λύση

- (α) Η τυχαία μεταβλητή X είναι διωνυμική με παραμέτρους: $(n, p) = (25.600, 1/2)$.
- (β) Η μέση τιμή και η διασπορά της τ.μ X γίνονται:

$$E[X] = np = 25.600 \cdot \frac{1}{2} = 12.800$$

και

$$\text{var}(X) = np(1 - p) = 25.600 \cdot \left(\frac{1}{2}\right)^2 = 6.400$$

- (γ) Έχουμε ότι: $X + Y = 25.600$. Άρα, οι τ.μ X, Y δεν μπορούν να είναι ανεξάρτητες.
- (δ) Έχουμε ότι: $Z = X - Y$. Όμως: $X + Y = 25.600 \Rightarrow Y = 25.600 - X$.
Η μέση τιμή της τ.μ Z γίνεται:

$$E[Z] = E[X] - E[Y] = E[X] - E[25.600 - X] = 2 \cdot E[X] - 25.600 = 2 \cdot 12.800 - 25.600 = 0$$

Για τη διασπορά της τ.μ Z έχουμε:

$$\text{var}(Z) = \text{var}(X - Y) = \text{var}(X - 25.600 + X) = \text{var}(2X - 25.600) = 4 \cdot \text{var}(X) = 4 \cdot 6.400 = 25.600$$

- (ε) Η Z μπορεί να προσεγγιστεί από μια κανονική τ.μ με $Z \sim N(0, 25.600)$. Σύμφωνα με το Κεντρικό Οριακό Θεώρημα, η πιθανότητα γίνεται:

$$P(\text{"Ο Β κερδίζει"}) = P(Z \leq 536) \approx \Phi\left(\frac{536 - 0}{\sqrt{25.600}}\right) = \Phi(3.35) \approx 0.9996$$

Άσκηση 2.

Έστω n ανεξάρτητες ομοιόμορφα κατανομημένες τ.μ. $X_1, X_2, \dots, X_n$, με τυπική απόκλιση ίση με 1. Θεωρήστε τον δειγματικό μέσο $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ και υπολογίστε τα παρακάτω ερωτήματα.

- (i) Εάν η τυπική απόκλιση του M_n είναι τουλάχιστον 0.1, ποιό είναι το μέγεθος των μεταβλητών n ;
- (ii) Εάν η εκτίμηση του δειγματικού μέσου όρου είναι το πολύ 0.01 με πιθανότητα τουλάχιστον 0.80, ποιό είναι το μέγεθος των μεταβλητών n ; (**Υπόδειξη:** Χρησιμοποιήστε την ανισότητα *Chebyshev*.)

Λύση

- (i) Με βάση τη θεωρία, ξέρουμε ότι η τυπική απόκλιση του M_n δίνεται από τον εξής τύπο:

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}$$

Με βάση την εκφώνηση της άσκησης, έχουμε:

$$\sigma_{M_n} \leq 0.1 \Rightarrow \frac{1}{\sqrt{n}} \leq 0.1 \Rightarrow \sqrt{n} \geq \frac{1}{0.1} \Rightarrow \sqrt{n} \geq 10 \Rightarrow n \geq 100$$

- (ii) Με βάση την εκφώνηση της άσκησης, θέλουμε να ισχύει:

$$P(|M_n - h| \leq 0.01) \geq 0.80$$

Κάνοντας χρήση της ανισότητας *Chebyshev*, έχουμε:

$$\begin{aligned} P(|M_n - h| \leq 0.01) &\stackrel{h=E[M_n]}{=} P(|M_n - E[M_n]| \leq 0.01) \\ &= 1 - P(|M_n - E[M_n]| \geq 0.01) \\ &\stackrel{\sigma_{M_n}^2 = \frac{1}{n}}{\geq} 1 - \frac{\frac{1}{n}}{(0.01)^2} \end{aligned}$$

Άρα, τελικά, έχουμε:

$$\begin{aligned} P(|M_n - h| \leq 0.01) &\geq 0.80 \Rightarrow 1 - \frac{\frac{1}{n}}{(0.01)^2} \geq 0.80 \\ &\Rightarrow \frac{\frac{1}{n}}{(0.01)^2} \leq 0.20 \Rightarrow \frac{1}{n} \leq 0.00002 \\ &\Rightarrow n \geq \frac{1}{0.00002} \Rightarrow n \geq 50000 \end{aligned}$$

Άσκηση 3

Τα μήλα μιας δενδροκαλλιέργειας έχουν μέσο βάρος $\mu = 50$ γρ με τυπική απόκλιση $\sigma = 10$ γρ. Τα μήλα συσκευάζονται σε κιβώτια και προωθούνται στην αγορά. Σε κάθε κιβώτιο τοποθετούνται 400 τυχαία επιλεγμένα μήλα. Μπορούμε να υπολογίσουμε κατά προσέγγιση ποιο ποσοστό των κιβωτίων περιέχει μήλα με μέσο βάρος: μεγαλύτερο των 51.25γρ;

Λύση

Τα μήλα κάθε κιβωτίου είναι ένα τυχαίο δείγμα μεγέθους $n = 400$ από τον πληθυσμό των μήλων της παραγωγής. Η μορφή της κατανομής των βαρών των μήλων δεν μας είναι γνωστή. Μπορεί να είναι οποιαδήποτε. Για την άγνωστη αυτή κατανομή γνωρίζουμε μόνο τη μέση τιμή της $\mu = 50\text{gr}$ και την τυπική απόκλιση της $\sigma = 10\text{gr}$. Τα ερωτήματά μας μπορούν να επαναδιατυπωθούν ως εξής: Ποιο κατά προσέγγιση ποσοστό των δειγματικών μέσων είναι μεγαλύτερο των 51.25gr .

