



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

Ασυμπτωτικά Θεωρήματα για Μεγάλες Αποκλίσεις

Διπλωματική Εργασία

Σπυρόπουλος Χρήστος

Επιβλέπων Καθηγητής: Μιχάλης Λουλάκης

Αθήνα, Οκτώβριος 2015



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

Ασυμπτωτικά Θεωρήματα για Μεγάλες Αποκλίσεις

Διπλωματική Εργασία

Σπυρόπουλος Χρήστος

Τριμελής επιτροπή

Μιχάλης Λουλάκης(Επιβλέπων)

Ιωάννης Σπηλιώτης

Δημήτρης Φουσκάκης

Αθήνα, Οκτώβριος 2015

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα της διπλωματικής εργασίας μου, κ. Μιχάλη Λουλάκη, Επίκουρο Καθηγητή του Εθνικού Μετσόβιου Πολυτεχνείου, για την βοήθεια και την καθοδήγηση που μου προσέφερε καθόλη τη διάρκεια της εργασίας. Ακόμα θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου για την βοήθεια που μου παρείχαν όλα τα χρόνια των σπουδών μου.

Πρόλογος

Στην εργασία αυτή θα ασχοληθούμε με το θέμα ασυμπτωτικά θεωρήματα για μεγάλες αποκλίσεις, και θα ασχοληθούμε κυρίως με το Θεώρημα Cramer, το οποίο υπολογίζει το βέλτιστο ρυθμό εκθετικής μείωσης της πιθανότητας πολύ σπάνιων ενδεχομένων. Ουσιαστικά πόσο γρήγορα η πιθανότητα αυτών των ενδεχομένων πηγαίνει στο 0.

Η Αρχή των μεγάλων αποκλίσεων συνίσταται στην εκτίμηση σπάνιων ενδεχομένων, δηλαδή ενδεχομένων που αποκλίνουν από την τυπική συμπεριφορά στο θεωρούμενο χώρο πιθανότητας. Το γεγονός αυτό την καθιστά χρήσιμη σε πολλές διαφορετικές επιστημονικές περιοχές που θα αναφέρουμε αργότερα.

Συνεπώς αρχικά θα αναφερθούμε σε βασικές έννοιες όπως το σπάνιο ενδεχόμενο, θα ορίσουμε την συνάρτηση ταχύτητας $I(x)$ έτσι ώστε να έχουμε όλα τα εφόδια για να αποδείξουμε το Θεώρημα του Cramer για ανέξαρτητες και ισόνομες τυχαίες μεταβλητές. Στην συνέχεια θα αναφέρουμε στο πρώτο κεφάλαιο κάποιες εφαρμογές του Θεωρήματος Cramer σε γνωστές κατανομές όπως η εκθετική και άλλες.

Στο δεύτερο κεφάλαιο θα αναφερθούμε στην Αρχή των μεγάλων αποκλίσεων. Αρχικά θα δώσουμε τον ορισμό της και στη συνέχεια θα παραθέσουμε το Λήμμα Varadhan και το Θεώρημα Gartner-Ellis.

Στο τρίτο κεφάλαιο με τη χρήση του μαθηματικού πακέτου της R θα προσπαθήσουμε να υπολογίσουμε τον αριθμό $\pi \approx 3,14$, για να το κάνουμε αυτό θα χρησιμοποιήσουμε δύο μεθόδους την Monte Carlo και την Importance Sampling.

Ακόμα στο τελευταίο κεφάλαιο θα μιλήσουμε για τις Μαρκοβιανές Αλυσίδες, θα δώσουμε τον ορισμό τους και θα αναφέρουμε μερικές εφαρμογές τους στις σύγχρονες επιστήμες. Στο τέλος θα αποδείξουμε το Θεώρημα Cramer για Μαρκοβιανές Αλυσίδες με την βοήθεια των Θεωρημάτων Gartner-Ellis και Perron-Frobenius.

Περιεχόμενα

Κεφάλαιο 1	Θεώρημα Cramer και Εφαρμογές	
1.1	Συνάρτηση ταχύτητας $I(x)$	6
1.2	Θεώρημα Cramer	14
1.3	Εφαρμογές του Θεωρήματος Cramer σε γνωστές Κατανομές	21
Κεφάλαιο 2	Αρχή Μεγάλων Αποκλίσεων	
2.1	Αρχή των μεγάλων αποκλίσεων	31
2.2	Απόδειξη LDP με χρήση Θεωρήματος Cramer	32
2.3	Λήμμα Varadhan	34
2.4	Θεώρημα Gartner-Elis	36
Κεφάλαιο 3	Προσομοίωση σπάνιων ενδεχομένων	
3.1	Μέθοδος Monte Carlo	41
3.2	Τεχνική της δειγματοληψίας κατά σπουδαιότητα	57
Κεφάλαιο 4	Θεώρημα Cramer για Μαρκοβιανές Αλυσίδες	
4.1	Μαρκοβιανές Αλυσίδες	64
4.2	Εφαρμογές Μαρκοβιανών Αλυσίδων στις Σύγχρονες Επιστήμες	66
4.3	Εργοδικό Θεώρημα για Μαρκοβιανές Αλυσίδες	70
4.4	Θεώρημα Perron-Frobenius	71
4.5	Θεώρημα Cramer για Μαρκοβιανές Αλυσίδες	73
Βιβλιογραφία		75

Κεφάλαιο 1

Θεώρημα Cramer και εφαρμογές

1.1 Συνάρτηση ταχύτητας $I(x)$

Η Αρχή μεγάλων αποκλίσεων πραγματεύεται την εκτίμηση της πιθανότητας πολύ σπάνιων ενδεχομένων. Αν λόγου χάρη πάρουμε μία ακολουθία ανεξάρτητων ισόνομων τυχαίων μεταβλητών $\{X_i\}_{i \in \mathbb{N}}$ με μηδενική μέση τιμή, παρατηρούμε από τον ασθενή νόμο των μεγάλων αριθμών ότι η πιθανότητα ο μέσος όρος τους να βρίσκεται μακριά από το 0 μικραίνει όσο το n μεγαλώνει

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i\right| \geq \delta\right) \rightarrow 0$$

Πόσο γρήγορα μικραίνει όμως αυτή η πιθανότητα, δηλαδή επί της ουσίας τι μπορούμε να πούμε για την ταχύτητα σύγκλισης αυτής της πιθανότητας στο 0;

Ουσιαστικά αυτό είναι το πεδίο της Αρχής των μεγάλων αποκλίσεων. Να μελετά την πιθανότητα σπάνιων ενδεχομένων και το πόσο γρήγορα τρέχει η πιθανότητα τους στο 0. Το γεγονός αυτό την κάνει πολύ χρήσιμη σε διάφορες άλλες επιστήμες όπως Στατιστική, Θεωρία Πληροφορίας, Τηλεπικοινωνίες και Στατιστική Μηχανική. Βέβαια επειδή είναι χρήσιμη σε τόσες διαφορετικές επιστήμες αντιμετωπίζει ανάλογα μεγάλη γκάμα διαφορετικών προβλημάτων οπότε είναι λογικό στο εκάστοτε πρόβλημα να χρειάζεται να χρησιμοποιηθεί διαφορετική μέθοδος επίλυσης. Αρχικά μπορούμε να παραθέσουμε τον ορισμό του σπάνιου ενδεχομένου.

Ορισμός 1.1

Έστω ακολουθία $\{Y_n\}_{n \in \mathbb{N}}$ πραγματικών τυχαίων μεταβλητών ορισμένων σε κάποιο χώρο πιθανότητας (Ω, \mathcal{L}, P) τέτοια ώστε να συγκλίνει κατά πιθανότητα σε κάποια εκφυλισμένη τυχαία μεταβλητή $Y \rightarrow y \in \mathbb{R}$. Ένα Borel ενδεχόμενο $A \subseteq \mathbb{R}$ θα λέγεται σπάνιο ενδεχόμενο για την $\{Y_n\}_{n \in \mathbb{N}}$ αν

$$P(Y_n \in A) \xrightarrow{n \rightarrow \infty} 0$$

Εμείς σε αυτό το κεφάλαιο θα μιλήσουμε για το Θεώρημα Cramer και κάποιες εφαρμογές του. Το Θεώρημα Cramer προσπαθεί να υπολογίσει την ταχύτητα σύγκλισης στο 0 όταν το A είναι σπάνιο.

Ορισμός 1.1.2

Η ροπογεννήτρια συνάρτηση μιας κατανομής μ στο \mathbb{R} είναι η συνάρτηση $M_\mu: \mathbb{R} \rightarrow [0, \infty]$

$$M_\mu(\lambda) = \int e^{\lambda x} d\mu(x)$$

και $A_\mu = \log(M_\mu)$ λέγεται η λογαριθμική ροπογεννήτρια της κατανομής μ .

Ορισμός 1.1.3

Το σύνολο $D_{M_\mu} := \{\lambda \in \mathbb{R} | M_\mu(\lambda) < \infty\}$ είναι το ουσιώδες πεδίο ορισμού της M_μ .

Ορισμός 1.1.4

Μια συνάρτηση $f: \mathbb{R} \rightarrow (-\infty, +\infty]$ λέγεται κυρτή αν για κάθε $x, y \in \mathbb{R}$ και κάθε $t \in [0, 1]$ έχουμε ότι

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

βέβαια η κυρτότητα εξασφαλίζει την συνέχεια της f μόνο στο εσωτερικό του ουσιώδους πεδίου ορισμού της D_f . Την συνέχεια της f σε όλο το D_f την εξασφαλίζει η συνθήκη της κάτω ημισυνέχειας.

Ορισμός 1.1.5

Μια συνάρτηση $f: \mathbb{R} \rightarrow (-\infty, +\infty]$ λέγεται κάτω ημισυνεχής αν το σύνολο

$$f^{-1}((-\infty, c]) = \{x \in \mathbb{R} | f(x) \leq c\}$$
 είναι κλειστό για κάθε $c \in \mathbb{R}$

Ορισμός 1.1.6

Έστω μ μια κατανομή στο \mathbb{R} . Η μ έχει εκθετικές ροπές αν $\int e^{\varepsilon|x|} d\mu(x) < \infty$ για κάποιο $\varepsilon > 0$

Ορισμός 1.1.7

Έστω μ κατανομή στο \mathbb{R} με εκθετικές ροπές. Θα λέμε συνάρτηση ταχύτητας της μ τη συνάρτηση

$I_\mu: \mathbb{R} \rightarrow [0, +\infty]$ με τύπο

$$I(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \log M_\mu(\lambda)) = \sup_{\lambda \in D_{M_\mu}} (\lambda x - \log M_\mu(\lambda))$$

Το σύνολο $D_{M_\mu} := \{\lambda \in \mathbb{R} | M_\mu(\lambda) < \infty\}$ είναι το ουσιώδες πεδίο ορισμού της M_μ .

Ακόμα θα ήταν χρήσιμο να αναφέρουμε τις ανισότητες Chebyshev και Jensen που θα χρησιμοποιήσουμε στην συνέχεια ως μαθηματικό εργαλείο σε πολλές αποδείξεις.

Ανισότητα Chebyshev

Δίνεται ότι η τυχαία μεταβλητή X έχει πεπερασμένη μέση τιμή $\mu = E[X]$ και διασπορά $\sigma^2 = Var[X]$. Τότε για κάθε $a > 0$ θα ισχύει ότι

$$P\{|X - \mu| \geq a\} \leq \frac{Var(X)}{a^2}$$

Ανισότητα Jensen

Έστω τυχαία μεταβλητή X και κυρτη συνάρτηση f . Τότε θα ισχύει ότι $f(E[X]) \leq E[f(X)]$

Πρόταση 1.1

Έστω μ κατανομή στο \mathbb{R} με εκθετικές ροπές και μέση τιμή \bar{x} και έστω I, M, Λ η συνάρτηση ταχύτητας, η ροπογεννήτρια και η λογαριθμική ροπογεννήτρια της κατανομής μ αντίστοιχα. Τέλος $-\infty \leq a < 0 < b \leq \infty$ τα άκρα του D_M .

(α) I κυρτή και κάτω ημισυνεχής.

(β) $I(x) = 0$ αν $x = \bar{x}$.

(γ) $I(x) = \sup_{\lambda \geq 0} (\lambda x - \log M_\mu(\lambda))$ για κάθε $x > \bar{x}$ και $I(x) = \sup_{\lambda \leq 0} (\lambda x - \log M_\mu(\lambda))$ για κάθε $x < \bar{x}$.

(δ) Η I είναι φθίνουσα στο $(-\infty, \bar{x})$ και αύξουσα στο $(\bar{x}, +\infty)$.

(ε) Για κάθε $c \geq 0$ το σύνολο υποστάθμης $I^{-1}[0, c] = \{x \in \mathbb{R} | I(x) \leq c\}$ είναι συμπαγές.

