Μάθημα 25 - 30.05.2023
ΓΡΑΦΕΙΣ
Γιάννης Βασιλόπουλος
Σωτηρία Χριστοδουλοπούλου

Πιθανότητες 1, Οριακά Θεωρήματα

30 Μαΐου 2023

Πώς συμπεριφέρται ένα άθροισμα πολλών (ανεξάρτητων) μεταβλητών

- Τιμές μετοχών στο χρηματιστήριο
- Κλίμα/Μετεωρολογία
- Κατανομή ταχυτήτων σε ένα αέριο
  Έστω μια ακολουθία $X_1, X_2, ..., X_n$ από τυχαίες μεταβλητές ανεξάρτητες και ισόνομες με μέση τιμή μ και διασπορά σ^2.

Πώς συμπεριφέρεται η $S_n = \sum_{i=1}^{n} X_i$ ως συνάρτηση του n

Έχουμε δει ότι: $E[S_n] = n\mu$ και $Var(S_n) = n\sigma^2$ (από ανεξαρτησία μεταβλητών).

- Κανονικοποιημένο άθροισμα: $\boxed{Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}}$ Η Z_n «κινείται» γύρω από τη μέση τιμή της $E[Z_n] = 0$ με διασπορά $Var(Z_n) = 1$.

  Έχουμε ήδη λίγη πληροφορία στατιστικής φύσεων για τη Z_n, αλλά δεν αρκεί!
  Αυτό που θα θέλαμε είναι η κατανομή $\boxed{P(Z_n \geq z)}$ ή την $\boxed{f_n(z) \equiv f_{Z_n}(z)}$.

- Μας ενδιαφέρουν οι ουρές $\boxed{P(|Z_n| \geq z)}$, δηλαδή η πιθανότητα να έχουμε απόκλιση πάνω από z από την κανονικοποιημένη μέση τιμή.
- Μας ενδιαφέρει η συμπεριφορά αυτών των ουρών για $n \to \infty$.

Πώς εξάγουμε πληροφορία για τις ουρές μιας μεταβλητής από τα στατιστικά της

- Ανισότητα $Markov$:

  Έστω τυχαία μεταβλητή $X > 0$. Τότε: $\boxed{P(X \geq a) \leq \dfrac{E[X]}{a}}$ $\forall a > 0$.

- Απόδειξη (για τη συνεχή περίπτωση):

  $E[X] = \int_0^\infty xf(x)dx = \int_0^a xf(x)dx + \int_a^\infty xf(x)dx \geq \int_a^\infty xf(x)dx \geq a\int_a^\infty af(x)dx = aP(X \geq a) \Rightarrow \boxed{P(X \geq a) \leq \dfrac{E[X]}{a}}$

- Από $Markov: P(|Z_n| \geq z) \leq \frac{E[Z_n]}{z} \left(\leq \frac{\sqrt{E[Z_n^2]}}{z} = \frac{1}{z} \right)$.
  Το φράγμα $O(\frac{1}{z})$ δεν είναι αρκετά ακριβές.

- Ανισότητα $Chebychev$: Αν η τυχαία μεταβλητή X έχει $E[X] < \infty$ και $E[x^2] < \infty$, τότε, $\forall a > 0$, έχουμε: $\boxed{P(|X - \mu| \geq a) \leq \dfrac{\sigma^2}{a^2}}$, όπου $\mu = E[X]$ και $\sigma^2 = Var(X)$.

- Απόδειξη:

  Έστω $U = (X - \mu)^2$, οπότε:

  $$P(|X-\mu| \geq a) = P((X-\mu)^2 \geq a^2) = P(U \geq a^2) \leq \frac{E[U]}{a^2} = \frac{E[(x-\mu)^2]}{a^2} = \frac{Var(X)}{a^2} = \frac{\sigma^2}{a^2}$$

  .

- Από ανισότητα $Chebychev$ στην Z_n, προκύπτει ότι $\boxed{P(|Z_n| \geq z) \leq \dfrac{1}{z^2}}$.