Για να μπορέσουμε να απαντήσουμε στα ερωτήματα που θέσαμε πρέπει να γνωρίζουμε την κατανομή των δειγματικών μέσων. Το Κ.Ο.Θ. μας βεβαιώνει ότι, παρότι δε γνωρίζουμε την κατανομή των βαρών των μήλων, εντούτοις, γνωρίζουμε την κατανομή των δειγματικών μέσων αφού τα δείγματά μας έχουν μέγεθος αρκετά μεγάλο. Δηλαδή, αρκεί μόνο, το ότι γνωρίζουμε τη μέση τιμή και την τυπική απόκλιση της κατανομής των βαρών των μήλων.

Έτσι, αν συμβολίσουμε με X την τυχαία μεταβλητή που εκφράζει το βάρος ενός τυχαία επιλεγμένου μήλου της καλλιέργειας, και με $\bar{X}$ την τυχαία μεταβλητή που εκφράζει τους δειγματικούς μέσους, το Κ.Ο.Θ. μας βεβαιώνει ότι η $\bar{X}$ ακολουθεί κατά προσέγγιση κανονική κατανομή με μέση τιμή $\mu_{\bar{X}} = \mu = 50\text{gr}$ και διασπορά $\sigma_{\bar{X}}^2 = \sigma^2/n = 0.25\text{gr}^2$. Αφού, πλέον, γνωρίζουμε ότι $\bar{X} \sim N(50, 0.5^2)$, η απάντηση στα ερωτήματά μας είναι πολύ απλή. Ζητάμε τις πιθανότητες: $P(X > 51.25)$. Έτσι έχουμε:

$$P(\bar{X} > 51.25) = 1 - P(\bar{X} \leq 51.25) = 1 - P(Z \leq \frac{51.25 - 50}{\sqrt{0.25}}) = 1 - \Phi(2.5) = 0.0062.$$

Άρα, το ποσοστό των κιβωτίων που περιέχουν μήλα με μέσο βάρος μεγαλύτερο των 51.25gr είναι κατά προσέγγιση 0.62% .

Άσκηση 4

Ένα δίκαιο εξαέδρο ζάρι ρίχνεται n φορές. Έστω X_i , το αποτέλεσμα της i -οστής ρίψης. Υποθέτουμε ότι οι τ.μ X_i είναι ανεξάρτητες και ορίζουμε το άθροισμα των ρίψεων: $S_n = X_1 + X_2 + \dots + X_n$. Υπολογίστε ένα άνω όριο για την πιθανότητα $P(S_{100} \geq 400)$.

- (α) Χρησιμοποιώντας την ανισότητα Markov.
- (β) Χρησιμοποιώντας την ανισότητα Chebyshev.
- (γ) Χρησιμοποιώντας προσέγγιση με βάση το κεντρικό οριακό θεώρημα.

Τι παρατηρείτε;

Λύση

- (α) Η μέση τιμή των τ.μ X_i γίνεται:

$$E[X_i] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$

Σύμφωνα με την ανισότητα Markov έχουμε:

$$P[S_{100} \geq 400] \leq \frac{100 \cdot 3.5}{400}$$

- (β) Υπολογίζουμε τη διασπορά των τ.μ X_i ως εξής:

$$\text{var}(X_i) = E[X_i^2] - (E[X_i])^2$$

Όμως, έχουμε ότι:

$$E[X_i^2] = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = 15$$

Επομένως, η διασπορά γίνεται:

$$\text{var}(X_i) = E[X_i^2] - (E[X_i])^2 = 15 - 3.5^2 = 2.75$$

Σύμφωνα με την ανισότητα του Chebyshev έχουμε:

$$P[S_{100} \geq 400] = P[|S_{100} - 350| \geq 50] \leq \frac{100 \cdot 2.75}{50^2} = 0.11$$

(γ) Με βάση το Κεντρικό Οριακό Θεώρημα έχουμε:

$$\begin{aligned} P[S_{100} \geq 400] &= 1 - P[S_{100} \leq 400] = 1 - P\left[\frac{S_{100} - 350}{\sqrt{2.75 \cdot 100}} \leq \frac{400 - 350}{\sqrt{2.75 \cdot 100}}\right] \\ &\approx 1 - \Phi\left(\frac{400 - 350}{\sqrt{2.75 \cdot 100}}\right) \\ &\approx 1 - \Phi(3.015) \\ &\approx 0.0009 \end{aligned}$$

Άσκηση 5

Έστω $X_1, X_2, \dots, X_n$ ανεξάρτητες τ.μ. Bernoulli με πιθανότητα επιτυχίας $p = 0.2$.