(στ) Έστω $x > \bar{x}$ ($x < \bar{x}$ αντίστοιχα). Αν $b \notin D_M$ ($a \notin D_M$ αντίστοιχα) τότε αν $\mu(x, +\infty) > 0$

($\mu(-\infty, x) > 0$ αντίστοιχα) το supremum στον ορισμό του $I(x)$ πιάνεται σε μοναδικό $\theta_x \in (0, b)$

($\theta_x \in (a, 0)$ αντίστοιχα) και το θ_x λύνει την εξίσωση

$$x = \frac{(M'(\theta_x))}{(M(\theta_x))} \quad (1)$$

ενώ αν $\mu(x, +\infty) = 0$ ($\mu(-\infty, x) = 0$ αντίστοιχα) το οποίο μπορεί να συμβεί μόνο αν $b = \infty$ ($a = -\infty$ αντίστοιχα), τότε το supremum στον ορισμό του $I(x)$ πάνεται καθώς το $\lambda \uparrow +\infty$ ($\lambda \downarrow -\infty$ αντίστοιχα) και σε αυτή την περίπτωση

$$I(x) = -\log \mu\{x\}$$

Αν $b \in D_M$ ($a \in D_M$ αντίστοιχα) τότε το supremum στον ορισμό του $I(x)$ πάνεται σε μοναδικό $\theta_x \in (0, b]$ ($\theta_x \in [a, 0)$ αντίστοιχα), $\bar{x} < A'_-(b)$ ($A'_+(a) < \bar{x}$ αντίστοιχα) και αν $x \in (\bar{x}, A'_-(b))$ ($x \in (A'_+(a), \bar{x})$ αντίστοιχα) τότε το θ_x λύνει την εξίσωση (1) ενώ αν $x \geq A'_-(b)$ ($x \leq A'_+(a)$ αντίστοιχα) τότε έχουμε ότι $I(x) = bx - A(b)$ ($I(x) = ax - A(a)$ αντίστοιχα) για κάθε $x \geq A'_-(b)$ ($x \leq A'_+(a)$ αντίστοιχα).

Απόδειξη

(α) Η I είναι το supremum των συναρτήσεων $\varphi_\lambda(x) = \lambda x - \log M_\mu(\lambda)$, $\lambda \in \mathbb{R}$

Οι φ_λ είναι προφανώς συνεχής άρα η I είναι και αυτή κάτω ημισυνεχής ως supremum συνεχών συναρτήσεων. Επίσης οι φ_λ είναι γραμμικές και άρα κυρτές.

Αυτό έχει ως αποτέλεσμα η I να είναι κυρτή ως supremum κυρτών συναρτήσεων

(β) Για κάθε $\lambda \in \mathbb{R}$ η συνάρτηση $\mathbb{R} \ni x \rightarrow e^{\lambda x}$ είναι κυρτή, και άρα από την ανισότητα Jensen έχουμε ότι

$$e^{\lambda \bar{x}} \leq \int e^{\lambda x} d\mu(x) = M_\mu(\lambda)$$

Αφού $e^{\lambda \bar{x}} \leq M_\mu(\lambda)$ έχουμε ότι $\lambda \bar{x} \leq \log M_\mu(\lambda)$ για κάθε $\lambda \in \mathbb{R}$ και άρα $I(\bar{x}) \leq 0$. Αυτό αποδεικνύει ότι $I(\bar{x}) = 0$ θα δείξουμε τώρα ότι το $I(x) = 0 \Rightarrow x = \bar{x}$. Έστω $x \geq \bar{x}$ τέτοιο ώστε $I(x) = 0$. Τότε $\lambda \bar{x} \leq \log M_\mu(\lambda)$ για κάθε $\lambda \in \mathbb{R}$ και άρα για κάθε $\lambda > 0$ έχουμε ότι

$$\frac{(e^{\lambda \bar{x}} - 1)}{\lambda} \leq \frac{(e^{\lambda x} - 1)}{\lambda} \leq \frac{(M(\lambda) - 1)}{\lambda}$$

όπου αν πάρουμε το όριο καθώς το λ πηγαίνει προς το 0 τότε θα έχουμε ότι

$$\bar{x} \leq x \leq \bar{x} \quad \text{άρα δείξαμε ότι } x = \bar{x} .$$

(γ) Έστω $x \geq \bar{x}$. Τότε, για κάθε $\lambda \leq 0$ έχουμε ότι

$$\lambda x - \log M_{\mu}(\lambda) \leq \lambda \bar{x} - \log M_{\mu}(\lambda) \leq I(\bar{x}) = 0$$

συνεπώς αφού για το $I(x) \geq 0$ τα $\lambda \leq 0$ δεν συνεισφέρουν στο supremum στον ορισμό του $I(x)$.

Άρα

$$I(x) = \max\left(\sup_{\lambda \leq 0}(\lambda x - \log M_{\mu}(\lambda)), \sup_{\lambda \geq 0}(\lambda x - \log M_{\mu}(\lambda))\right)$$

$$I(x) = \max(0, \sup_{\lambda \geq 0}(\lambda x - \log M_{\mu}(\lambda))) = \sup_{\lambda \geq 0}(\lambda x - \log M_{\mu}(\lambda))$$

Ομοίως βλέπουμε και ότι αν $x \leq \bar{x}$ τότε $I(x) = \sup_{\lambda \leq 0}(\lambda x - \log M_{\mu}(\lambda))$

(δ) Έστω $x < y < \bar{x}$. Τότε για κάθε $\lambda \leq 0$ έχουμε ότι $\lambda y - \log M_{\mu}(\lambda) \leq \lambda x - \log M_{\mu}(\lambda)$

και άρα από το (γ) $I(y) = \sup_{\lambda \leq 0}(\lambda y - \log M_{\mu}(\lambda)) \leq \sup_{\lambda \leq 0}(\lambda x - \log M_{\mu}(\lambda)) = I(x)$

Την ίδια διαδικασία ακολουθούμε για να δείξουμε ότι η I είναι αύξουσα στο $(\bar{x}, +\infty)$

(ε) Αφού η I είναι κάτω ημισυνεχής, το $I^{-1}[0, c]$ είναι κλειστό για κάθε $c \geq 0$.

Αρκεί λοιπόν να δείξουμε ότι είναι φραγμένο. Αυτό όμως είναι προφανές, αφού λόγω του ότι η μ έχει εκθετικές ροπές υπάρχει $\varepsilon > 0$ τέτοιο ώστε $M(-\varepsilon) \vee M(\varepsilon) < +\infty$ και έτσι αν $I(x) \leq c$ τότε

$$(\varepsilon x - A(\varepsilon)) \vee (-\varepsilon x - A(-\varepsilon)) \leq I(x) \leq c$$

Έτσι το $I^{-1}[0, c]$ είναι φραγμένο αφού

$$I^{-1}[0, c] \subseteq \left[\frac{-(c + A(-\varepsilon))}{\varepsilon}, \frac{(c + A(\varepsilon))}{\varepsilon} \right] \subseteq \mathbb{R}$$

(στ) $b \notin D_M$: Αν $b < +\infty$ τότε προφανώς $\mu(x, +\infty) > 0$ για κάθε $x > \bar{x}$ και απο την κάτω ημισυνέχεια των ροπογεννητριών συναρτήσεων έχουμε ότι $\lim_{\lambda \uparrow b} M(\lambda) = +\infty$ και άρα $\lim_{\lambda \uparrow b} (\lambda x - \log M(\lambda)) = -\infty$. Αν $b = \infty$ και $\mu(x, +\infty) > 0$ τότε $\mu(x + \varepsilon, +\infty) > 0$ για κάποιο $\varepsilon > 0$ και άρα για κάθε $\lambda \geq 0$ έχουμε ότι

$$M_\mu(\lambda) = \int e^{\lambda t} d\mu(t) \geq \int_{x+\varepsilon}^{+\infty} e^{\lambda t} d\mu(t) \geq e^{\lambda(x+\varepsilon)} \mu(x+\varepsilon, +\infty)$$

Συνεπώς έχουμε οτι

$$\lambda x - \log M_\mu(\lambda) \leq -\lambda \varepsilon - \log \mu(x + \varepsilon, +\infty) \rightarrow -\infty$$

για λ να πηγαίνει στο $+\infty$

Άρα σε κάθε περίπτωση έχουμε οτι αν $b \notin D_M$ $\lim_{\lambda \uparrow b} (\lambda x - \log M(\lambda)) = -\infty$ και άρα

$\lambda x - \log M_\mu(\lambda) < 0$ για κάθε $\lambda \geq \lambda_0$ για κάποιο $\lambda_0 \in [0, b)$. Από το **(γ)** όμως έχουμε ότι

$I(x) = \sup_{\lambda \geq 0} (\lambda x - \log M_\mu(\lambda))$ και άρα το supremum στον ορισμό του $I(x)$ πιάνεται σε κάποιο

πεπερασμένο $\theta_x \in (0, b]$, οπότε $I(x) = \theta_x - \log M(\theta_x) < +\infty$. Τώρα για κάθε θ_x στο οποίο

πιάνεται το supremum στον ορισμό του $I(x)$, έχουμε από το **(β)** ότι $\theta_x > 0$ και

άρα κάθε τέτοιο θ_x είναι σημείο τοπικού ακροτάτου της $(0, b) \ni \lambda \rightarrow \lambda x - \log M(\lambda)$.

Συνεπώς έχουμε οτι

$$x - \frac{(M'(\theta_x))}{(M(\theta_x))} = \frac{d}{d\lambda} \Big|_{\lambda=\theta_x} (\lambda x - \log M(\lambda)) = 0$$

Τέλος σε όλες τις παραπάνω περιπτώσεις το μ δεν μπορεί να είναι κατανομή Dirac οπότε από τις ιδιότητες των ροπογεννητριών συναρτήσεων έχουμε οτι η συνάρτηση

$(0, b) \ni \lambda \rightarrow \lambda x - A(\lambda)$ είναι αυστηρά κοίλη, το οποίο αποδεικνύει τη μοναδικότητα του θ_x .

Εξετάζουμε τώρα την περίπτωση όπου $\mu(x, +\infty) = 0$. Τότε

$$\lambda x - \log M_\mu(\lambda) = -\log e^{-\lambda x} - \log \int_{-\infty}^x e^{\lambda t} d\mu(t) = -\log \left(e^{-\lambda x} \int_{-\infty}^x e^{\lambda t} d\mu(t) \right)$$

$$\lambda x - \log M_\mu(\lambda) = -\log \int_{(-\infty, x]} 1_{(-\infty, x]} e^{-\lambda(x-t)} d\mu(t) \quad (2)$$

Όμως $1_{(-\infty, x]} e^{-\lambda(x-t)} \downarrow 1_x(t)$ καθώς $\lambda \uparrow +\infty$ και άρα από το θεώρημα φραγμένης σύγκλισης έχουμε ότι

$$\int 1_{(-\infty, x]} e^{-\lambda(x-t)} d\mu(t) \downarrow \mu\{x\} \quad \text{καθώς το } \lambda \uparrow +\infty$$

Έτσι, από την (2) η ποσότητα $\lambda x - \log M_\mu(\lambda)$ αυξάνει καθώς το $\lambda \uparrow +\infty$ και άρα έχουμε ότι

$$\sup_{\lambda \geq 0} (\lambda x - \log M_\mu(\lambda)) = \lim_{\lambda \uparrow b} (\lambda x - \log M_\mu(\lambda)) = -\log \mu\{x\}$$

Τώρα θα πάρουμε την περίπτωση όπου $b \in D_M$

$b \in D_M$: Αφού $b \in D_M$ έχουμε ότι $b < +\infty$ και άρα $\mu(x, +\infty) > 0$ κάθε $x > \bar{x}$ και το μ δεν είναι κάποιο μέτρο Dirac. Συνεπώς η Λ' είναι γνησίως αύξουσα στο D_μ^o και άρα

$$\Lambda'_+(\alpha) < \Lambda'(0) = \bar{x} < \Lambda'_-(b)$$

Αφού $b \in D_M$ $I(x) = \sup_{\lambda \in [0, b]} (\lambda x - A(\lambda)) < +\infty$ και άρα το supremum στον ορισμό του

$I(x)$ πιάνεται σε κάποιο $\theta_x \in [0, b]$ και το θ_x είναι μοναδικό αφού η

$$(0, b) \ni \lambda \mapsto \lambda x - A(\lambda)$$

είναι αυστηρά κοίλη. Μάλιστα, εδώ έχουμε ότι $I(y) = \sup_{\lambda \in [0, b]} (\lambda y - A(\lambda)) < +\infty$ για κάθε $y \geq \bar{x}$ οπότε $[\bar{x}, +\infty) \subseteq D_I$. Από το (β) πάλι $\theta_x > 0$ και άρα $\theta_x \in (0, b)$.

Έστω τώρα $\bar{x} < x < \Lambda'_-(b)$. Αφού $\Lambda' > 0$ στο $(0, b)$ και η Λ είναι λεία έπεται ότι

$A' : (0, b) \rightarrow A'(0, b)$ είναι αύξουσα αμφιδιαφορίσιμη. Επίσης από την κυρτότητα της Λ έχουμε ότι $A'_- = A'$ είναι συνεχής από αριστερά στο $(0, b]$ και άρα $A'(0, b) = (\bar{x}, A'_-(b))$. Συνεπώς για κάθε $\bar{x} < x < A'_-(b)$ υπάρχει $\lambda_x \in (0, b)$ τέτοιο ώστε

$$A'(\lambda_x) = x. \text{ Τώρα προφανώς } \varphi'_x(\lambda_x) = 0 \text{ και άρα το } \lambda_x \text{ είναι κρίσιμο σημείο } \varphi_x.$$

Όμως $\varphi''_x(\lambda_x) = -A(\lambda_x) \leq 0$, οπότε το λ_x είναι σημείο μεγίστου της φ_x και άρα από τη μοναδικότητα του θ_x έχουμε ότι $\lambda_x = \theta_x$ και ο θ_x ικανοποιεί την (1)

Από την άλλη, αν $x \geq A'_-(b)$ τότε από την κυρτότητα της Λ έχουμε ότι

$$x \geq A'_-(b) \geq \frac{A(b) - A(\lambda)}{b - \lambda}$$

για κάθε $\lambda < b$ από όπου έπεται ότι

$$I(x) = \sup_{\lambda \in [0, b]} (\lambda x - A(\lambda)) < bx - A(b),$$

ενώ από την άλλη αφού $b \in D_M$ προφανώς $I(x) > bx - A(b)$

1.2 Θεώρημα Cramer

1.2 Θεώρημα Cramer

Έστω $\{X_i\}_{i \in \mathbb{N}}$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με κοινή κατανομή μ και έστω $I: \mathbb{R} \rightarrow [0, +\infty]$ η συνάρτηση ταχύτητας της μ . Αν η μ έχει εκθετικές ροπές και

$\bar{x} := \int_{\mathbb{R}} x d\mu(x)$, τότε για κάθε $x > \bar{x}$ ο αριθμός $I(x)$ είναι ο βέλτιστος ρυθμός εκθετικής μείωσης της πιθανότητας $P(S_n \geq nx)$ με την έννοια ότι

$$P(S_n \geq nx) \leq e^{-nI(x)} \quad (1)$$

για κάθε $n \in \mathbb{N}$

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \geq -I(x) \quad (2)$$

Ομοίως, για $x \leq \bar{x}$ το σπάνιο ενδεχόμενο $(-\infty, x]$ ικανοποιεί τις (1), (2).