- Αφού $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$, θα έχουμε:
  $P(|Z_n| \geq z) = P(\frac{S_n - n\mu}{\sigma\sqrt{n}} \geq z) = P(|S_n - n\mu| \geq z\sqrt{n}\sigma) \leq \frac{1}{z^2}$.

  Θέτοντας $a = z\sqrt{n}\sigma$, παίρνουμε: $\boxed{P(|S_n - n\mu| \geq a) \leq \dfrac{n\sigma^2}{a^2}}$.

Μελέτη του δειγματικού μέσου όρου

$$\overline{\mathrm{X_n}} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n}$$

Εκτίμηση του μέσου όρου μ μιας τυχαίας μεταβλητής X από μια δειγματοληψία μεγέθους n:
$P(|\overline{\mathrm{X_n}} - \mu| \geq a) = P(|\frac{S_n}{n} - \mu| \geq a) = P(|S_n - n\mu| \geq an) \leq \frac{n\sigma^2}{n^2a^2} = \frac{\sigma^2}{na^2}$

Δείξαμε, λοιπόν, ότι: $\boxed{P(|S_n - n\mu| \geq an)\dfrac{\sigma^2}{na^2}}$

Ασθενής Νόμος Μεγάλων Αριθμών ($Bernoulli, Khinchin$)

Έστω $X_1, ..., X_n$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με $E[X^2] < \infty$.
Τότε: $\forall \epsilon > 0, P(|\overline{\mathrm{X}} - \mu| \geq \epsilon) \to 0, n \to \infty$.

Λέμε ότι η $\overline{\mathrm{X_n}}$ συγκλίνει στη μ κατά πιθανότητα και γράφουμε ότι $\overline{\mathrm{X_n}} \xrightarrow{p} \mu, n \to \infty$.
(Καθώς $n \to \infty$, η τυχαιότητα «εξαφανίζεται».)

Ισχυρός Νόμος Μεγάλων Αριθμών ($Borel, Kolmogorov$)

Έστω $X_1, ..., X_n$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με μέση τιμή μ και διασπορά σ^2.
Τότε: $\boxed{P(\overline{\mathrm{X_n}} \to \mu, \text{ καθώς } n \to \infty) = 1}$.

Λέμε ότι η $\overline{\mathrm{X_n}}$ συγκλίνει στη μ σχεδόν βεβαίως.

<u>Σημειώσεις:</u>

1. Ισχυρός Ν.Μ.Α. $\Rightarrow$ Ασθενή Ν.Μ.Α.
   (Σχεδόν βέβαιη σύγκλιση $\Rightarrow$ Σύγκλιση κατά πιθανότητα)
2. Ωστόσο, υπάρχουν περιπτώσεις (χαλαρότερες υποθέσεις) στις οποίες ισχύει ο Ασθενής Ν.Μ.Α., αλλά όχι ο Ισχυρός Ν.Μ.Α..

Πώς συμπεριφέρεται, τελικά, η $P(|Z_n| \geq z$ για μεγάλο n

<u>Δύο εφαρμογές:</u>

1. Τυχαίοι περίπατοι
2. Πλήθος των πρώτων διαιρετών
   Διαλέγουμε έναν τυχαίο αριθμό X ομοιόμορφα στην τύχη από το $\{1, ..., 10^{12}\}$.
   Έστω D ο αριθμός των πρώτων διαιρετών του X.
   π.χ.: Αν $X = 30$, τότε $D_X = 3$.

Και στις δύο περιπτώσεις, η εμπειρική κατανομή είναι κανονική! (καθώς $n \to \infty$)
Θέλουμε να εκτιμήσουμε την $P(|Z_n| \geq z)$.

