

Κάποια ντετερμινιστικά αποτελέσματα στις πιθανότητες

Δασκαλόπουλος Κυριάκος

Υπο κατασκευή

1η Έκδοση: Σεπτέμβριος 2020

Εισαγωγή

Οι νόμοι των μεγάλων αριθμών (Ασθενής και Ισχυρός), καθώς και το Κεντρικό Οριακό Θεώρημα μας επιτρέπουν να βγάσουμε ντετερμινιστικά συμπεράσματα για φαινόμενα τα οποία είναι στην βάση τους στοχαστικά. Το γεγονός αυτό είναι ο λόγος που η Θεωρία Πιθανοτήτων έχει γνωρίσει, από την δεκαετία του 1930 μέχρι σήμερα, τεράστια ανάπτυξη.

Θα ξεκινήσουμε με ένα παράδειγμα. Έστω ότι ένας παίκτης πάει σε ένα καζίνο και παίζει ρουλέτα, ποντάροντας 1 ευρώ στο κόκκινο σε κάθε γύρο. Θέλουμε να καταλάβουμε αν με την συγκεκριμένη διαδικασία θα έχει κέρδος.

Θέτουμε X_n το κέρδος του παίκτη στον γύρο n (μπορεί να είναι και αρνητικό). Η πιθανότητα να έρθει κόκκινο στην ρουλέτα είναι $18/37$ ¹. Άρα η X_n είναι μία τυχαία μεταβλητή για κάθε n , με τύπο

$$X_n = \begin{cases} X_n + 1, p = 18/37 \\ X_n - 1, p = 19/37 \end{cases}$$

Μας ενδιαφέρει η ποσότητα $X_1 + \dots + X_n$ (το συνολικό κέρδος μέχρι και την n -οστή ρίψη) ή καλύτερα η

$$\bar{X}_n := \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

δηλαδή το ποσοτό κέρδους στους πρώτους n γύρους.

Έστω ότι, για να απαντήσουμε στο ερώτημα, παρακολουθούμε τον παίκτη, και ύστερα από 20 γύρους, η ρουλέτα έφερε 14 κόκκινα και 6

¹Η ρουλέτα έχει συνολικά 37 αριθμούς, από τους οποίους 18 είναι κόκκινοι, 18 είναι μαύροι και ένας είναι πράσινος. Όταν ποντάρουμε να έρθει κόκκινο, πρέπει να έρθει ένας από τους 18 κόκκινους για να κερδίσουμε.

μάυρα, και έτσι το συνολικό του κέρδος είναι +8 ευρώ. Υπολογίζουμε το ποσοστό κέρδους στους 20 γύρους και βρίσκουμε ότι

$$\bar{X}_n = 14/20 = 70\%$$

Εκ πρώτης όψεως, φαίνεται ότι η διαδικασία είναι κερδοφορα. Όμως, η διαίσθηση μας λέει ότι, αν για παράδειγμα ποντάρουμε να έρθει κορώνα κατα την ρίψη ενός κάλπικου νόμισματος με πιθανότητα να έρθει κορώνα $p = 0.483 = 18/37$, τότε μάλλον θα χάσουμε(αν όχι την πρώτη φορά, σίγουρα μετά απο μερικές ρίψεις).

Οι νόμοι των μεγάλων αριθμών ερχονται να επιβεβαιώσουν την διαίσθηση μας(σπάνιο στις θετικές επιστήμες) απαντώντας στο παραπάνω φαινομενικό "παράδοξο" με την εύρεση της οριακής συμπεριφοράς της ποσότητας $\bar{X}_n$. Εδώ θα ασχοληθούμε μόνο με τον ασθενή νόμο των μεγάλων αριθμών. Η διαφορά του με τον Ισχυρό είναι το είδος της σύγκλισης που μας δίνει.

Εμπειρικά αποτελέσματα με την γλώσσα R

Πριν προχωρήσουμε στις απόδειξεις, θα δούμε κάποια αποτελέσματα που θα μας βοηθήσουν να κατανοήσουμε διαισθητικά τα επόμενα αποτελέσματα.