Απόδειξη

Η ανισότητα (1) έπεται από την ανισότητα του Markov αφού για κάθε $n \in \mathbb{N}$ και κάθε $\lambda > 0$ έχουμε ότι

$$P(S_n \geq nx) \leq e^{-\lambda nx} \mathbb{E}(e^{\lambda S_n})$$

$$e^{-\lambda nx} \mathbb{E}(e^{\lambda S_n}) = e^{-\lambda nx} (M(\lambda))^n$$

$$e^{-\lambda nx} (M(\lambda))^n = e^{-\lambda nx + n \log M(\lambda)}$$

$$e^{-\lambda nx + n \log M(\lambda)} \leq \inf_{\lambda} e^{-\lambda nx + n \log M(\lambda)} \quad \text{η τελευταία ανισότητα προκύπτει από το (γ) της Πρότασης 1.1}$$

$$\inf_{\lambda} e^{-\lambda nx + n \log M(\lambda)} = e^{-n \sup_{\lambda} (\lambda x + \log M(\lambda))} = e^{-n I(x)}$$

Τώρα για να ικανοποιήσουμε την (2) αρκεί να εξετάσουμε την περίπτωση όπου το μ δεν είναι κατανομή Dirac και αρκεί να αποδείξουμε ότι για κάθε $\delta > 0$ έχουμε ότι

$$\liminf_{n \rightarrow +\infty} \left(\frac{1}{n} \right) \log P(nx \leq S_n \leq nx + n\delta) \geq -I(x)$$

Και σε αυτή τη περίπτωση θέτουμε $b := \sup D_M$. Εξετάζουμε πρώτα την περίπτωση στην οποία το supremum στον ορισμό του $I(x)$ πιάνεται σε κάποιο $\theta \in (0, +\infty) \cap D_M$ τέτοιο ώστε να ικανοποιείται

$$x = \frac{(M'(\theta_x))}{(M(\theta_x))} \quad \text{δηλαδή}$$

$$x = \int y d\mu_\theta(y) =: E[\mu_\theta] \quad (1)$$

Όπου μ_θ είναι η θ -εκθετική διαταραχή του μ . Για κάθε $n \in \mathbb{N}$ συμβολίζουμε με $s_n: \mathbb{R}^n \rightarrow \mathbb{R}$ τη συνάρτηση με τύπο $s_n(y) = y_1 + \dots + y(n)$, οπότε $S_{n*} P = s_{n*} \mu^{\otimes n}$, όπου $\mu^{\otimes n} := \mu \otimes \dots \otimes \mu$ n -φορές, και

$$d\mu_\theta^{\otimes n}(y) = \frac{(e^{\theta s_n(y)})}{(M^n(\theta))} d\mu^{\otimes n}(y)$$

Τότε προφανώς για κάθε $0 < \varepsilon < \delta$ έχουμε ότι

$$P(nx \leq S_n \leq nx + n\delta) = \mu^{\otimes n}(s_n^{-1}[nx, nx + n\delta]) \geq \int_{s_n^{-1}[nx, nx + n\varepsilon]} 1 d\mu^{\otimes n}$$

Στο σύνολο $s_n^{-1}[nx, nx + n\varepsilon]$ όμως έχουμε ότι $\theta s_n - n\theta x - n|\theta|\varepsilon \leq 0$ και άρα από την παραπάνω ανισότητα έχουμε ότι

$$P(nx \leq S_n \leq nx + n\delta) \geq \int_{s_n^{-1}[nx, nx + n\varepsilon]} e^{\theta s_n(y) - n\theta x - n|\theta|\varepsilon} d\mu^{\otimes n}(y)$$

$$= e^{-n\theta x - n|\theta|\varepsilon} M^n(\theta) \int_{s_n^{-1}[nx, nx + n\varepsilon]} d\mu_\theta^{\otimes n}$$

Τώρα παίρνοντας και από τα δύο μέλη λογαρίθμους και διαιρώντας με n έχουμε

$$\frac{1}{n} \log P(x \leq \frac{S_n}{n} \leq x + \delta) \geq -\theta x - |\theta| \varepsilon + \log M(\theta) + \frac{1}{n} \log \int 1_{s^{-1}[nx, nx+n\varepsilon]} d\mu_\theta^{\otimes n}$$

Έστω τώρα το $P^\theta := \mu_\theta^{\otimes \infty} = \mu_\theta \otimes \mu_\theta \otimes \dots$ είναι το άπειρο γινόμενο στο μετρήσιμο χώρο γινόμενο $(\mathbb{R}^{\mathbb{N}}, B_{\mathbb{R}}^{\mathbb{N}})$ και έστω $x^i: \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ οι φυσικές προβολές $x^i(x) = x(i)$. Όπως ξέρουμε το P^θ καθορίζεται από τις

$$(x^1, \dots, x^n)_* P^\theta = \mu_\theta^{\otimes n}, \quad \forall n \in \mathbb{N}$$

όπου φυσικά $(x^1, \dots, x^n): \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^n$ είναι η προβολή στις πρώτες n συντεταγμένες και άρα

$$\begin{aligned} \int 1_{s^{-1}[nx, nx+n\varepsilon]} d\mu_\theta^{\otimes n} &= \int 1_{s^{-1}[nx, nx+n\varepsilon]}(x^1, \dots, x^n) dP^\theta \\ &= P^\theta(0 \leq \frac{1}{n} \sum_{i=1}^n x^i - x \leq \varepsilon) \end{aligned}$$

Η ακολουθία $(x^i)_{i \in \mathbb{N}}$ στο χώρο πιθανότητας $(\mathbb{R}^{\mathbb{N}}, B_{\mathbb{R}}^{\mathbb{N}}, P^\theta)$ είναι ακολουθία ανεξάρτητων ισόνομων και τυχαίων μεταβλητών με κοινή κατανομή μ_θ και από την (1) έχουμε ότι

$E^{P^\theta} x^1 = \int y d\mu_\theta(y) = x$. Αν η μ έχει θετική διασπορά $\sigma > 0$ τότε και η μ_θ έχει θετική διασπορά $\sigma_\theta > 0$ και άρα από το κεντρικό οριακό θεώρημα έχουμε ότι

$$\begin{aligned} P^\theta(0 \leq \frac{1}{n} \sum_{i=1}^n x^i - x \leq \varepsilon) &= P^\theta(0 \leq \frac{1}{(\sigma_\theta \sqrt{n})} \sum_{i=1}^n x^i - x \leq (\varepsilon \frac{\sqrt{n}}{\sigma_\theta})) \\ &\xrightarrow{n \rightarrow +\infty} N(0,1)([0, +\infty)) = \frac{1}{2} \end{aligned}$$

Συνεπώς έχουμε ότι

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \log \int 1_{s^{-1}[nx, nx+n\varepsilon]} d\mu_\theta^{\otimes n} \right) = 0 \log\left(\frac{1}{2}\right) = 0$$

οπότε έχουμε ότι

$$\lim_{n \rightarrow +\infty} \left(\inf \frac{1}{n} \log P \left(x \leq \frac{S_n}{n} \leq x + \delta \right) \geq -\theta x - |\theta| \varepsilon + \log M(\theta) \right)$$

Αφού αυτό ισχύει για κάθε $\varepsilon > 0$ έχουμε ότι

$$\lim_{n \rightarrow +\infty} \left(\inf \frac{1}{n} \log P \left(x \leq \frac{S_n}{n} \leq x + \delta \right) \geq -\theta x + \log M(\theta) \right) = -I(x)$$

από τον ορισμό του θ . Συνεπώς το ζητούμενο ισχύει για κάθε $x > \bar{x}$ τέτοιο ώστε το supremum στον

ορισμό του $I(x)$ να πιάνεται σε κάποιο $\theta \in (0, +\infty) \cap D_M$ που ικανοποιεί την $x = \frac{M'(\theta_x)}{M(\theta_x)}$. Όπως

έχουμε δει στην πρόταση 1 στο (ε) ερώτημα αυτό μπορεί να μην ισχύει μόνο αν είτε $b = +\infty$ και

$\mu(x, +\infty) = 0$, οπότε $I(x) = -\log \mu(x)$, είτε $b \in D_M$ και $x \geq A'(b)$, οπότε

$I(x) = bx - \log M(b)$. Αν $b = +\infty$ και $\mu(x, +\infty) = 0 = P(X_n > x) = 0$, τότε

$$P(nx \leq S_n \leq nx + n\delta) = P(X_1 = \dots = X_n = x) = P(X_1 = x)^n = (\mu(x))^n$$

οπότε παίρνοντας λογαριθμούς και διαιρώντας με n έχουμε ότι

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \log P \left(x \leq \frac{S_n}{n} \leq x + \delta \right) = \log \mu(x) \right) = -I(x)$$

Εώς τώρα έχουμε αποδείξει το ζητούμενο στην περίπτωση όπου $b \notin D_M$. Για να δείξουμε το

ζητούμενο στην περίπτωση όπου $b \in D_M$ και $x > A'(b)$ θα αναχθούμε στην περίπτωση όπου

$b \notin D_M$ προσεγγίζοντας την κοινή κατανομή μ των X_i με κατάλληλη ακολουθία $(\bar{\mu}_N)$ τέτοια

ώστε $\sup D_{M_{\bar{\mu}_N}} = +\infty \notin \sup D_{M_{\bar{\mu}_N}}$. Ορίζουμε καταρχάς μέτρα $\mu_N, N \in \mathbb{N}$, από την $d\mu_N = 1_{(-\infty, N]} d\mu$.

Αφού $\sup D_{M_{\bar{\mu}_N}} < +\infty$ έχουμε ότι $\mu(x, +\infty) > 0$ και άρα $\mu((-\infty, N] \cap (x, +\infty)) > 0$ για κάθε

$N \geq N_1$ για κάποιο $N_1 \in \mathbb{N}$. Ειδικότερα $\mu_N \neq 0$ για κάθε $N \geq N_1$ και έτσι από το N_1 και

πέρα ορίζονται οι κατανομές

$$\bar{\mu}_N := \frac{1}{(\mu_N(\mathbb{R}))} \mu_N$$

Προφανώς $\sup D_{M_{\bar{\mu}_N}} = +\infty$ και $\mu_N(x, +\infty) > 0$ για κάθε $N \geq N_1$. Επίσης, για κάθε $N \geq N_1$ και κάθε Borel $A \subseteq \mathbb{R}$ έχουμε ότι

$$\mu(A) \geq \mu_N(A) = \mu((-\infty, N]) \bar{\mu}_N(A),$$

και άρα για κάθε $n \in \mathbb{N}$, $N \geq N_1$ και κάθε Borel $A \subseteq \mathbb{R}$ έχουμε ότι

$$\mu^{*n}(nA) = \int 1_{nA}(s_n) d\mu^{\otimes n} = \int \dots \int 1_{nA}(s_n) d\mu \dots d\mu$$

$$\mu^{*n}(nA) \geq \mu((-\infty, N])^n \int \dots \int 1_{nA}(s_n) d\bar{\mu}_N \dots d\bar{\mu}_N$$

$$\mu^{*n}(nA) \geq \mu((-\infty, N])^n \bar{\mu}_N^{*n}(nA)$$

Συνεπώς

$$\liminf \frac{1}{n} \log \mu^{*n}(nA) \geq \log \mu((-\infty, N]) + \liminf \frac{1}{n} \log \bar{\mu}_N^{*n}(nA)$$

για κάθε $N \geq N_1$. Προφανώς για κάθε $\lambda \geq 0$ έχουμε ότι $M_{\bar{\mu}_N}(\lambda) \leq e^{\lambda N} < +\infty$ και άρα

$\sup D_{M_{\bar{\mu}_N}} = +\infty$ οπότε το θεώρημα του Cramer ισχύει για τις $\bar{\mu}_N$ και άρα αν

$A := [x, x + \delta)$, $x > \bar{x}$, $\delta > 0$ τότε

$$\liminf \frac{1}{n} \log \mu^{*n}(nA) \geq \log \mu((-\infty, N]) - I_{\mu_N}(x)$$

για κάθε $N \geq N_1$, οπότε

$$\liminf \frac{1}{n} \log \mu^{*n}(nA) \geq \limsup_{N \rightarrow +\infty} (-I_{\mu_N}(x)) = - \liminf_{N \rightarrow +\infty} (I_{\mu_N}(x))$$

Αρκεί λοιπόν να αποδείξουμε ότι

$$I(x) := \liminf_{N \rightarrow +\infty} (I_{\mu_N}(x)) \leq I(x)$$

Έστω A_N, \bar{A}_N είναι οι λογαριθμικές ροπογεννήτριες των μ_N και $\bar{\mu}_N$ αντίστοιχα. Τότε η

$(A_N)_N$ είναι αύξουσα και από το θεώρημα μονότονης σύγκλισης έχουμε ότι $A_N(\lambda) \uparrow A(\lambda)$ για κάθε

$\lambda \in \mathbb{R}$. Συνεπώς επίσης

$$\overline{A}_N(\lambda) = -\log \mu(-\infty, N] + A_N(\lambda) \rightarrow A(\lambda)$$