- Γνωρίζουμε ότι $E[Z_n] = 0$ και $E[Z_n^2] = 1$.
- Ιδανικά, θα θέλαμε να γνωρίζουμε όλες τις ροπές $E[Z_n^r]$ της Z_n.
(ή όσο περισσότερες γίνεται)
- Εναλλακτικά, θα θέλουμε να υπολογίζουμε τη ροπογεννήτρια συνάρτηση $M_n(t) := E[exp(tZ_n)]$.
- Εξ ορισμού, έχουμε:
$M_n(t) = E\left[exp(tZ_n)\right] = E\left[exp\left(t\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)\right] = E\left[exp\left(t\frac{X_1+...+X_n}{\sigma\sqrt{n}}\right)\right] exp\left(-\frac{\mu\sqrt{n}t}{\sigma}\right)$.
- Για απλότητα, θα συνεχίσουμε με την περίπτωση $\mu = 0$ και $\sigma = 1$.
$M_n(T) = E\left[exp\left(t\frac{X_1+...+X_n}{\sigma\sqrt{n}}\right)\right] = E\left[exp\left(t\frac{X_1}{\sqrt{n}}\right) \cdot ... \cdot exp\left(t\frac{X_n}{\sqrt{n}}\right)\right] =$
$= E\left[exp\left(t\frac{X_1}{\sqrt{n}}\right)\right] \cdot ... \cdot E\left[exp\left(t\frac{X_n}{\sqrt{n}}\right)\right] = E\left[exp\left(t\frac{X}{\sqrt{n}}\right)\right]^n = \left[M_X\left(\frac{t}{\sqrt{n}}\right)\right]^n$

για X μια τυχαία μεταβλητή ισόνομη με $X_1, ...X_n$.

Εξ ορισμού:
$M_X(u) = E[exp(uX)] = E[1] + uE[X] + \frac{u^2}{2}E[X^2] + O(u^3) =$
$= 1 + \frac{u^2}{2} + O(u^3), E[X] = 0, E[X^2] = 1 \Rightarrow$

$$M_n(t) = \left[1 + \frac{t^2}{2n} + O\left(\frac{t^3}{n^{3/2}}\right)\right]^n \to e^{\frac{t^2}{2}}$$

- Δείξαμε ότι οι ροπογεννήτριες συναρτήσεις $M_n(t)$ συγκλίνουν
$\forall t \in \mathbb{R}, M_n(t) \to e^{\frac{t^2}{2}}$ καθώς $n \to \infty$,
δηλαδή στη ροπογεννήτρια συνάρτηση της $f(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{x^2}{2}}$.
- Λήμμα *Levy*: Αν η ροπογεννήτρια συνάρτηση $M_n(t)$ της Z_n συγκλίνει στη ροπογεννήτρια συνάρτηση $M(t)$ της Z, τότε:
$F_{Z_n}(z) \to F_Z(z)$ καθώς $n \to \infty$ σε κάθε σημείο συνέχειας της F_Z.

Κεντρικό Οριακό Θεώρημα

Έστω ότι $X_1, ..., X_n$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με μέσο μ και διασπορά σ^2. Τότε, η κατανομή του κανονικοποιημένου αθροίσματος $Z_n = \frac{X_1+...+X_n *}{\sigma\sqrt{n}}$ τείνει καθώς $n \to \infty$ στην τυπική κανονική κατανομή $* -n\mu$
$f_Z(z) \to \frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}$ καθώς $n \to \infty, \forall z \in \mathbb{R}$.

30/5/23

Οριακά Θεωρήματα [κεφ. 8 από Ross]

Πως συμπεριφέρεται ένα άθροισμα πολλών (ανεξάρτητων) τυχαίων μεταβλητών?

π.χ. Εφαρμογή
- → Τιμές μετοχών στο χρηματιστήριο
- → Κλίμα / Μετεωρολογία
- → κατανομή ταχυτήτων σε ένα αέριο

⋙ Έστω μια ακολουθία $X_1, X_2, \dots, X_n$ από τυχαίες μεταβλητές που είναι
i) ανεξάρτητες
ii) ισόνομες
} Δεν μας είναι απαραίτητα πάντα αλλά τα υποθέτουμε πάντοτε

με μέση τιμή μ και διασπορά σ^2

⋙ Πως συμπεριφέρεται η $S_n = \sum_{i=1}^{n} X_i$ ως συνάρτηση του n?