Ένας εύκολος τρόπος να καταλάβουμε την οριακή συμπεριφορά του αριθμητικού μέσου $\bar{X}_n$ δεδομένης κάποιας κατανομής, είναι να πάρουμε διαδοχικά δείγματα $(Y_{i_1}, \dots, Y_{i_k})$ αυξανόμενου μεγέθους (για παράδειγμα $Y_{100} = (X_1, X_2, \dots, X_{100})$ είναι ένα δείγμα μεγέθους 100) και να παρατηρήσουμε αν η ακολουθία των αριθμητικών μέσων $(\bar{X}_{n1}, \dots, \bar{X}_{nk})$, όπου $(\bar{X}_{nm} = (\sum_{j=1}^{i_m} X_j)/n)$, παρουσιάζει κάποιο μοτίβο όσο το n αυξάνει.

Για παράδειγμα, έστω ότι θέλουμε να εξετάσουμε το παραπάνω για την τυπική κανονική κατανομή. Με τον ακόλουθο κώδικα θα πάρουμε διαδοχικά δείγματα από μία τυχαία μεταβλητή $X \sim \mathcal{N}(0, 1)$ μεγέθους 10, 100, 1000 και 10000 αντίστοιχα, και ύστερα θα εκτυπώσουμε τις τιμές $\bar{X}_{nm}, i_m = 10, 100, 1000, 10000$.

```
y<-rnorm(10)
x<-rnorm(100)
w<-rnorm(1000)
z<-rnorm(10000)
mean(y)
mean(x)
mean(w)
mean(z)
```

```
[1] 0.2147669
```

```
[1] -0.03125359
```

```
[1] -0.06155245
```

```
[1] 0.003833976
```

Βλέπουμε ότι όσο ο αριθμός των παρατηρήσεων που έχουμε μεγαλώνει, τόσο ο αριθμητικός μέσος πλησιάζει το 0, δηλαδή την μέση τιμή της κατανομής.

Παρατήρηση Να σημειωθεί εδώ ότι αν X_i είναι n τυχαίες μεταβλητές με $X_i \sim \mathcal{N}(\mu, \sigma^2)$ τότε $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$.

Εφαρμόζουμε τα παραπάνω για την κατανομή Poisson(3).

```
y<-rpois(10, lambda = 3)
x<-rpois(100, lambda = 3)
w<-rpois(1000, lambda = 3)
z<-rpois(10000, lambda = 3)
mean(y)
mean(x)
mean(w)
mean(z)
```

```
[1] 3.2
```

```
[1] 3.16
```