για κάθε $\lambda \in \mathbb{R}$. Αφού η $(A_N)_N$ είναι αύξουσα και $A_N \leq A$ για κάθε $N \in \mathbb{N}$ και

$$\overline{T}_N(x) := I_{\overline{\mu}_N}(x) = I_N(x) + \log \mu(-\infty, N] \leq I_N(x)$$

για κάθε $N \geq N_1$. Συνεπώς

$$\overline{T}(x) := \liminf_{N \rightarrow +\infty} (\overline{T}_N(x)) = \liminf_{N \rightarrow +\infty} (I_N(x)) + \lim_{N \rightarrow +\infty} \log \mu(-\infty, N]$$

$$\overline{T}(x) = \downarrow \lim_{N \rightarrow +\infty} (I_N(x)) \geq I(x)$$

Από το Θεώρημα κυριαρχημένης σύγκλισης έχουμε ότι

$$\overline{x}_N := \int x d\overline{\mu}_N(x) = \frac{1}{(\mu(-\infty, N])} \int_{-\infty}^N x d\mu(x) \rightarrow \overline{x}$$

καθώς το $N \rightarrow +\infty$ και άρα αφού $\overline{x} < x$, από κάποιο N_2 και πέρα έχουμε ότι $x > \overline{x}_N$ οπότε απο την πρόταση 1 το (γ) ερώτημα έχουμε ότι

$$\overline{T}_N(x) = \sup_{\lambda \geq 0} (\lambda x - \overline{A}_N(\lambda)) = \sup_{\lambda \geq 0} (\lambda x - A_N(\lambda)) + \log \mu(-\infty, N]$$

Ειδικότερα για $N \geq N_2$ έχουμε ότι $I_N(x) = \sup_{\lambda \geq 0} (\lambda x - A_N(\lambda))$. Ορίζουμε τώρα για κάθε

$N \geq N_0 := N_1 \vee N_2$ τα σύνολα

$$I_N(x) = \langle \lambda \geq 0 \mid ((\lambda x - A_N(\lambda)) \geq \overline{T}_N(x)) \rangle$$

Παρατηρούμα ότι αν δείξουμε ότι $\bigcap_{N \geq N_0} A_N \neq \emptyset$ τότε επιλέγοντας $\lambda_0 \in A$ έχουμε ότι

$$I(x) \geq (\lambda_{0x} - A(\lambda_0)) = \lim_{N \rightarrow +\infty} (\lambda_{0x} - A_N(\lambda_0)) \geq \overline{I}(x) \quad \text{το οποίο ουσιαστικά}$$

ολοκληρώνει την απόδειξη.

Είναι εύκολο να δούμε ότι η $(A_N)_{N \geq N_0}$ είναι φθίνουσα ακολουθία μη-κενών κλειστών διαστημάτων. Πράγματι, το ότι η $(A_N)_{N \geq N_0}$ είναι φθίνουσα έπεται από το ότι η $(A_N)_{N \geq N_0}$ είναι αύξουσα, το ότι τα (A_N) είναι κλειστά και κυρτά έπεται από το ότι η $[0, +\infty) \ni \lambda \rightarrow (\lambda_x - A_N(\lambda))$ είναι συνεχής και κοίλη αντίστοιχα, και τέλος, για κάθε $N \geq N_0$ έχουμε ότι $\mu_N(x, +\infty) > 0$ οπότε από τον ορισμό του supremum στον ορισμό του $I_N(x)$ πιάνεται σε κάποιο πεπερασμένο $\lambda_N > 0$ και άρα

$$\lambda_N x - A_N(\lambda_N) = I_N(x) \geq \overline{I}(x)$$

δηλαδή $\lambda_N \in A_N$.

1.3 Εφαρμογές του Θεωρήματος Cramer σε γνωστές Κατανομές

Εφαρμογή του Θεωρήματος Cramer για την κατανομή Bernoulli

Το προβλημά μας εδώ είναι η εκτίμηση της ταχύτητας σύγκλισης της πιθανότητας

$$P(S_n \geq nx) \rightarrow 0$$

για $q > p$ η ταχύτητα σύγκλισης της πιθανότητας

$$P(S_n \leq nx) \rightarrow 0$$

Κάνοντας μια απλή ανισότητα Chebyshev ($P(|X - E(X)| \geq a) \leq \text{Var} \frac{(X)}{a^2}$)

καταλήγουμε ότι η ταχύτητα σύγκλισης είναι τουλάχιστον γραμμική αλλά αυτό το γνωρίζουμε ούτως η άλλως γιατί οι X_i έχουν εκθετικές ροπές δηλαδή $E(e^{\lambda X_i}) < +\infty$ και αυτό ισχύει για κάθε $\lambda \in \mathbb{R}$ έτσι μπορούμε να πούμε ότι η ταχύτητα σύγκλισης είναι τουλάχιστον εκθετική.

$$P(S_n \geq nq) \leq e^{-\lambda nq} E(e^{\lambda S_n}) = e^{-\lambda nq} (E[e^{\lambda X_1}])^n$$

$$e^{-\lambda nq} (pe^{\lambda} + 1 - p)^n = e^{-\lambda nq} e^{n \log(pe^{\lambda} + 1 - p)}$$

$$e^{-\lambda nq} (pe^{\lambda} + 1 - p)^n = e^{-\lambda nq + n \log(pe^{\lambda} + 1 - p)} = \inf_{\lambda} e^{-\lambda nq + n \log(pe^{\lambda} + 1 - p)}$$

$$\inf_{\lambda} e^{-\lambda nq + n \log(pe^{\lambda} + 1 - p)} = e^{-n \sup_{\lambda} (\lambda q - \log(pe^{\lambda} + 1 - p))} = e^{-nI(q)}$$

Άρα βρήκαμε ότι το $I(q)$ είναι $I(q) = \sup_{\lambda \in \mathbb{R}} [\lambda q - \log(pe^{\lambda} + 1 - p)] = \sup_{\lambda \geq 0} [\lambda q - \log(pe^{\lambda} + 1 - p)]$

Η τελευταία ισότητα προκύπτει από την ανισότητα Jensen. Αποδείξαμε λοιπόν ότι η ταχύτητα σύγκλισης της πιθανότητας είναι εκθετική με ρυθμό μάλιστα $I(q)$. Τώρα ήρθε η ώρα να υπολογίσουμε αυτό το ρυθμό.

Παραγωγίζουμε την σχέση που βρήκαμε για το $I(q)$ ως προς λ

$$\frac{d}{d\lambda}(\lambda q - \log(p e^\lambda + 1 - p)) = q - \frac{(p e^\lambda)}{(p e^\lambda + 1 - p)} = q - 1 + \frac{(1 - p)}{(p e^\lambda + 1 - p)}$$

Τώρα θα διακρίνουμε περιπτώσεις για το q

(1) $q \geq 1$ Είναι προφανές ότι η παραγωγός της είναι θετική για κάθε $\lambda \in \mathbb{R}$

$$I(q) = \lim_{\lambda \rightarrow +\infty} (\lambda q - \log(p e^\lambda + 1 - p)) = \log \frac{1}{p} \quad \text{για } q=1$$

για $q > 1$ προφανώς και είναι $+\infty$

(2) για $p < q < 1$ εδώ έχουμε μόνο ένα κρίσιμο σημείο το $\lambda = \log\left(\frac{q(1-p)}{p(1-q)}\right) > 0$

Τώρα όταν παραγωγίζουμε την $I(q)$ ως προς q στο $(p, 1)$ βλέπουμε ότι είναι γνησίως αύξουσα και άρα $I(q) > 0$ για κάθε $q \in (p, 1)$

(3) Το ίδιο ακριβώς κάνουμε και για $0 < q < p$ και μας βγαίνει ότι η $I(q)$ είναι φθίνουσα και για $q \leq 0$ έχουμε ότι

$$I(q) = \lim_{\lambda \rightarrow +\infty} (\lambda q - \log(p e^\lambda + 1 - p)) = \log \frac{1}{(1-p)} \quad \text{για } q=0$$

και για $q < 0$ είναι $+\infty$

$$\text{Άρα } I(q) = q \log \frac{q}{p} + (1-q) \log \frac{(1-q)}{(1-p)} \quad q \in [0, 1]$$

και αν το q δεν ανήκει εκεί τότε ο ρυθμός είναι ίσος με $+\infty$

$$I(q) = +\infty \quad q \notin [0, 1]$$

Το μόνο που μας μένει να κάνουμε είναι να δείξουμε ότι ο I είναι ο βέλτιστος ρυθμός.

Τα βήματα είναι τα εξής :

Αρχικά θα δείξουμε ότι αν υπάρχουν $n_0 \in \mathbb{N}$ και $\theta > 0$ ώστε

$$n \geq n_0 \Rightarrow P(S_n \geq nx) \leq e^{-n\theta} \quad (i)$$

τότε το $\theta < I(q)$.

Άρα πρέπει να δείξουμε

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \geq -I(x)$$

Βέβαια από την σχέση (I) προκύπτει

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \leq -\theta$$

άρα όλο μαζί μπορούμε να το γράψουμε ως εξής

$$-I(q) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \leq -\theta$$

Ακόμα μπορούμε να πούμε ότι το άθροισμα S_n ως άθροισμα τυχαίων μεταβλητών Bernoulli με παράμετρο $p \in (0,1)$ ακολουθεί διωνυμική κατανομή με παράμετρο p .

Επειδή όμως είναι δύσκολο να υπολογίσουμε το όριο του παραπάνω αθροίσματος θα το υπολογίσουμε για $p=q$ με την βοήθεια του κεντρικού οριακού θεωρήματος.

$$\sum_{k > np} \binom{n}{k} p^k (1-p)^{n-k} = P(S_n \geq np) = P\left(\frac{(S_n - np)}{(p(1-p)\sqrt{n})} \geq 0\right)$$

και από το κεντρικό οριακό θεώρημα έχουμε

$$\lim_{n \rightarrow +\infty} \sum_{k > np} \binom{n}{k} p^k (1-p)^{n-k} = N(0,1)([0, +\infty]) = \frac{1}{2}$$

Τώρα για $q > p$ γράφουμε ότι

$$P(S_n \geq nq) = \sum_{k \geq nq} \binom{n}{k} q^k (1-q)^{(n-k)} \left(\frac{p}{q}\right)^k \left(\frac{1-p}{1-q}\right)^{(n-k)}$$

$$P(S_n \geq nq) = \left(\frac{1-p}{1-q}\right)^n \sum_{k \geq nq} \binom{n}{k} q^k (1-q)^{(n-k)} \left(\frac{p(1-q)}{q(1-p)}\right)^k \quad (1)$$

Τώρα απο γνωστή πρόταση έχουμε ότι

$$\sum_{k > n(p+\varepsilon)} \binom{n}{k} p^k (1-p)^{(n-k)} = P(S_n \geq n(p+\varepsilon)) = P\left(\frac{S_n - np}{(p(1-p)\sqrt{n})} \geq \frac{(\varepsilon\sqrt{n})}{\sigma}\right) \rightarrow 0$$

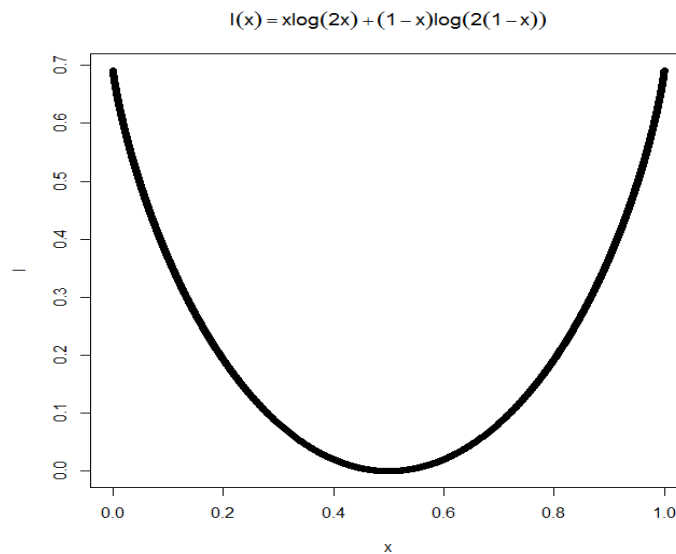
$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \geq \log \frac{(1-p)}{(1-q)} + (q+\varepsilon) \log \frac{(p(1-p))}{(q(1-q))}$$

και αφού αυτό ισχύει γαι κάθε $\varepsilon > 0$

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \geq \log \frac{(1-p)}{(1-q)} + q \log \frac{(p(1-p))}{(q(1-q))} = -I(q) \quad \text{αυτο που θέλαμε δηλαδή.}$$

Αυτό που κάνουμε ουσιαστικά στη σχέση (1) είναι να αλλάζουμε την κατανομή των X_i στο προβλημα μας. Αυτό το κάνουμε έτσι ώστε αυτό που θεωρούσαμε ως μη τυπική συμπεριφορά τώρα να γίνει τυπική συμπεριφορά. Αυτή η μέθοδος χρησιμοποιείται συχνά στη θεωρία αποκλίσεων.

Θεωρητικό γράφημα της κατανομής μ αν ακολουθεί Bernoulli(p) με $p = \frac{1}{2}$



Εφαρμογή του Θεωρήματος Cramer για την Εκθετική κατανομή

Το προβλημα μας και εδώ είναι η εκτίμηση της ταχύτητας σύγκλισης της πιθανότητας

$$P(S_n \geq nx) \rightarrow 0$$

Άρα πρέπει να υπολογίσουμε την $I(x) = \sup_{\lambda} (\lambda x - \log M_{\mu}(\lambda))$.