Έχουμε δει ότι $\mathbb{E}[S_n] = n\mu$ (δεν χρειάζομαι ανεξαρτησία) $Var(S_n) = n\sigma^2$ (χρειάζομαι ανεξαρτησία)

Κανονικοποίηση αθροίσματος: $\boxed{Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}}$
→ κανονικοποιημένο άθροισμα

Η Z_n "κινείται" γύρω από τη μέση τιμή της $\mathbb{E}[Z_n] = 0$ με διασπορά $Var(Z_n) = 1$

Έχουμε ήδη λίγη πληροφορία στατιστικής φύσεως για την Z_n αλλά ΔΕΝ αρκεί
Αυτό που θα θέλαμε είναι η κατανομή
$\underbrace{\mathbb{P}(Z_n \geq z)}_{\text{ουρές}}$ ή την $F_n(z) \equiv F_{Z_n}(z)$

⋙ Μας ενδιαφέρουν οι ουρές $\boxed{\mathbb{P}(|Z_n| \geq z)}$ δηλαδή η πιθανότητα να έχουμε απόκλιση πάνω από z από την καν/νη μέση τιμή.
Μας ενδιαφέρει η συμπεριφορά αυτών των ουρών για $n \to \infty$

⋙ Πως εξάγουμε πληροφορία για τις ουρές μιας μεταβλητής από τα στατιστικά της?

Ανισότητα Markov: Έστω τ.μ. $X \geq 0$. Τότε

$$\boxed{\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}} \quad \forall a > 0$$

→ είναι ανάλογη για διακριτές
απόδειξη (για την συνεχής περίπτωση)

$$\mathbb{E}[X] = \int_{0 \to \text{όχι } -\infty}^{\infty} x f(x)\,dx = \underbrace{\int_0^a x f(x)\,dx}_{\geq 0 \text{ άρα το βγάζω}} + \int_a^{\infty} x f(x)\,dx \geq$$

αφού $X \geq 0$

$$\geq \int_a^{\infty} \underbrace{x}_{\geq a} f(x)\,dx \geq \int_a^{\infty} a f(x)\,dx = a \int_a^{\infty} f(x)\,dx = a\mathbb{P}(X \geq a)$$

$$\Rightarrow \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

∎

⋙ Στην $\mathbb{P}(|Z_n| \geq z)$: Από Markov: $\mathbb{P}(|Z_n| \geq z) \leq \frac{\mathbb{E}[|Z_n|]}{z}$

$$\left(\underset{\substack{\uparrow \\ \text{Από Jenson}}}{\leq} \frac{\sqrt{\mathbb{E}[Z_n^2]}}{z} = \frac{1}{z} \right)$$

⋙ Το φράγμα $O(1/z)$ δεν είναι αρκετά ακριβές

Ανισότητα Chebyshev: Αν η τ.μ. X έχει
- $\mathbb{E}[X] < \infty$
- $\mathbb{E}[X^2] < \infty$

Τότε $\forall a > 0$ έχουμε: $\boxed{\mathbb{P}(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}}$

όπου $\mu = \mathbb{E}[X]$, $\sigma^2 = Var(X)$

<u>απόδειξη:</u> Έστω $U = (X-\mu)^2$ οπότε

$$\mathbb{P}(|X-\mu| \geq a) = \mathbb{P}((X-\mu)^2 \geq a^2) = \mathbb{P}(U \geq a^2)$$

$$\mathbb{P}(U \geq a^2) \overset{\text{Markov}}{\leq} \frac{\mathbb{E}[U]}{a^2} = \frac{\mathbb{E}[(X-\mu)^2]}{a^2} = \frac{Var(X)}{a^2} = \frac{\sigma^2}{a^2}$$

>>> Στην $\mathbb{P}(|Z_n| \geq z)$: Από Chebyshev: $\boxed{\mathbb{P}(|Z_n| \geq z) \leq \frac{1}{z^2}}$