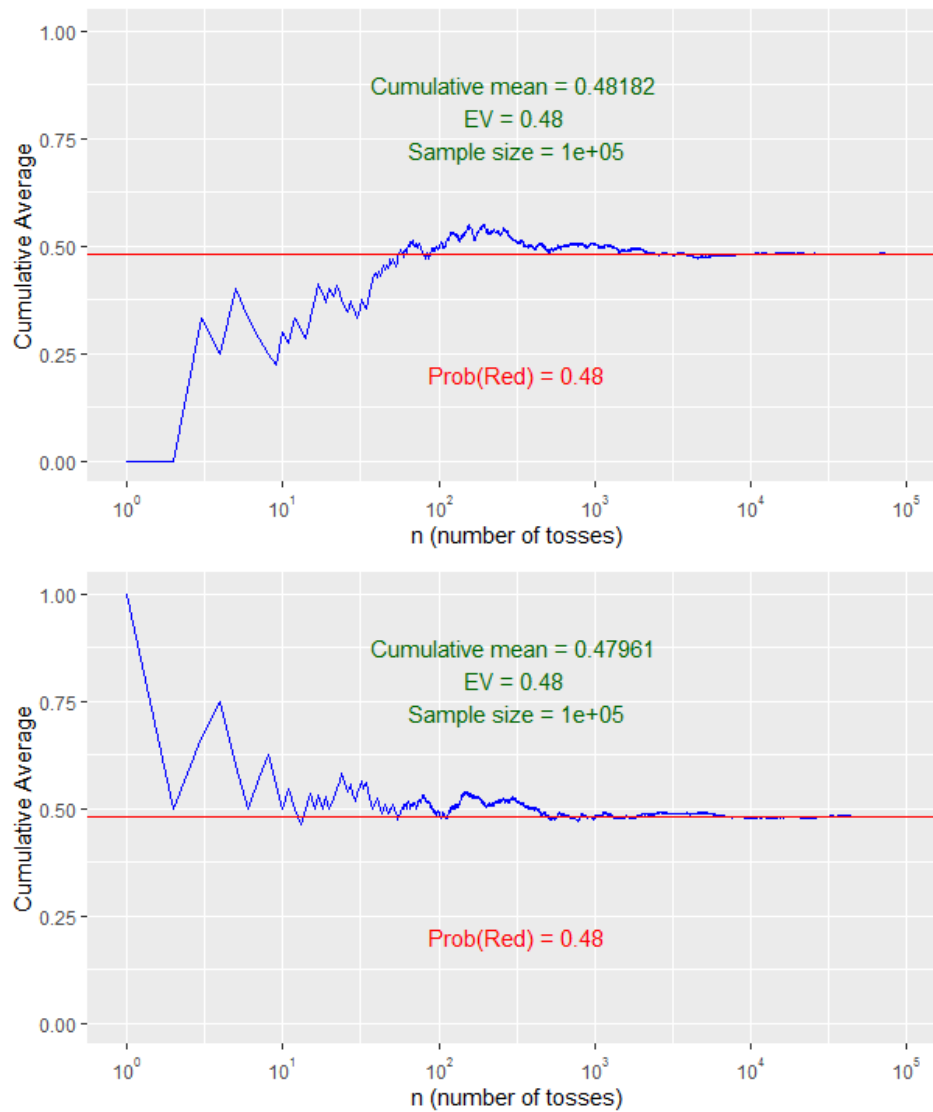
```
[1] 3.087
```

```
[1] 2.986
```

Πάλι παρατηρούμε, ότι η απόλυτη τιμή της απόστασης του αριθμητικού μέσου από την μέση τιμή της κατανομής, μικραίνει όσο μεγαλώνει το πλήθος των παρατηρήσεων. Δηλαδή για δείγμα μεγέθους 10 το "σφάλμα" είναι 0.2, ενώ για δείγμα μεγέθους 10000 είναι 0.014.

Επιστρέφοντας στο πρόβλημα του παίκτη ρουλέτας, θέλουμε να καταλάβουμε το μακροχρόνιο ποσοστό κέρδους, δηλαδή το που περίπου θα "συγκλίνει" η ακολουθία $(\bar{X}_n)_{n \in \mathbb{N}}$. Με τα παρακάτω γραφήματα(ο

κώδικας δίνεται στο τέλος) βλέπουμε ότι όσο το n μεγαλώνει, η τιμή του αριθμητικού μέσου, ανεξάρτητα από την αρχική απόκλιση που παρουσιάζει, είναι με μεγάλη ακρίβεια κοντά στο 0.48.



Προβλήματα

Το παραπάνω γεγονός δεν ισχύει για όλες τις κατανομές. Για παράδειγμα, για την κατανομή Cauchy(κατανομή για την οποία δεν ορίζεται μέση τιμή) η ποσότητα $\bar{X}_n$ δεν πλησιάζει κάποια σταθερά καθώς το $n \rightarrow \infty$. Όπως βλέπουμε εμπειρικά στον παρακάτω κώδικα, οι τιμές της δεν παρουσιάζουν κάποιο μοτίβο(δεν έχουν κάποια "προτίμηση") όσο μεγαλώνει το δείγμα.

```
y<-rcauchy(100, location= 0, scale=1)
x<-rcauchy(1000, location= 0, scale=1)
w<-rcauchy(9999, location= 0, scale=1)
z<-rcauchy(10000, location= 0, scale=1)
e<-rcauchy(99999, location= 0, scale=1)
q<-rcauchy(1000000, location= 0, scale=1)
p<-rcauchy(5362218, location= 0, scale=1)
mean(y)
mean(x)
mean(w)
mean(z)
mean(e)
mean(q)
mean(p)
```

```
[1] 0.37857
```

```
[1] 0.8043228
```

```
[1] 0.4393708
```

```
[1] 0.7667333
```

```
[1] 0.2473142
```

```
[1] -2.101454
```

```
[1] 0.8399794
```

Άρα βλέπουμε ότι πρέπει να βάλουμε κάποιους περιορισμούς στις κα-

τανομές απο τις οποίες παίρνουμε δείγμα, ώστε η ακολουθία $\bar{X}_n$ να συγκλίνει σε κάποιον αριθμό(θα δούμε παρακάτω τι σύγκλιση θα έχουμε). Στην επόμενη παράγραφο θα αναπτυχθούν όλα αυτά αυστηρά.