Αρχικά υπολογίζουμε την ροπογεννήτρια $M_{\mu}(\lambda)$

$$M_{\mu}(\lambda) = \int e^{\lambda x} d\mu(x) = \int_0^{+\infty} e^{\lambda x} \gamma e^{-\gamma x} dx = \frac{\gamma}{(\gamma - \lambda)} \text{ για } \lambda < \gamma \text{ για } \lambda \geq \gamma M_{\mu}(\lambda) = +\infty.$$

Τώρα λοιπόν η συνάρτηση ταχύτητας έχει πάρει την παρακάτω μορφή

$$I(x) = \sup_{\lambda < \gamma} (\lambda x - \log \frac{\gamma}{(\gamma - \lambda)}) \text{ και θα πάρουμε την παράγωγο της}$$

συνάρτησης $A(\lambda, x) = \lambda x - \log \frac{\gamma}{(\gamma - \lambda)}$ ώστε να βρούμε το μέγιστο της $I(x)$.

$$\frac{d}{d\lambda} A(\lambda, x) = \frac{d}{d\lambda} (\lambda x - \log(\frac{\gamma}{(\gamma - \lambda)})) = x - \frac{1}{(\gamma - \lambda)}.$$

Από την παράγωγο βλέπουμε λοιπόν ότι όσο το λ μεγαλώνει και πάει κοντά στο γ , τόσο η παράγωγος γίνεται αρνητική γιατί ο δεύτερος όρος πηγαίνει στο άπειρο αφού ο παρανομαστής πάει στο 0. Από την άλλη όσο το λ μικραίνει, τόσο ο δεύτερος όρος πηγαίνει στο 0 οπότε η παράγωγος εξαρτάται κυρίως από το x . Στην συγκεκριμένη περίπτωση όμως βλέπουμε ότι το x για να έχει τοπικό ακρότατο η παράγωγος πρέπει να είναι μεγαλύτερο από το 0 αυστηρά, γιατί αλλιώς η παράγωγος θα είναι παντού αρνητική, αφού ο δεύτερος όρος είναι παντού αρνητικός, οπότε θα είναι γνησιώς φθίνουσα παντού για $x < 0$. Άρα μιλάμε μόνο για $x > 0$ και ψάχνουμε να βρούμε τοπικό ακρότατο της συνάρτησης. Τώρα η

παράγωγος μηδενίζεται στο $\lambda_* = \gamma - \frac{1}{x}$ όπου είναι τοπικό μέγιστο της συνάρτησης, γιατί όπως είπαμε

και παραπάνω όσο μικραίνει το λ η παράγωγος αυξάνεται άρα γίνεται θετική και όσο το λ αυξάνεται η παράγωγος γίνεται αρνητική. Οπότε το λ_* είναι το σημείο στο οποίο η συνάρτηση πιάνει το τοπικό μέγιστο και πάμε να το αντικαταστήσουμε έτσι ώστε να πάρουμε την $I(x)$.

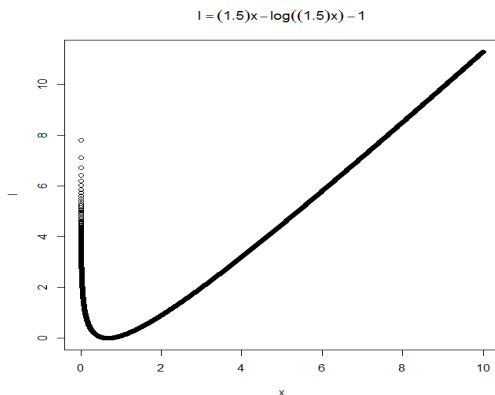
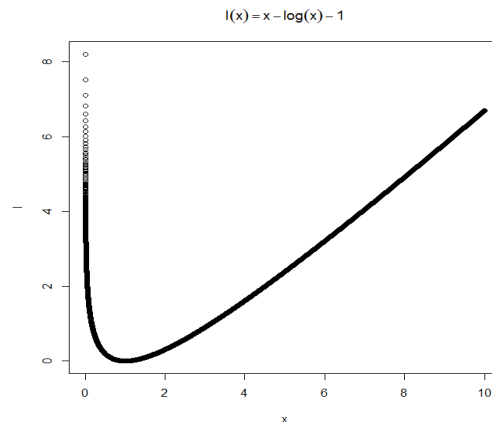
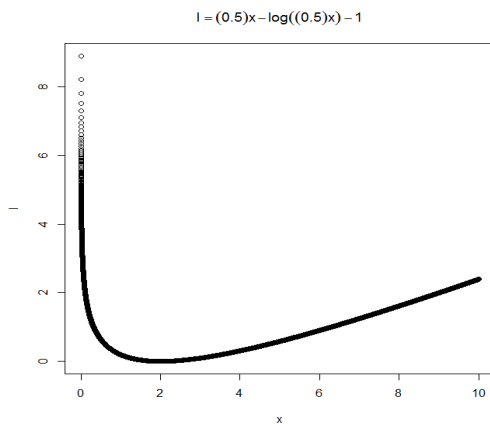
$$I(x) = \sup_{\lambda < \gamma} (\lambda x - \log \frac{\gamma}{\gamma - \lambda}) \quad x > 0$$

$$I(x) = \lambda_* x - \log \frac{\gamma}{\gamma - \lambda_*} \quad x > 0$$

Τώρα αντικαθιστούμε το $\lambda_* = \gamma - \frac{1}{x}$ έτσι ώστε να πάρουμε την τελική μορφή της συνάρτησης ταχύτητας $I(x)$

$$I(x) = \gamma x - 1 - \log(\gamma x) \quad x > 0 \quad \text{και} \quad I(x) = +\infty \quad x \leq 0$$

Παραθέτουμε και τρία παραδείγματα στην R για την θεωρητική γραφική παράσταση της συνάρτησης ταχύτητας για εκθετική κατανομή για τρεις τιμές του γ . Η πρώτη για $\gamma=0.5$, η δεύτερη για $\gamma=1$ και η τρίτη για $\gamma=1.5$



Εφαρμογή του Θεωρήματος Cramer για την Poisson κατανομή

Το πρόβλημά μας και εδώ είναι η εκτίμηση της ταχύτητας σύγκλισης της πιθανότητας

$$P(S_n \geq nx) \rightarrow 0$$

Άρα πρέπει να υπολογίσουμε την $I(x) = \sup_{\lambda} (\lambda x - \log M_{\mu}(\lambda))$.

Αρχικά υπολογίζουμε την ροπογεννήτρια $M_{\mu}(\lambda)$

$$M_{\mu}(\lambda) = E[e^{\lambda x}] = \sum_0^{+\infty} e^{\lambda x} e^{-\theta} \frac{\theta^k}{(k!)}$$

$$M_{\mu}(\lambda) = e^{-\theta} \sum_0^{+\infty} \left(\frac{e^{\lambda} \theta}{(k!)} \right)^k = e^{-\theta} e^{e^{\lambda} \theta} = e^{\theta(e^{\lambda} - 1)}$$

Οπότε τώρα το $\log M_{\mu}(\lambda) = \theta(e^{\lambda} - 1)$

Τώρα λοιπόν η συνάρτηση ταχύτητας είναι η εξής $I(x) = \sup_{\lambda} (\lambda x - \theta(e^{\lambda} - 1))$ και θα πάρουμε την παράγωγο της συνάρτησης $A(\lambda, x) = \lambda x - \theta(e^{\lambda} - 1)$ ώστε να βρούμε το μέγιστο της $I(x)$.

$$\frac{d}{d\lambda} (A(\lambda, x)) = \frac{d}{d\lambda} (\lambda x - \theta(e^{\lambda} - 1)) = x - \theta e^{\lambda}$$

Από την παράγωγο βλέπουμε λοιπόν ότι όσο το λ μεγαλώνει, τόσο η παράγωγος γίνεται αρνητική. Από την άλλη όσο το λ μικραίνει, τόσο ο δεύτερος όρος πηγαίνει στο 0 οπότε η παράγωγος εξαρτάται κυρίως από το x . Στην συγκεκριμένη περίπτωση όμως βλέπουμε ότι το x για να έχει τοπικό ακρότατο η συνάρτηση πρέπει να είναι μεγαλύτερο από το 0 ή ίσο, γιατί αλλιώς η παράγωγος θα είναι παντού αρνητική, αφού ο δεύτερος όρος είναι παντού αρνητικός εκτός από την περίπτωση που το $\lambda \rightarrow -\infty$, τότε ο δεύτερος όρος απλά γίνεται 0, οπότε θα είναι γνησιώς φθίνουσα παντού για $x < 0$. Άρα μιλάμε μόνο για $x \geq 0$ και ψάχνουμε να βρούμε τοπικό ακρότατο της συνάρτησης. Σε αυτή τη περίπτωση η

παράγωγος μηδενίζεται για $\lambda_* = \log \frac{x}{\theta}$ όπου είναι τοπικό μέγιστο της συνάρτησης γιατί

όσο το λ μεγαλώνει η παράγωγος μικραίνει και όσο το λ μικραίνει η παράγωγος μεγαλώνει. Οπότε το λ_* είναι το σημείο τοπικού μεγίστου και πάμε να το να το αντικαταστήσουμε έτσι ώστε να πάρουμε την $I(x)$.

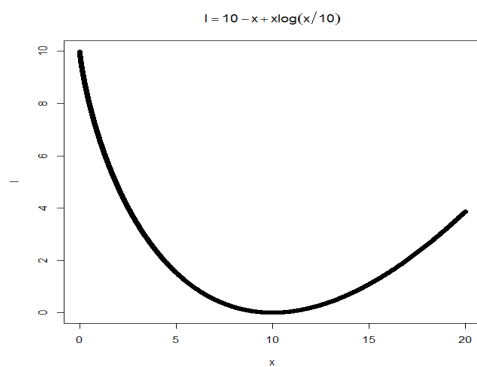
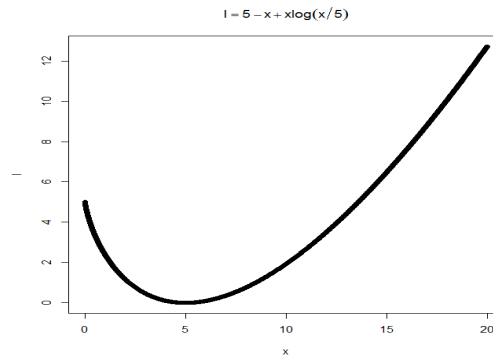
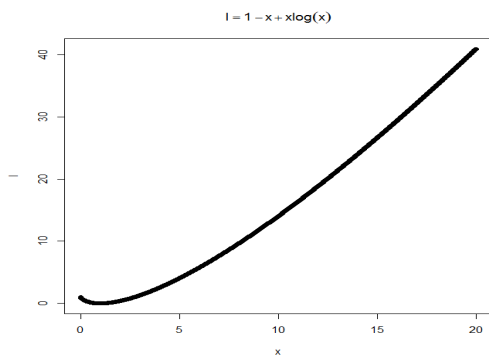
$$I(x) = \sup_{\lambda} (\lambda x - \theta(e^{\lambda} - 1)) \quad x \geq 0$$

$$I(x) = \lambda_* x - \theta(e^{\lambda_*} - 1) \quad x \geq 0$$

Τώρα αντικαθιστούμε το $\lambda_* = \log \frac{x}{\theta}$ έτσι ώστε να πάρουμε την τελική μορφή της συνάρτησης ταχύτητας $I(x)$.

$$I(x) = \theta - x + x \log\left(\frac{x}{\theta}\right) \quad x \geq 0 \quad \text{και} \quad I(x) = +\infty \quad x < 0$$

Θα παραθέσουμε τρία γραφήματα με τη βοήθεια του στατιστικού πακέτου της R για την γραφική απεικόνιση της συνάρτησης ταχύτητας για Poisson κατανομή με $\theta=1$ για το πρώτο γράφημα με $\theta=5$ για το δεύτερο γράφημα και με $\theta=10$ για το τρίτο γράφημα.



Εφαρμογή του Θεωρήματος Cramer για την Κανονική κατανομή

Το προβλημα μας και εδώ είναι η εκτίμηση της ταχύτητας σύγκλισης της πιθανότητας και το πρόβλημα μας αφορά κανονική κατανομή με διασπορά σ^2 και μέση τιμή μ .

$$P(S_n \geq nx) \rightarrow 0$$

Άρα πρέπει να υπολογίσουμε για μια ακόμα φορά την $I(x) = \sup_{\lambda} (\lambda x - \log M_{\mu}(\lambda))$.