Αφού $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ θα έχουμε

$$\mathbb{P}(|Z_n| \geq z) = \mathbb{P}\left(\frac{|S_n - n\mu|}{\sqrt{n}\sigma} \geq z\right) = \mathbb{P}(|S_n - n\mu| \geq \sqrt{n}z\sigma)$$

Από Chebyshev: $\mathbb{P}(|S_n - n\mu| \geq \underbrace{\sqrt{n}z\sigma}_{a}) \leq \frac{1}{z^2}$

$$\Rightarrow \boxed{\mathbb{P}(|S_n - n\mu| \geq a) \leq \frac{n\sigma^2}{a^2}}$$

>>> Μελέτη του δειγματικού μέσου όρου $\overline{X}_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$

(εκτίμηση του μέσου μ μιας τ.μ. X από μία δειγματοληψία μεγέθους n)

$$\Rightarrow \mathbb{P}(|\overline{X}_n - \mu| \geq a) = \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq a\right) =$$

$$\boxed{= \mathbb{P}(|S_n - n\mu| \geq an) \leq \frac{n\sigma^2}{n^2a^2} = \frac{\sigma^2}{na^2}}$$

Δείξαμε λοιπόν ότι: $\boxed{\mathbb{P}(|\overline{X}_n - \mu| \geq \alpha) \leq \frac{\sigma^2}{n\alpha^2}}$

Ασθενής νόμος Μεγάλων Αριθμών (Bernoulli Khinchin)

Έστω $X_1, \dots, X_n$ ακολουθία ανεξ. ισονόμων μετ. με $\mathbb{E}[X_i^2] < \infty$

Τότε $\forall \varepsilon > 0$ $\boxed{\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \longrightarrow 0 \text{ καθώς } n \to \infty}$

Λέμε ότι η $\overline{X}_n$ συγκλίνει στην μ κατά πιθανότητα και γράφουμε ότι $\overline{X}_n \xrightarrow{P} \mu$ καθώς $n \to \infty$

Καθώς $n \to \infty$ η τυχαιότητα "εξαφανίζεται"

Ισχυρός νόμος Μεγάλων Αριθμών (Borel Kolmogorov)

Έστω $X_1, \dots, X_n$ ακολουθία ισονόμων και ανεξάρτητων τ.μ. με μέση τιμή μ και διασπορά σ.

Τότε $\boxed{\mathbb{P}(\overline{X}_n \to \mu \text{ καθώς } n \to \infty) = 1}$

Λέμε ότι η $\overline{X}_n$ συγκλίνει στην μ σχεδόν βέβαια

Σημειώσεις

1) Ισχυρός ΝΜΑ $\Rightarrow$ Ασθενή ΝΜΑ
(σχεδόν βέβαιη σύγκλιση $\Rightarrow$ σύγκλιση κατά πιθανότητα)

2) Ωστόσο υπάρχουν περιπτώσεις (χαλαρότερες υποθέσεις) όπου ισχύει ο ασθενής ΝΜΑ αλλά όχι ο ισχυρός ΝΜΑ

→→→ Πως συμπεριφέρεται τελικά η $\mathbb{P}(|Z_n| \geq z)$ για μεγάλο n?

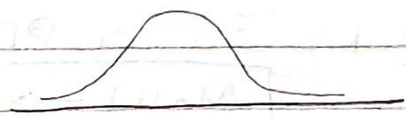
→ Δύο εφαρμογές:

- τυχαίοι περίπατοι
- πλήθος πρώτων διαιρετών

Διαλέγουμε έναν αριθμό ομοιόμορφα στην τύχη από το $\{1, \dots, N = 10^{12}\}$. Έστω D ο αριθμός των πρώτων διαιρετών του X.