Παρατήρηση Να σημειωθεί ότι η μοντελοποίηση μέσω της R , ενώ είναι χρήσιμη και μας βοηθάει να κατανοήσουμε κάποια αποτελέσματα, δεν πρέπει να χρησιμοποιείται ως απόδειξη. Μας παρέχει μία εμπειρική εποπτεία που μας προϊδεάζει για κάποιο αποτέλεσμα, η οποία όμως πολλές φορές οδηγεί σε σοβαρά λάθη όταν δεν υπάρχει το θεωρητικό υπόβαθρο.

Αποδείξεις

Για να αποδείξουμε τον ασθενή νόμο, πρέπει πρώτα να δείξουμε την ανισότητα του Chebyshev, η οποία είναι ειδική περίπτωση της ανισότητας Markov.

Στα παρακάτω X θα είναι μία τυχαία μεταβλητή με $\mathbb{E}(X) < \infty$.

Λήμμα (Ανισότητα Markov) Έστω X μη αρνητική και t ένας αριθμός. Τότε ισχύει

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t} \quad (1)$$

Απόδειξη: Ορίζουμε Y τυχαία μεταβλητή ως εξής:

$$Y = \begin{cases} 0, & X < t \\ t, & X \geq t \end{cases}$$

Άρα ισχύει ότι

$$\mathbb{E}(Y) = t\mathbb{P}(Y = t) + 0\mathbb{P}(Y = 0) = t\mathbb{P}(Y = t) = t\mathbb{P}(X \geq t)$$

Όμως $X \geq Y$, και επομένως $\mathbb{E}(X) \geq \mathbb{E}(Y)$. Τελικά καταλήγουμε στο ότι

$$\mathbb{E}(X) \geq \mathbb{E}(Y) = t\mathbb{P}(X \geq t)$$

απο την οποία έπεται το ζητούμενο.

Έστω τώρα Y τυχαία μεταβλητή με $\mathbb{E}(Y) = \mu$ και $Var(Y) < \infty$. Θέτοντας $X = (Y - \mu)^2$, $k = t^2$ στην 1 και απλοποιώντας τα τετράγωνα μέσα στην πιθανότητα παίρνουμε ότι

$$\mathbb{P}(|Y - \mu| \geq t) \leq \frac{\mathbb{E}((Y - \mu)^2)}{t^2} := \frac{Var(Y)}{t^2} \quad (2)$$

Η παραπάνω ανισότητα ονομάζεται ανισότητα του Chebyshev

Θεώρημα 1 (Ασθενής Νόμος των μεγάλων αριθμών)

Θεωρούμε n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, \dots, X_n$. Έστω $\mathbb{E}(X_1) = \mu$, $\text{Var}(X_1) = \sigma^2$ και $\bar{X} = (\sum_{i=1}^n X_i)/n$. Τότε, για κάθε $\varepsilon > 0$ υπάρχουν ακέραιος n και $\delta > 0$ τέτοια ώστε

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) < \delta \quad (3)$$

Απόδειξη:

Είναι εύκολο να δούμε ότι

$$\mathbb{E}(\bar{X}) = \frac{1}{n} \mathbb{E}(\sum_{i=1}^n X_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} n\mu = \mu$$

και

$$\text{Var}(\bar{X}_n) := \mathbb{E}\{(\bar{X}_n - \mathbb{E}(\bar{X}_n))^2\} = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}$$

Εφαρμόζοντας την 2 για την τυχαία μεταβλητή $\bar{X}_n - \mu$, παίρνουμε την σχέση

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n - \mu)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

Για να δείξουμε το θεώρημα αρκεί να επιλέξουμε $n > \sigma^2/\delta\varepsilon^2$.

Το παραπάνω θεώρημα ουσιαστικά μας λέει ότι αν έχουμε n δεδομένα από μία κατανομή X , τότε ο αριθμητικός μέσος τους είναι μία πολύ καλή εκτίμηση της μέσης τιμής της X .