Αρχικά υπολογίζουμε την ροπογεννήτρια $M_{\mu}(\lambda)$

$$M_{\mu}(\lambda) = \int e^{\lambda x} d\mu(x) = \int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) e^{\lambda x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) e^{\frac{2\sigma^2 \lambda x - (x-\mu)^2}{2\sigma^2}}$$

$$M_{\mu}(\lambda) = \int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) e^{\frac{2\sigma^2 \lambda x - (x-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \int_{-\infty}^{+\infty} e^{\frac{2\sigma^2 \lambda x - x^2 + 2x\mu - \mu^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \int_{-\infty}^{+\infty} e^{\frac{2\sigma^2 \lambda x - x^2 + 2x\mu - \mu^2}{2\sigma^2}}$$

$$M_{\mu}(\lambda) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \int_{-\infty}^{+\infty} e^{\frac{-(x-(\mu+\sigma^2\lambda))^2 + \sigma^4\lambda^2 + 2\sigma^2\mu\lambda}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) e^{\mu\lambda + \frac{(\sigma^2\lambda^2)}{2}} \int_{-\infty}^{+\infty} e^{\frac{-(x-(\mu+\sigma^2\lambda))^2}{2\sigma^2}}$$

$M_{\mu}(\lambda) = e^{\mu\lambda + \frac{(\sigma^2\lambda^2)}{2}}$ γιατί το άλλο ολοκλήρωμα μαζί με τον όρο που έχουμε μπροστά κάνει 1 αφού είναι ολοκλήρωμα κανονικής κατανομής με μέση τιμή $\mu + \sigma^2\lambda$ και διασπορά σ^2

Τώρα λοιπόν το $\log M_{\mu}(\lambda) = \mu\lambda + \left(\frac{\sigma^2\lambda^2}{2} \right)$

Άρα είμαστε πλέον έτοιμοι να βρούμε την συνάρτηση ταχύτητας

$$I(x) = \sup_{\lambda} \left(\lambda x - \mu\lambda - \left(\frac{\sigma^2\lambda^2}{2} \right) \right)$$

και θα πάρουμε την παράγωγο της συνάρτησης $A(x) = \lambda x - \mu\lambda - \left(\frac{\sigma^2\lambda^2}{2} \right)$ για να βρεθεί το μέγιστο της $I(x)$

$$\frac{d}{d\lambda} (A(\lambda, x)) = \frac{d}{d\lambda} \left(\lambda x - \mu\lambda - \left(\frac{\sigma^2\lambda^2}{2} \right) \right) = x - \mu - \sigma^2\lambda$$

Από το αποτέλεσμα λοιπόν εξάγουμε το συμπέρασμα ότι όσο το λ αυξάνει τόσο η παράγωγος

μειώνεται και όσο το λ μειώνεται τόσο η παράγωγος αυξάνεται, οπότε το τοπικό ακρότατο που έχουμε είναι μέγιστο. Η παράγωγος μηδενίζεται στο σημείο $\lambda_* = \frac{(x-\mu)}{\sigma^2}$ άρα αυτό είναι το σημείο τοπικού

μεγίστου οπότε πάμε να το να το αντικαταστήσουμε έτσι ώστε να πάρουμε την $I(x)$.

$$I(x) = \sup_{\lambda} \left(\lambda x - \mu\lambda - \left(\frac{\sigma^2\lambda^2}{2} \right) \right) \quad x \in \mathbb{R}$$

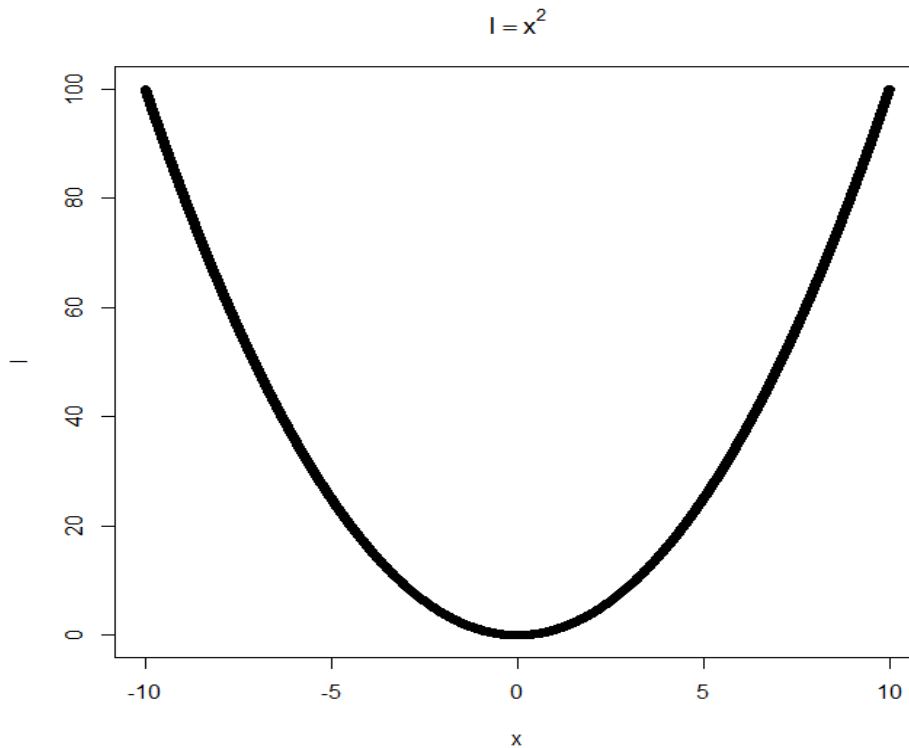
Τώρα αντικαθιστούμε το $\lambda_* = \frac{(x-\mu)}{\sigma^2}$ έτσι ώστε να πάρουμε την συνάρτηση $I(x)$

$$I(x) = \sup_{\lambda} (\lambda x - \mu \lambda - (\frac{\sigma^2 \lambda^2}{2})) \quad x \in \mathbb{R}$$

$$I(x) = (\frac{x-\mu}{\sigma^2})x - \mu(\frac{x-\mu}{\sigma^2}) - (\frac{\sigma^2(\frac{x-\mu}{\sigma^2})^2}{2}) \quad x \in \mathbb{R}$$

$$I(x) = \frac{(x-\mu)^2}{2\sigma^2} \quad x \in \mathbb{R}$$

Τέλος θα παραθέσουμε και ένα γράφημα της συνάρτησης ταχύτητας για την κανονική κατανομή με $\mu=0$ και $\sigma^2=1$.



Κεφάλαιο 2

2.1 Αρχή των Μεγάλων Αποκλίσεων

Ορισμός Αρχής των Μεγάλων Αποκλίσεων(LDP)

Λέμε ότι μια οικογένεια κατανομών $\{\mu_\varepsilon\}_{\varepsilon>0}$ ικανοποιεί την LDP με συνάρτηση ταχύτητας I όταν ισχύουν τα παρακάτω

$$(\alpha) \quad \lim_{n \rightarrow +\infty} \sup \frac{1}{n} \log \mu_n(K) \leq \inf_{x \in K} I(x) \quad , \quad \forall K \subset \mathbb{R}^d \text{ κλειστά}$$

$$(\beta) \quad \lim_{n \rightarrow +\infty} \inf \frac{1}{n} \log \mu_n(O) \geq \inf_{x \in O} I(x) \quad , \quad \forall O \subset \mathbb{R}^d \text{ ανοιχτά}$$

2.2 Απόδειξη LDP με χρήση του Θεωρήματος Cramer

Θα χρησιμοποιήσουμε το Θεώρημα Cramer για να δείξουμε ότι, αν μ_n είναι η κατανομή της

$\frac{(X_1 + \dots + X_n)}{n}$ τότε η $\{\mu_n\}_n$ ικανοποιεί την LDP με συνάρτηση ταχύτητας I , δηλαδή θα έχουμε

$$(α) \quad \lim_{n \rightarrow +\infty} \sup \frac{1}{n} \log \mu_n(K) \leq \inf_{x \in K} I(x) \quad , \quad \forall K \subset \mathbb{R}^d \text{ κλειστά}$$

$$(β) \quad \lim_{n \rightarrow +\infty} \inf \frac{1}{n} \log \mu_n(O) \geq \inf_{x \in O} I(x) \quad , \quad \forall O \subset \mathbb{R}^d \text{ ανοιχτά}$$

Απόδειξη

Δείχνουμε πρώτα το κάτω φράγμα. Έστω λοιπόν $O \subset \mathbb{R}^d$ ανοιχτό. Αν $\bar{x} := \int x d\mu(x) \in O$ τότε $\mu_n(O) \rightarrow 1$ καθώς $n \rightarrow +\infty$ και $\inf_{x \in O} I(x) = 0$ οπότε

$$\lim_{n \rightarrow +\infty} \inf \frac{1}{n} \log \mu_n(O) = 0 = -\inf_{x \in O} I(x)$$

Έστω τώρα ότι $\bar{x} \notin O$. Για κάθε $x \in O$ επιλέγουμε $\delta_x > 0$ τέτοιο ώστε $(x - \delta_x, x + \delta_x) \subset O$ και όπως έχουμε δει στην απόδειξη του Θεωρήματος Cramer για κάθε $x \in O$ έχουμε ότι

$$\lim_{n \rightarrow +\infty} \mu_n(O) \geq \lim_{n \rightarrow +\infty} \mu_n(x - \delta_x, x + \delta_x) \geq \lim_{n \rightarrow +\infty} \mu_n[x, x + \delta_x] \geq -I(x)$$

Παίρνοντας στην παραπάνω ανισότητα supremum ως προς όλα τα $x \in O$ έπεται ότι

$$\lim_{n \rightarrow +\infty} \mu_n(O) \geq \sup_{x \in O} -I(x) = -\inf_{x \in O} I(x)$$

Αποδεικνύουμε τώρα το άνω φράγμα. Αν $\bar{x} \in K$ τότε προφανώς έχουμε ότι

$$\limsup_{n \rightarrow +\infty} \mu_n(K) \leq 0 = -\inf_{x \in K} I(x)$$

Έστω λοιπόν ότι $\bar{x} \notin K$. Αφού το K είναι κλειστό υπάρχει $\varepsilon > 0$ τέτοιο ώστε $(\bar{x} - \varepsilon, \bar{x} + \varepsilon) \cap K = \emptyset$.

Έτσι αν (a, b) είναι το μέγιστο ανοικτό διάστημα J που περιέχει το \bar{x} και $K \cap J = \emptyset$ τότε

$a < \bar{x} < b$ και $F \subset (-\infty, a] \cup [b, +\infty)$ οπότε από

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log c_n = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log(a_n \vee b_n) \text{ θα έχουμε ότι}$$

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(K) \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n((-\infty, a] \cup [b, +\infty)) = -[I(a) \vee I(b)]$$

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(K) \leq -\inf_{x \in K} I(x)$$

όπου η τελευταία ισότητα έπεται από το ότι $a, b \in K$ και $K \cap (a, b) = \emptyset$

2.3 Λήμμα Varadhan

Πριν διατυπώσουμε και αποδείξουμε το λήμμα Varadhan να αναφέρουμε ότι μια συνάρτηση ταχύτητας $I(x)$ ονομάζεται “καλή” αν για κάθε $M > 0$ τα σύνολα $\{x : I(x) \leq M\}$ είναι συμπαγή

Λήμμα Varadhan

Έστω μια οικογένεια μέτρων πιθανότητας $\{\mu_\varepsilon\}_{\varepsilon > 0}$ που ακολουθούν την LDP με μία “καλή” συνάρτηση ταχύτητας I . Τότε για κάθε συνάρτηση $F : X \rightarrow \mathbb{R}$ άνω φραγμένη και συνεχή έχουμε :

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \int_X e^{\frac{F(x)}{\varepsilon}} d\mu_\varepsilon(x) = \sup_{x \in X} [F(x) - I(x)] \quad (2)$$

Απόδειξη

Κάτω φράγμα : Από την συνέχεια της $F(x)$ για κάθε $x \in X$ και $\delta > 0$ υπάρχει περιοχή U_x του x τέτοια ώστε : $F(y) > F(x) - \delta$, για κάθε $y \in U_x$. Έτσι,

$$\int e^{\frac{F(y)}{\varepsilon}} d\mu_\varepsilon(y) \geq \int_{U_x} e^{\frac{F(y)}{\varepsilon}} d\mu_\varepsilon(y) \geq e^{\frac{F(x) - \delta}{\varepsilon}} \mu_\varepsilon(U_x)$$

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log \int_X e^{\frac{F(x)}{\varepsilon}} d\mu_\varepsilon(x) \geq F(x) - \delta - \inf_{U_x} I(y) \geq F(x) - I(x) - \delta$$

Παίρνοντας $\delta \rightarrow 0$ και το supremum για $x \in X$ στο δεξί σκέλος, έχουμε το κάτω φράγμα οπότε το δείξαμε για το κάτω φράγμα.

Άνω φράγμα :Για κάθε $x \in X$ υπάρχει (από την συνέχεια της F και την κάτω ημισυνέχεια της I) περιοχή $U_x \ni x$ ώστε

$$F(y) < F(x) + \delta \text{ και } I(y) > I(x) - \delta, \text{ για κάθε } y \in U_x$$

Για κάθε $M > 0$ το σύνολο $\Psi_I(M)$ είναι συμπαγές και η $\cup_{x \in \Psi_I(M)} U_x$ είναι μια ανοιχτή κάλυψη του.

Ας θεωρήσουμε λοιπόν μια πεπερασμένη υποκάλυψη $\{U_{x_i}\}_{1 \leq i \leq N}$ και το κλειστό σύνολο

$$F = (\cup_{i=1}^N U_{x_i})^c.$$

$$\int e^{\frac{(F(y))}{\varepsilon}} d\mu_\varepsilon(y) = \sum_{i=1}^N \int_{U_{x_i}} e^{\frac{(F(y))}{\varepsilon}} d\mu_\varepsilon(y) + \int_F e^{\frac{(F(y))}{\varepsilon}} d\mu_\varepsilon(y)$$

$$\int e^{\frac{(F(y))}{\varepsilon}} d\mu_\varepsilon(y) \leq \sum_{i=1}^N e^{\frac{(F(x_i + \delta))}{\varepsilon}} \mu_\varepsilon(U_{x_i}) + \int_F e^{\frac{(F(y))}{\varepsilon}} d\mu_\varepsilon(y)$$

Από το άνω φράγμα των ανισοτήτων των μεγάλων αποκλίσεων έχουμε

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \int_X e^{\frac{(F(x))}{\varepsilon}} d\mu_\varepsilon(x) \leq \max \left\{ F(x_i) + \delta - \inf_{U_{x_i}} I(y), (\sup F - \inf_{x \in F} I(x)) \right\}$$

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \int_X e^{\frac{(F(x))}{\varepsilon}} d\mu_\varepsilon(x) \leq \max \left\{ F(x_i) + 2\delta - I(x_i), (\sup F - M) \right\}$$

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \int_X e^{\frac{(F(x))}{\varepsilon}} d\mu_\varepsilon(x) \leq \max \left\{ \sup_{x \in X} (F(x) - I(x)), (\sup F - M) \right\} + 2\delta$$

Αφήνουμε τώρα το δ να πάει στο 0 και το M στο άπειρο και η απόδειξη ολοκληρώθηκε.