- (α) Έστω η τ.μ. $S_n = X_1 + X_2 + \dots + X_n$, για $n = 40$. Δημιουργείστε $N = 10000$ τυχαία δείγματα της τ.μ. S_n .
- (β) Θεωρείστε το δειγματικό μέσο όρο τους $M_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{S_n}{n}$ και απεικονίστε το ιστόγραμμα του για 20 bins. Τι παρατηρείτε;
- (γ) Ορίστε την τ.μ. $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ και δώστε το ιστόγραμμα της για 20 bins. Τι παρατηρείτε;

Λύση

- (α) Η τ.μ. $S_n = X_1 + X_2 + \dots + X_n$ θα ακολουθεί τη Διωνυμική κατανομή με παραμέτρους $n = 40$ και $p = 0.2$.
Δημιουργούμε $N = 10000$ τυχαία δείγματα της Διωνυμικής τ.μ. S_n , χρησιμοποιώντας τη συνάρτηση `binornd` της Matlab, όπως φαίνεται στον κώδικα παρακάτω.

```
% Initialize variables
```

```
N = 10000;
```

```
n = 40;
```

```
p = 0.2;
```

```
n_bins = 20;
```

```
% a) Binomial random variable Sn = X1+X2+...+Xn
```

```
Sn = binornd(n,p,N,1);
```

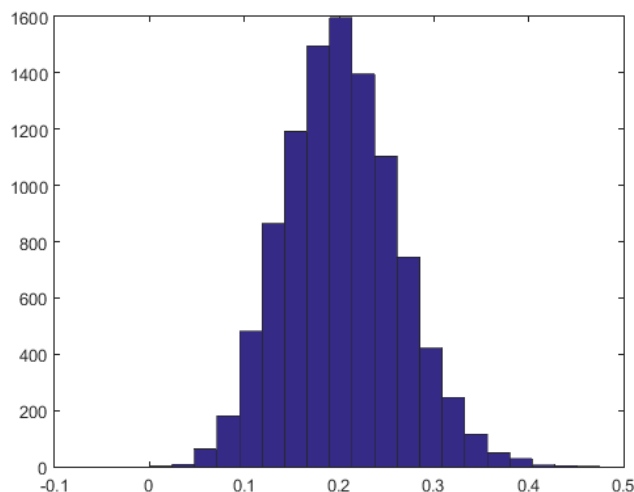
- (β) Υπολογίζουμε το δειγματικό μέσο όρο $M_n = \frac{S_n}{n}$ των δειγμάτων που δημιουργήσαμε, όπως φαίνεται στον παρακάτω κώδικα.

```
% b) Mn
```

```
Mn = Sn/n;
```

```
figure; hist(Mn,n_bins);
```

Στο σχήμα 1 φαίνεται το ιστόγραμμα του για 20 bins. Όπως περιμέναμε από τη θεωρία, η M_n ακολουθεί κανονική κατανομή γύρω από τη μέση τιμή $\mu = p = 0.2$ των X_i .

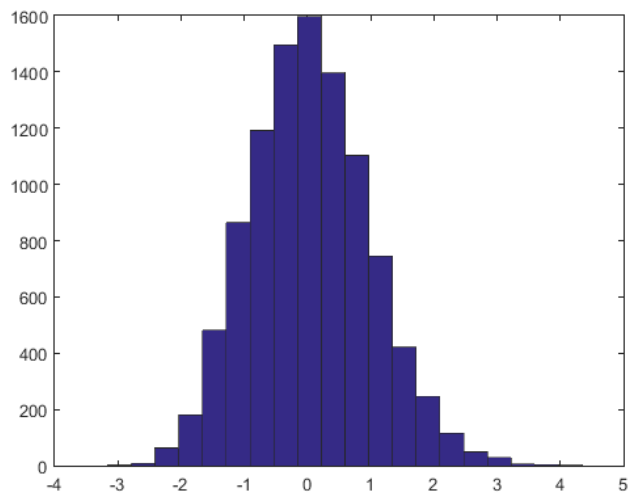


Σχήμα 1: Ιστόγραμμα του δειγματικού μέσου όρου M_n .

(γ) Υπολογίζουμε την τ.μ. $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$, με μέση τιμή $\mu = p$ και διασπορά $\sigma^2 = p(1-p)$ των τ.μ. Bernoulli X_i , όπως φαίνεται και στον παρακάτω κώδικα.

```
% c) Random Variable Zn = (Sn-n*mu)/(sigma*sqrt(n))
mu = p;
sigma = sqrt(p*(1-p));
Zn = (Sn-n*mu)/(sigma*sqrt(n));
figure; hist(Zn,n_bins);
```

Στο σχήμα 2 φαίνεται το ιστόγραμμά του για 20 bins. Σύμφωνα και με το θεώρημα κεντρικού ορίου, παρατηρούμε ότι η Z_n συγκλίνει στην τυπική κανονική κατανομή με μέση τιμή 0 και διασπορά 1.



Σχήμα 2: Ιστόγραμμα της τ.μ. Z_n .