*Εναλλακτική απόδειξη μέσω χαρακτηριστικών

Εκτός από την παραπάνω μέθοδο, μπορούμε να αποδείξουμε το Θεώρημα 1 με την βοήθεια του ακόλουθου θεωρήματος, του οποίου την απόδειξη μπορεί κανείς να την βρει στο [1](Chapter 19) ή στα ελληνικά στο [2](Κεφάλαιο 15). Εδώ το παρουσιάζουμε σε απλοποιημένη μορφή.

Θεώρημα 2 (Θεώρημα συνέχειας του Levy) Έστω $X_n, n \geq 1$ και X τυχαίες μεταβλητές με κατανομές F_{X_n}, F_X και χαρακτηριστικές συναρτήσεις $\phi_{X_n}(t) = \mathbb{E}(e^{itX_n}), \phi_X(t) = \mathbb{E}(e^{itX})$ αντίστοιχα. Τότε ισχύει το εξής:

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \phi_X(t) \implies \lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

σε όλα τα σημεία συνέχειας της F_X .

Το Θεώρημα 1 είναι ισοδύναμο με το

Θεώρημα 3 Θεωρούμε n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, \dots, X_n$. Έστω $\mathbb{E}(X_1) = \mu$, $Var(X_1) = \sigma^2$ και $\bar{X} = (\sum_{i=1}^n X_i)/n$. Τότε, για κάθε $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0 \quad (4)$$

Αυτο το βλέπουμε εύκολα απο τον ορισμό του ορίου ακολουθίας. Δηλαδή,

$$4 \implies \forall \delta > 0, \exists N > 0 : \forall n \geq N, \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) < \delta$$

Η τελευταία ανισότητα είναι η 3.

Πριν προχωρήσουμε στην απόδειξη της 4, αποδεικνύουμε το ακόλουθο χρήσιμο για αυτήν Λήμμα:

Λήμμα Έστω X τυχαία μεταβλητή με $\mathbb{E}(X) < \infty$. Τότε

$$\lim_{t \rightarrow 0} \frac{\log \phi_X(t) - i\mu t}{t} = 0 \quad (5)$$

Απόδειξη: Η $\phi_X(t)$ είναι παραγωγίσιμη και $\phi_X(0) = 1$. Άρα η $\log \phi_X(t)$ είναι καλά ορισμένη και $\log \phi_X(0) = 0$. Απο την σχέση $\phi_X^{(n)}(t) = i^n \mathbb{E}(X^n)$ παίρνουμε $\phi_X'(0) = i\mu$. Άρα,

$$\log \phi_X'(0) = \frac{\phi_X'(0)}{\phi_X(0)} = i\mu$$

Επομένως,

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\log \phi_X(t) - i\mu t}{t} &= \lim_{t \rightarrow 0} \frac{\log \phi_X(t)}{t} - i\mu = \lim_{t \rightarrow 0} \frac{\log \phi_X(t) - \log \phi_X(0)}{t - 0} - i\mu = \\ &= \log \phi_X'(0) - i\mu = 0 \end{aligned}$$

Απόδειξη: (της 4) Η χαρακτηριστική συνάρτηση της $\bar{X}_n - \mu$ είναι η

$$\begin{aligned} \phi_{\bar{X}_n - \mu}(t) &= \mathbb{E}(e^{it(\bar{X}_n - \mu)}) = \mathbb{E}(e^{it\bar{X}_n} e^{-it\mu}) = \\ &= e^{-it\mu} (\phi_{X_1}(t/n))^n = \exp(n(\log \phi_{X_1}(t/n)) - i\mu(t/n)) \end{aligned}$$

Θα δείξουμε ότι $\lim_{n \rightarrow \infty} n(\log \phi_{X_1}(t/n) - i\mu(t/n)) = 0$.