2.4 Θεώρημα Gartner-Ellis

Θεώρημα Gartner-Ellis

Πρέπει να δείξουμε ότι το όριο $\Lambda(t) = \lim_{\varepsilon \rightarrow 0} \varepsilon \Lambda_\varepsilon\left(\frac{t}{\varepsilon}\right) = \lim_{\varepsilon \rightarrow 0} \varepsilon \log \int e^{\frac{\langle x, t \rangle}{\varepsilon}} d\mu_\varepsilon(x)$ υπάρχει για

κάθε $t \in \mathbb{R}^d$ και είναι Λ παραγωγίσιμη συνάρτηση του t .

Θα ορίσουμε αρχικά τον μετασχηματισμό Legendre του I η οποία είναι ο εξής

$$I(x) = A^*(x) = \sup_{t \in \mathbb{R}^d} [\langle x, t \rangle - I(t)] \quad , \quad x \in \mathbb{R}^d$$

$$A^*(S) = \inf_{x \in S} A^*(x) \quad , \quad S \subset \mathbb{R}^d \quad \text{όπου για την } I(x) \text{ ισχύουν οι παρακάτω οι υποθέσεις}$$

Επίσης υποθέτουμε ότι είτε $D_I^\circ = \mathbb{R}^d$ είτε για την I ισχύει ότι $\lim_{t \rightarrow \partial D_I; t \in \partial D_I} |\nabla \Lambda(t)| = \infty$ και $0 \in D_I^\circ$

$$(\alpha) \quad \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(C) \leq -A^*(C) = \inf_{x \in C} A^*(x) \quad , \quad \forall C \subset \mathbb{R}^d \text{ κλειστά}$$

$$(\beta) \quad \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(O) \geq -A^*(O) = \inf_{x \in O} A^*(x) \quad , \quad \forall O \subset \mathbb{R}^d \text{ ανοιχτά}$$

Απόδειξη

(α) Πρώτα λοιπόν θα κάνουμε την απόδειξη για το πάνω φράγμα . Θα πάρουμε ένα $\delta > 0$. Όπου για $x \in \mathbb{R}^d$ θα ισχύει

$$A_\delta^*(x) = \min(A^*(x) - \delta, \frac{1}{\delta})$$

Για κάθε $x \in \mathbb{R}^d$ υπάρχει ένα $t_x \in \mathbb{R}^d$ έτσι ώστε

$$[\langle x, t_x \rangle - I(t_x)] \geq A_\delta^*(x)$$

Ακόμα για κάθε $x \in \mathbb{R}^d$ υπάρχει μια γειτονία του A_x όπου για κάθε x ισχύει οτι

$$\inf_{y \in A_x} \langle y - x, t_x \rangle \geq -\delta$$

Χρησιμοποιώντας την ανισότητα Chebyshev προκύπτει

$$P_n(A_x) = P(Z_n \in A_x) \leq P(\langle Z_n - x, t_x \rangle \geq -\delta)$$

$$P_n(A_x) \leq e^{\delta n} E(e^{n \langle Z_n - x, t_x \rangle}) = e^{\delta n} \varphi_n(nt_x) e^{-n \langle x, t_x \rangle}$$

και έστω ένα σύνολο $K \subset \mathbb{R}^d$ συμπαγές. Τότε το κάλυμμα $\cup_{x \in K} A_x$ του K έχει πεπερασμένο υποκάλυμμα $\cup_{i=1, \dots, N} A_{x_i}$

$$\frac{1}{n} \log P_n(K) \leq \frac{1}{n} \log [N \max_{i=1, \dots, N} P_n(A_{x_i})]$$

$$\frac{1}{n} \log P_n(K) \leq \frac{1}{n} \log N + \delta - \min_{i=1, \dots, N} [\langle x_i, t_{x_i} \rangle - \frac{1}{n} \log \varphi_n(nt_{x_i})]$$

Για $n \rightarrow +\infty$ παίρνουμε ότι

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(K) \leq \delta - \min_{i=1, \dots, N} [\langle x_i, t_{x_i} \rangle - I(t_{x_i})]$$

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(K) \leq \delta - \min_{i=1, \dots, N} A_\delta^*(x_i) \leq \delta - A_\delta^*(K)$$

Με το $\delta \downarrow 0$ παίρνουμε

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(K) \leq -A^*(K)$$

Η επέκταση απο συμπαγή σύνολα σε κλειστά μας δείχνει την εκθετική “σφιχτότητα” της P_n .

Φυσικά αν C είναι κλειστό τότε $C \cap [-N, N]^d$ είναι συμπαγής για όλα τα $N > 0$ και αυτό θα έχει ως αποτέλεσμα

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(C) \leq \max(-A^*(C \cap [-N, N]^d), -M_N) \quad \forall N > 0$$

όπου το M_N είναι το ελάχιστο

$$-M_N = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(\mathbb{R}^d \setminus [-N, N]^d)$$

Εάν καταφέρουμε να αποδείξουμε ότι $\lim_{N \rightarrow +\infty} M_N = \infty$ στην συνέχεια θα ξέρουμε ότι

$$\lim_{n \rightarrow +\infty} -A^*(C \cap [-N, N]^d) = A^*(C)$$

Παίρνουμε τα μοναδιαία διανύσματα e^1, \dots, e^d του \mathbb{R}^d . Αφού $0 \in D_f^\circ$ τότε υπάρχει $\delta > 0$ ώστε

$I(-\delta e^i) < \infty$ και $I(\delta e^i) < \infty$ για όλα τα i . Από την Ανίσωτητα Chebyshev προκύπτει ότι για όλα τα i ισχύει

$$P(Z_n^i \leq -N) \leq e^{-n\delta N} \varphi_n(-n\delta e^i)$$

$$P(Z_n^i \geq N) \leq e^{-n\delta N} \varphi_n(n\delta e^i)$$

Για το M_N λοιπόν

$$-M_N = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_n(\exists i \in \{1, \dots, d\} : Z_n^i \notin [-N, N])$$

$$-M_N \leq -\delta N + \max_{i=1, \dots, d} \max(A(-\delta e_i), A(\delta e_i)),$$

και ουσιαστικά αποδείξαμε το κομμάτι για το πάνω φράγμα τώρα θα πάμε στην περίπτωση του κάτω φράγματος.

(α) Κάτω Φράγμα

Παίρνουμε την ανοιχτή σφαίρα $B_\varepsilon(x)$ με ακτίνα $\varepsilon > 0$ γύρω από το x και πρέπει να δείξουμε ότι

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(B_\varepsilon(x)) \geq -A^*(x) \quad \forall x \in E$$

Φυσικά ισχύει

$$P_n(O) \geq P_n(B_\varepsilon(x)) \quad \forall x \in O \quad \forall \varepsilon \leq \varepsilon_0(x)$$

Αρχικά φιξάρουμε ένα $x \in E$ και παίρνουμε ένα $\tau \in D_I^\circ$. Ακόμα $\varphi_n(n\tau) < \infty$ για αρκετά μεγάλο n

$$\frac{(d \hat{P}_n)}{(d P_n)}(y) = \frac{1}{(\varphi_n(n\tau))} e^{n\langle y, \tau \rangle} \quad y \in \mathbb{R}^d$$

$$\frac{1}{n} \log P_n(B_\varepsilon(x)) = \frac{1}{n} \log \int_{B_\varepsilon(x)} P_n(dy)$$

$$\frac{1}{n} \log P_n(B_\varepsilon(x)) = \frac{1}{n} \log \varphi_n(n\tau) + \frac{1}{n} \log \int_{B_\varepsilon(x)} e^{-n\langle y, \tau \rangle} \hat{P}_n(dy)$$

$$\frac{1}{n} \log P_n(B_\varepsilon(x)) \geq \frac{1}{n} \log \varphi_n(n\tau) - \langle x, \tau \rangle - \varepsilon |\tau| + \frac{1}{n} \log \hat{P}_n(B_\varepsilon(x))$$

Η τελευταία ανισότητα χρησιμοποιεί ότι $|y - x| \leq \varepsilon$ για $y \in B_\varepsilon(x)$

Ως εκ τούτου φτάνουμε στο ότι

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(B_\varepsilon(x)) \geq I(\tau) - \langle x, \tau \rangle + \lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \hat{P}_n(B_\varepsilon(x))$$

Επειδή $\lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_n(B_\varepsilon(x)) \geq -A^*(x)$ μας δίνει ότι

$$\lim_{\varepsilon \downarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \hat{P}_n(B_\varepsilon(x)) = 0$$

Για το θεώρημα Cramer για ανεξάρτητες τυχαίες μεταβλητές μπορεί να αποδειχτεί πολύ εύκολα εδώ όμως θα χρησιμοποιήσουμε κάτι διαφορετικό. Παίρνουμε την σχέση μεταξύ της φ και της $\hat{\varphi}$

$$\hat{\varphi}_n(nt) = \frac{(\varphi_n(n(t+\tau)))}{(\varphi_n(n\tau))}$$

Ως εκ τούτου

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \log \hat{\phi}_n(nt) \right) = \hat{I}(t) = I(t+\tau) - I(\tau) \quad t \in \mathbb{R}^d$$

Ακόμα έχουμε την σχέση

$$A^*(x) = \sup_{t \in \mathbb{R}^d} [\langle x, t \rangle - \hat{I}(t)] = A^*(x) - \langle x, \tau \rangle + I(\tau)$$

Επίσης αφού $B_\varepsilon(x)$ είναι ανοιχτό τότε το $\mathbb{R}^d \setminus B_\varepsilon(x)$ είναι κλειστό οπότε έχουμε ότι

$$\lim_{n \rightarrow +\infty} \sup \frac{1}{n} \log P_n(\mathbb{R}^d \setminus B_\varepsilon(x)) \leq -I^*(\mathbb{R}^d \setminus B_\varepsilon(x)) \quad \forall \varepsilon > 0$$

$$A^*(\mathbb{R}^d \setminus B_\varepsilon(x)) = A^*(x_0) \quad \text{για κάποιο } x_0 \neq x$$

Ακόμα για $x \in E$ και παίρνουμε ένα $\tau \in D_I^\circ$

$$\hat{A}^*(x_0) = A^*(x_0) - \langle x_0, \tau \rangle + I(\tau)$$

$$\hat{A}^*(x_0) = [A^*(x_0) - \langle x_0, \tau \rangle] + [A^*(x) - \langle x, \tau \rangle] > 0$$

Συνεπώς βρίσκουμε ότι

$$\lim_{n \rightarrow +\infty} \sup \frac{1}{n} \log \hat{P}_n(\mathbb{R}^d \setminus B_\varepsilon(x)) < 0 \quad \forall \varepsilon > 0$$

Ακόμα $\hat{P}_n(\mathbb{R}^d) = 1$ για όλα τα n , οπότε η απόδειξη τελείωσε.

Κεφάλαιο 3

Προσομοίωση σπάνιων ενδεχομένων

3.1 Μέθοδος Monte Carlo

Σε αυτό το κεφάλαιο θα προσπαθήσουμε να υπολογίσουμε το π με τη βοήθεια της μεθόδου Monte Carlo με χρήση του στατιστικού πακέτου της R. Για τον υπολογισμό του π θα πρέπει να φτιάξουμε ένα μικρο πρόγραμμα στην R που θα το υπολογίζει. Ουσιαστικά αυτό που θέλουμε να κάνει το πρόγραμμα είναι να υπολογίζει όσο δυνατόν καλύτερα το π για να το επιτύχουμε αυτό θα του δώσουμε δύο μεταβλητές x, y οι οποίες θα ακολουθούν και οι δύο την εξής κατανομή ($2*\text{runif}(1)$), δηλαδή την ομοιόμορφη κατανομή πολλαπλασιασμένη με το 2 και μειωμένη κατά 1, αυτό συμβαίνει γιατί θέλουμε αυτές οι δυο μεταβλητές σε ένα σύστημα Oxy να πέρνουν τιμές μόνο σε ένα τετράγωνο με πλευρά 2 και κορυφές του τα σημεία $(1,1), (1,-1), (-1,-1), (-1,1)$. Τώρα η σημαντικότερη συνθήκη για το πρόγραμμα είναι $x^2 + y^2 < 1$ και όταν αυτή η συνθήκη ισχύει, αυτό θα σημαίνει ότι το σημείο αυτό (x,y) θα ανήκει στο μοναδιαίο κύκλο που περιέχεται στο τετράγωνο που σχηματίσαμε. Όποτε λοιπόν συμβαίνει αυτό τότε το k του προγράμματος θα αυξάνεται κατά 1. Άρα αν για παράδειγμα έχουμε M παρατηρήσεις και οι N πέφτουν μέσα στο κύκλο τότε ο υπολογισμός του π θα είναι ο εξής

$$\frac{N}{M} \xrightarrow{M \rightarrow +\infty} \frac{\pi}{4}$$

άρα το $\pi \approx \frac{(4N)}{M}$

Ας κάνουμε πρώτα θεωρητικά τον υπολογισμό της συνάρτησης ταχύτητας $I(x)$. Η συνάρτηση ταχύτητας δίνεται από τον παρακάτω τύπο του Θεωρήματος Cramer

$$P(S_n \geq nx) \leq e^{-nI(x)}$$

Εμείς στο προβλημά μας έχουμε να υπολογίσουμε την πιθανότητα $P[|\hat{\pi}(\omega) - \pi| > x] = ?$. Για αυτή την πιθανότητα μπορούμε να γράψουμε ότι

$$P[|\hat{\pi}(\omega) - \pi| > x] = P[\hat{\pi}(\omega) - \pi > x] + P[\hat{\pi}(\omega) - \pi < -x]$$

$$P[|\hat{\pi}(\omega) - \pi| > x] = P[\hat{\pi}(\omega) > \pi + x] + P[\hat{\pi}(\omega) < \pi - x]$$

Ο πρώτος όρος του αθροίσματος πηγαίνει στο 0 όταν το $x > 4 - \pi$ ενώ ο δεύτερος όταν το $x > \pi$.