π.χ. αν $X = 30$, τότε $D_X = 3$

Και στις δύο περιπτώσεις η εμπειρική κατανομή είναι κανονική (καθώς $n \to \infty$)



→→→ Θέλουμε να εκτιμήσουμε την $\mathbb{P}(|Z_n| \geq z)$

- Γνωρίζουμε $\mathbb{E}[Z_n] = 0$
  $\mathbb{E}[Z_n^2] = 1$
- Ιδανικά θα θέλαμε να γνωρίζουμε όλες τις ροπές $\mathbb{E}[Z_n^r]$ της Z_n [ή όσο περισσότερες γίνεται]
- Εναλλακτικά θα θέλαμε να υπολογίσουμε τη ΡΓΣ
  $M_n(t) = \mathbb{E}[\exp(t Z_n)]$

Εξ' ορισμού έχουμε: $M_n(t) = \mathbb{E}[\exp(t Z_n)] = \mathbb{E}\left[\exp\left(t \frac{S_n - n\mu}{\sqrt{n}\,\sigma}\right)\right]$

$$M_n(t) = \mathbb{E}\left[\exp\left(t \frac{X_1 + \dots + X_n}{\sqrt{n}\,\sigma}\right)\right] \exp\left(-\frac{\mu\sqrt{n}\,t}{\sigma}\right)$$

Για απλότητα θα συνεχίσουμε με την περίπτωση $\mu = 0$ $\sigma = 1$

$$\Rightarrow M_n(t) = \mathbb{E}\left[\exp\left(t \frac{X_1 + \dots + X_n}{\sqrt{n}}\right)\right] = \mathbb{E}\left[\exp\left(t \frac{X_1}{\sqrt{n}}\right) \times \dots \times \exp\left(t \frac{X_n}{\sqrt{n}}\right)\right]$$

$(X_1, \dots, X_n)_{\perp\!\!\!\perp}$ Έστω X ισόνομη με $X_1, \dots, X_n$

$$\overset{\downarrow}{=} \mathbb{E}\left[\exp\left(t \frac{X_1}{\sqrt{n}}\right)\right] \times \dots \times \mathbb{E}\left[\exp\left(t \frac{X_n}{\sqrt{n}}\right)\right] \overset{\downarrow}{=} \underbrace{\mathbb{E}\left[\exp\left(t \frac{X}{\sqrt{n}}\right)\right]^n}_{M_X(t/\sqrt{n})}$$

$$\Rightarrow M_n(t) = [M_X(t/\sqrt{n})]^n \quad (*)$$

Εξ' ορισμού $M_X(u) = \mathbb{E}[\exp(uX)] = \mathbb{E}[1 + uX + \frac{u^2}{2}X^2 + O(u^3)]$

$$= \mathbb{E}[1] + u\,\mathbb{E}[X] + \frac{u^2}{2}\mathbb{E}[X^2] + O(u^3)$$

$\mu = 0$, $\sigma = 1$

$$= 1 + \frac{u^2}{2} + O(u^3)$$

$$(*) \Rightarrow M_n(t) = \left[1 + \frac{t^2}{2n} + O\left(\frac{t^3}{n^{3/2}}\right)\right]^n \longrightarrow e^{t^2/2} \qquad \lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a$$

Δείξαμε ότι οι ΡΓΣ $M_n(t)$ συγκλίνουν $\forall t \in \mathbb{R}$

$\boxed{M_n(t) \longrightarrow e^{t^2/2}}$ καθώς $n \to \infty$

ΡΓΣ της $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Λήμμα (Lévy):

Αν οι ΡΓΣ $M_n(t)$ της Z_n συγκλίνει στη ΡΓΣ $M(t)$ της Z, τότε $\boxed{F_{Z_n}(z) \xrightarrow[n\to\infty]{} F_Z(z)}$ σε κάθε σημείο συνέχειας της F_Z

Κεντρικό οριακό θεώρημα

Έστω οι $X_1, \dots, X_n$ ανεξάρτητες και ισόνομες τ.μ. με μέσο μ και διασπορά σ^2. Τότε η κατανομή του καν/νου αθροίσματος $Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\,\sigma}$ τείνει καθώς $n \to \infty$ στην τυπική κανονική κατανομή δηλαδή

$$\boxed{f_{Z_n}(z) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}} \qquad \forall z \in \mathbb{R}$$