Για $t = 0$ η ισότητα ισχύει. Για $t \neq 0$, γράφουμε το όριο ως

$$t \lim_{n \rightarrow \infty} \frac{\log \phi_{X_1}(t/n) - i\mu(t/n)}{t/n}$$

Αφού $t/n \xrightarrow{n \rightarrow \infty} 0$, έχουμε απο το Λήμμα 2 ότι το παραπάνω όριο είναι 0 (κάνουμε την αλλαγή μεταβλητής $t/n = k$, οπότε $k \xrightarrow{n \rightarrow \infty} 0$). οπότε

$$\lim_{n \rightarrow \infty} \phi_{\bar{X}_n - \mu}(t) = 1$$

Όμως 1 είναι η χαρακτηριστική της τετριμμένης τυχαίας μεταβλητής $X(\omega) = 0 \forall \omega \in \Omega$. Δηλαδή έχουμε $\lim_{n \rightarrow \infty} \phi_{\bar{X}_n - \mu}(t) = \phi_X(t)$. Απο το Θεώρημα 2 ισχύει ότι

$$\lim_{n \rightarrow \infty} \mathbb{P}(\bar{X}_n - \mu \leq \varepsilon) := \lim_{n \rightarrow \infty} F_{\bar{X}_n - \mu}(\varepsilon) = F_X = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

και αφού $\varepsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(\bar{X}_n - \mu \leq \varepsilon) = 1 \implies \lim_{n \rightarrow \infty} \mathbb{P}(\bar{X}_n - \mu > \varepsilon) = 0$.

Τέλος, προφανώς έχουμε $\lim_{n \rightarrow \infty} \mathbb{P}(\bar{X}_n - \mu \leq -\varepsilon) = F_X(-\varepsilon) = 0$, το οποίο ολοκληρώνει την απόδειξη.

Λήμμα (Borel-Cantelli)

Θεώρημα 4 (Ισχυρός Νόμος των μεγάλων αριθμών) Έστω $X_n, n \geq 1$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές ορισμένες στον ίδιο χώρο, με $\mathbb{E}(X_i) = \mu$ και $Var(X_i) = \sigma^2 < \infty$. Τότε

$$\lim_{n \rightarrow \infty} \bar{X}_n = \mu$$

$\mathbb{P}$ -σχεδόν παντού.

Απόδειξη. Ορίζουμε $E = \{\omega \in \Omega \mid \lim_{n \rightarrow \infty} \bar{X}_n = \mu\}$ και θα αποδείξουμε ότι $\mathbb{P}(E) = 1$.

Ένα ενδεχόμενο δεν ανήκει στο E αν ισχύει $|Z_n| > \varepsilon$ για κάποιο $\varepsilon > 0$ και για άπειρα $n \in \mathbb{N}$. Δηλαδή, αν και μόνο αν $\omega \in \limsup_n A_n$, $A_n = \{|Z_n| > n\varepsilon\}$. Απο το λήμμα Borel-Cantelli λοιπόν, αρκεί να δείξουμε ότι $\sum_n \mathbb{P}(A_n) < \infty$.

γενικευμένη ανισότητα τσεμπισιοφ

$\mathbb{E}(Z_n^4)$ Άρα

$$\mathbb{P}(A_n) = \mathbb{P}(|Z_n| > n\varepsilon) \leq \frac{1}{(n\varepsilon)^4} \mathbb{E}(Z_n^4) \leq \frac{C}{\varepsilon^4} \frac{1}{n^2}$$

Επομένως

$$\sum_{n \geq n_0} \mathbb{P}(A_n) \leq \frac{C}{\varepsilon^4} \sum_{n \geq n_0} \frac{1}{n^2} < \infty$$

□

Αντιπαραδείγματα

Κατανομή Cauchy

Ορισμός Έστω X πραγματική τυχαία μεταβλητή με στήριγμα το $\mathbb{R}$ και συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\beta\pi} \frac{\beta}{\beta + (x - \alpha)^2}$$

όπου α, β σταθερές τέτοιες ώστε $-\infty < \alpha < \infty, 0 < \beta < \infty$.

Τότε λέμε ότι η X έχει την κατανομή Cauchy με παραμέτρους α και β .

Ένα γεγονός που κάνει την κατανομή Cauchy ενδιαφέρουσα είναι ότι, για $\alpha = 0$ και $\beta = 1$, η μέση της τιμής *δεν υπάρχει*.

Για να το δούμε αυτό παρατηρούμε ότι $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$, όπου $X^+ = \max(X, 0)$ και $X^- = -\min(X, 0)$.

Η συνάρτηση πυκνότητας πιθανότητας γίνεται

$$f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$$

Έχουμε λοιπόν ότι:

$$\mathbb{E}(X^+) = \int_0^\infty x f_X(x) dx \geq \frac{1}{\pi} \int_1^\infty x f_X(x) dx \geq \frac{1}{\pi} \int_1^\infty \frac{1}{2x} dx = \infty$$

όπου χρησιμοποιήθηκαν διαδοχικά τα εξής:

1. Ισχύει ότι $X < 0$ για $x < 0$
2. $\frac{x}{1+x^2} \geq 0, x \geq 0$
3. $\frac{x}{1+x^2} \geq \frac{1}{2x}, x > 1$

Ομοίως:

$$\mathbb{E}(X^-) = \int_{-\infty}^0 \frac{1}{\pi} \frac{x}{1+x^2} dx = \int_0^{\infty} \frac{1}{\pi} \frac{y}{1+y^2} dy = \infty$$

όπου στην δεύτερη ισότητα έγινε η αλλαγή μεταβλητής $y = -x$. Άρα καταλήγουμε στο ότι

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) = \infty - \infty$$

,δηλαδή σε απροδιοριστία. Έτσι μέση τιμή της κατανομής δεν υπάρχει. Η έκφραση $\bar{X}_n \xrightarrow{n \rightarrow \infty} \mathbb{E}(X)$ δεν έχει κάποιο νόημα, δηλαδή δεν ισχύει σε αυτήν την περίπτωση ο ασθενής νόμος των μεγάλων αριθμών. Ένας άλλος τρόπος να το δούμε αυτό είναι κάνοντας την ακόλουθη παρατήρηση.

Πρόταση Έστω $X, Y \sim \text{Cauchy}(0, 1)$ ανεξάρτητες τυχαίες μεταβλητές. Τότε η $\frac{1}{2}(X + Y)$ ακολουθεί και αυτή την κατανομή $\text{Cauchy}(0, 1)$.

Απόδειξη. Η εύρεση της κατανομής του αθροίσματος $X + Y$ απαιτεί κάποια δουλειά. Μπορεί κανείς να την βρεί στο [3]. Η συνάρτηση πυκνότητας πιθανότητας είναι η $f_{X+Y}(x) = 2/\pi(4 + x^2)$.

Ύστερα βρίσκουμε την κατανομή του κλάσματος $A = \frac{1}{2}(X + Y)$ ως εξής:

$$F_A(x) = \mathbb{P}(A \leq x) = \mathbb{P}\left(\frac{1}{2}(X + Y) \leq x\right) = F_{X+Y}(2x)$$

Τότε, $f_A(x) = F'_A(x) = F'_{X+Y}(2x) = 2f_{X+Y}(2x) = \frac{1}{\pi} \frac{4}{4+4x^2} = \frac{1}{\pi} \frac{1}{1+x^2}$, η οποία είναι η σ.π. της κατανομής $\text{Cauchy}(0, 1)$. $\square$

Εφαρμόζοντας το παραπάνω αποτέλεσμα $n - 1$ φορές έχουμε ότι ο α-ριθμητικός μέσος $\text{Cauchy}(0, 1)$ τυχαίων μεταβλητών ακολουθεί την κατανομή $\text{Cauchy}(0, 1)$. Θέλουμε να εξετάσουμε αν ισχύει ότι $\bar{X}_n \xrightarrow{n \rightarrow \infty} \mathbb{E}(X)$. Όμως κάθε όρος της ακολουθίας $(\bar{X}_n)_{n \in \mathbb{N}}$ είναι, με βάση τα παραπάνω, τυχαία μεταβλητή με κατανομή $\text{Cauchy}(0, 1)$, και άρα, όσο και να μεγαλώνει το δείγμα, δεν αλλάζει η διασπορά των τιμών της ακολουθίας (για παράδειγμα, στην περίπτωση της κανονικής ισχύει $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$, με

$\sigma^2/n \xrightarrow{n \rightarrow \infty} 0$). Δηλαδή, δεν μπορεί να ισχύει ότι η ακολουθία $(\bar{X}_n)_{n \in \mathbb{N}}$ συγκλίνει.

Παρατήρηση Έστω κατανομή F για την οποία ισχύει ότι $\bar{X}_n \sim F$. Θεωρούμε την $\mathbb{E}(\bar{X}_n)$ ως μία τυχαία μεταβλητή με κατανομή F_Y τέτοια ώστε

$$Y = \begin{cases} \mathbb{E}(X), p = 1 \\ 0, p = 0 \end{cases}$$

Αφού $\bar{X}_n \sim F$ και $F \neq F_Y$ δεν γίνεται να ισχύει $\bar{X}_n \xrightarrow{P} \mathbb{E}(X)$. Καταλήξαμε στο ότι για τις κατανομές F των οποίων ο αριθμητικός μέσος οποιουδήποτε δείγματος ακολουθεί πάλι την κατανομή F (με τις ίδιες παραμέτρους!), δεν μπορεί να ισχύει ο Ασθενής νόμος των Μεγάλων Αριθμών.

Σύγκλιση τυχαίων μεταβλητών

Κεντρικό Οριακό Θεώρημα

*Κώδικας R

[4]

```
library(ggplot2)
library(scales)
tossCoin = function(n=30, p=0.48) {

  # create a probability distribution , a vector of outcomes (H/T are coded
  # using 0/1)
  # and their associated probabilities
  outcomes = c(0,1) # sample space
  probabilities = c(1-p,p)

  # create a random sample of n flips; this could also be done with
  # the rbinom() function , but sample() is perhaps more useful
  flips = sample(outcomes,n,replace=T,prob=probabilities)

  # now create a cumulative mean vector
  cum_sum = cumsum(flips)
  index = c(1:n)
  cum_mean = cum_sum / index

  # now combine the index , flips and cum_mean vectors
  # into a data frame and return it
  # return(data.frame(index ,flips ,cum_mean))
  return(data.frame(index,cum_mean))
}

ggplotCoinTosses = function(n=30, p=.48) {
  # visualize how cumulative average converges on p
  # roll the dice n times and calculate means
  trial1 = tossCoin(n,p)
  max_y = ceiling(max(trial1$cum_mean))
  if (max_y < .75) max_y = .75
  min_y = floor(min(trial1$cum_mean))
  if (min_y > .4) min_y = .4
```

```

# calculate last mean and standard error
last_mean = round(trial1$cum_mean[n],9)

# plot the results together
plot1 = ggplot(trial1, aes(x=index,y=cum_mean)) +
  geom_line(colour = "blue") +
  geom_abline(intercept=0.48,slope=0, color = 'red', size=.5) +
  theme(plot.title = element_text(size=rel(1.5)),
        panel.background = element_rect()) +
  labs(x = "n (number of tosses)",
       y = "Cumulative Average") +
  scale_y_continuous(limits = c(min_y, max_y)) +
  scale_x_continuous(trans = "log10",
                    breaks = trans_breaks("log10",function(x) 10^x),
                    labels = trans_format("log10",math_format(10^.x))) +
  annotate("text",
         label=paste("Cumulative mean=", last_mean,
                    "\nEV=", p,
                    "\nSample size=", n),
         y=(max_y - .20),
         x=10^(log10(n)/2), colour="darkgreen") +
  annotate("text",
         label=paste("P(Heads) with Coin=0.48"),
         y=(max_y - .80),
         x=10^(log10(n)/2), colour="red")

return(plot1)
}
# call the function; let's use a coin
ggplotCoinTosses(100000, .48)

```

Βιβλιογραφία

- [1] Jacod, J., Protter, P. (2004). Probability Essentials. In Universitext. Springer Berlin Heidelberg.
- [2] Χελιώτης, Δ., 2015. Ένα δεύτερο μάθημα στις πιθανότητες. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/2825>
- [3] Dwass, M. (1985). On the Convolution of Cauchy Distributions. The American Mathematical Monthly, 92(1), 55.
- [4] <https://rpubs.com/stephenmoore56/301283>
- [5] P., Hoel, S., Port, C. Stone(1971), Introduction to Probability Theory, Μεταφρασμένο στα ελληνικά(2002): Εισαγωγή στην Θεωρία Πιθανοτήτων, Γιαννόπουλος Απόστολος, ISBN:978-960-524-156-8