Τώρα αν λογαριθμίσουμε την παραπάνω σχέση και την διαιρέσουμε με το n σύμφωνα με το Θεώρημα

του Cramer θα πάρουμε ότι $(\frac{1}{n}) \log P[|\hat{\pi}(\omega) - \pi| > x] = -I(x)$ βέβαια μας είναι γνωστό από την

ανάλυση ότι $\log(a_n \vee b_n) \leq \log s_n \leq \log 2(a_n \vee b_n)$ αυτό ισχύει γιατί

$$\log(a_n \vee b_n) \leq \log s_n \leq \log 2(a_n \vee b_n)$$

$$\log(a_n \vee b_n) \leq \log s_n \leq \frac{\log 2}{n} + \log \frac{(a_n \vee b_n)}{n} \quad \text{όταν το } n \text{ λοιπόν πάει στο άπειρο}$$

τότε το όριο και του άνω και του κάτω φράγματος γίνονται ίσα. Σε αυτή την περίπτωση για μεγάλα n το όριο του αθροίσματος γίνεται ίσο με το μεγαλύτερο από τα δύο όρια δηλαδή στην δικιά μας περίπτωση ισχύει ότι

$$\left(\frac{1}{n}\right) \log P[|\hat{\pi}(\omega) - \pi| > x] = \max \left\{ -I_B\left(\frac{(x+\pi)}{4}\right), -I_B\left(\frac{(\pi-x)}{4}\right) \right\}$$

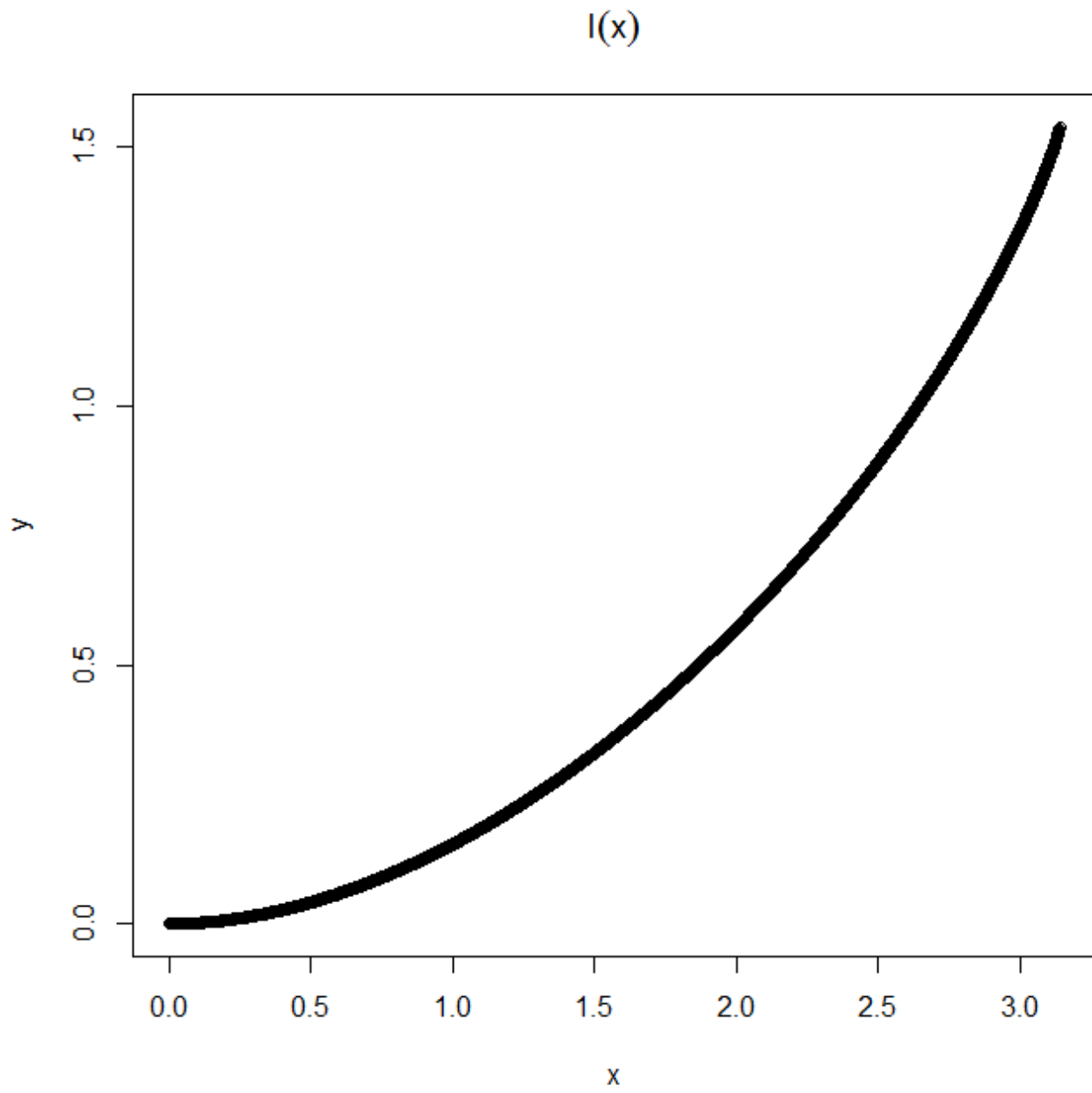
$$\left(\frac{1}{n}\right) \log P[|\hat{\pi}(\omega) - \pi| > x] = -\min \left\{ I_B\left(\frac{(x+\pi)}{4}\right), I_B\left(\frac{(\pi-x)}{4}\right) \right\}$$

τόρα ας κάνουμε ένα plot για να δούμε την θεωρητική γραφική παράσταση της $I(x)$. Ουσιαστικά πριν να κάνουμε το γράφημα γνωρίζουμε ότι στο πεδίο ορισμού της συνάρτησης ταχύτητας που είναι από το $-\pi$ μέχρι το π η I για το ένα διάστημα $(-\pi, 4-\pi)$ επιβιώνει μόνο το $I((x+\pi)/4)$ και στο διάστημα $(4-\pi, \pi)$ πέρνει τιμές μόνο από το $I((\pi-x)/4)$. Αυτό συμβαίνει γιατί ας πούμε στο πρώτο διάστημα $(-\pi, 4-\pi)$ το $I((\pi-x)/4)$ πηγαίνει στο άπειρο οπότε είναι προφανές ότι σε εκείνο το διάστημα είναι μεγαλύτερο από το $I((x+\pi)/4)$ ομοίως και για το διάστημα $(4-\pi, \pi)$. Βέβαια εμείς εν τέλει θα υπολογίσουμε μόνο το διάστημα $(0, \pi)$ γιατί δεν έχει νόημα να υπολογίσουμε την πιθανότητα $P[|\hat{\pi}(\omega) - \pi| > x]$ με $x < 0$ γιατί αυτή η πιθανότητα είναι 1.

x<-seq(0,pi,0.0001)

I<-ifelse(x<4-pi,I((x+pi)/4),I((pi-x)/4))

`plot(x,I)`



Τώρα μπορούμε να προχωρήσουμε στην δημιουργία του προγράμματος για τον υπολογισμό του π με τη βοήθεια του στατιστικού περιβάλλοντος της R.

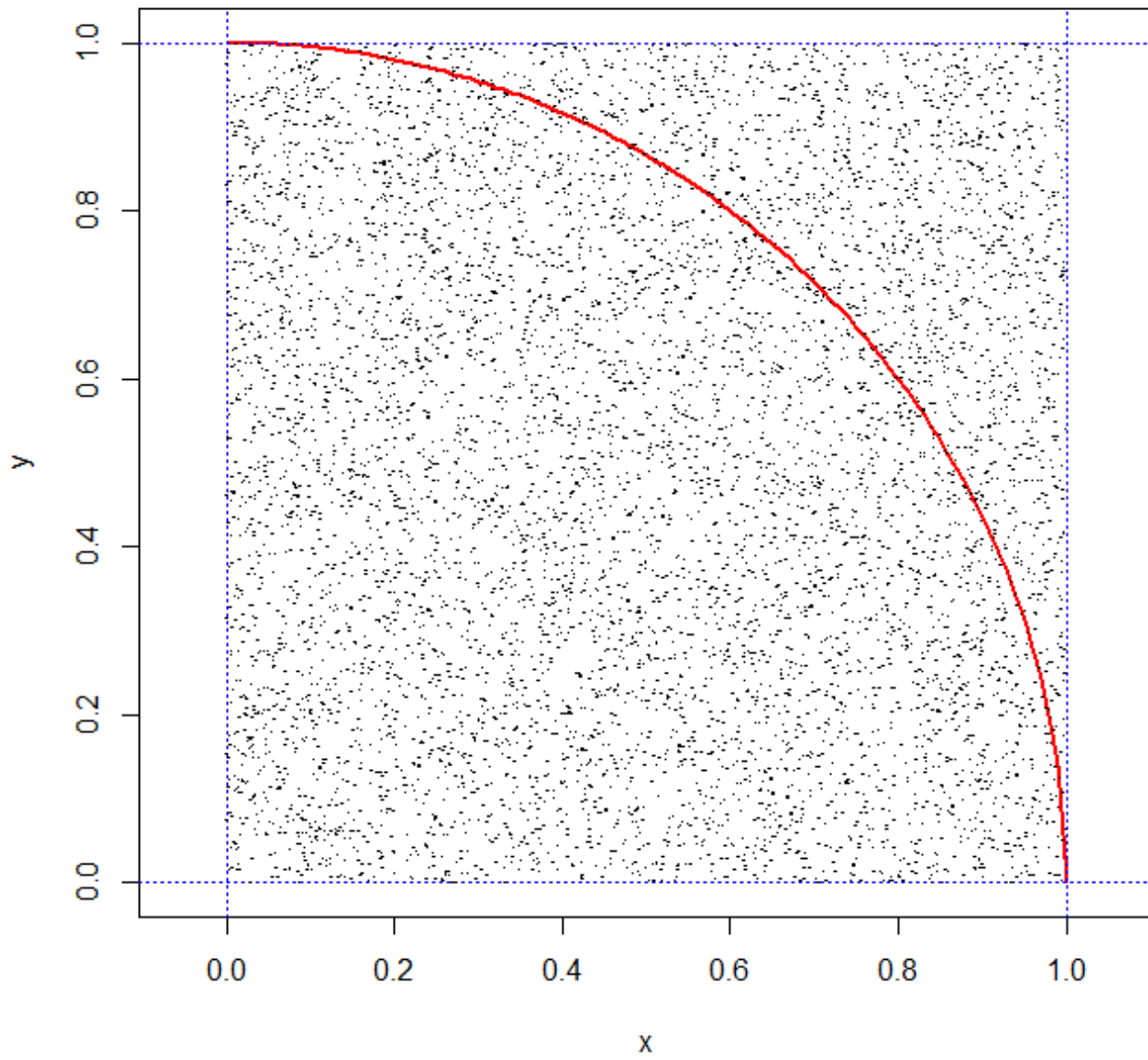
Δηλαδή η μορφή του προγράμματος που θέλουμε να φτιάξουμε σε διάγραμμα στο πρώτο τεταρτημόριο για x, y που ακολουθούν απλά την ομοιόμορφη κατανομή (ενώ στο δικό μας πρόβλημα θα ακολουθούν την ομοιόμορφη πολλαπλασιασμένη με το 2 και μειωμένη κατά 1) θα είναι κάπως έτσι :

Αρχικά ο κώδικας στην R θα είναι ο εξής

```
x <- seq(0, 1, by=0.01)
y <- sqrt(1-x^2)
plot(x, y, type="l", asp=1,xlim=c(0,1), ylim=c(0,1), col=2, lwd=2)
abline(h=0, col="blue", lty=3)
abline(v=0, col="blue", lty=3)
abline(h=1, col="blue", lty=3)
abline(v=1, col="blue", lty=3)

N <- 10000
x <- runif(N)
y <- runif(N)
points(x, y, type="p", cex=0.2, pch=18)
```

και το γράφημα που παίρνουμε είναι το παρακάτω



Το πρόγραμμα λοιπόν που θα φτιάξουμε είναι το εξής

```
p=function(n,M,t)
{
  dian=rep(1,n)
  for(i in 1:n)
  {
    k=0
    for (j in 1:M) {
      x <- 2*runif(1)-1 ; y <- 2*runif(1)-1;

      if (x^2+y^2<1)
      {
        k=k+1
      }
      dian[i]=4*k/M
    }
  }
  er=rep(1,n)
  for(i in 1:n)
  {
    er[i]=abs(dian[i]-pi)
  }
  v=0
  for(j in 1:n)
  {
    if(er[j]>t)
    {
      v=v+1
    }
  }
  return(v/n)
}
```

Αυτό τον αλγόριθμο θα τον χρησιμοποιήσουμε για αρκετά M για να υπολογίσουμε το π και για να δούμε κιόλας πόσο γρήγορα πάει στο 0 η παρακάτω πιθανότητα

$$P[|\hat{\pi}(\omega) - \pi| > x] = ?$$

Στο παραπάνω πρόγραμμα με t συμβολίζουμε το x .

Θα πάρουμε περιπτώσεις για $M = \{1000, 2000, 4000, 8000, 10000, 20000\}$

Με σταθερό $x = \frac{1}{100}$

Επίσης από το Θεώρημα του Cramer ξέρουμε ότι

$$-\left(\frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right)\right) = I(x)$$

Για να βρούμε τώρα το $I(x)$ θα χρησιμοποιήσουμε το `plot(-logp,M)`.

```
M<-c(1000,2000,4000,8000,10000,20000)
```

```
logp<-c(0.0728,0.1040,0.1536,0.2204,0.2823,0.4168)
```

```
plot(M,logp)
```

```
title(expression(I(x)))
```

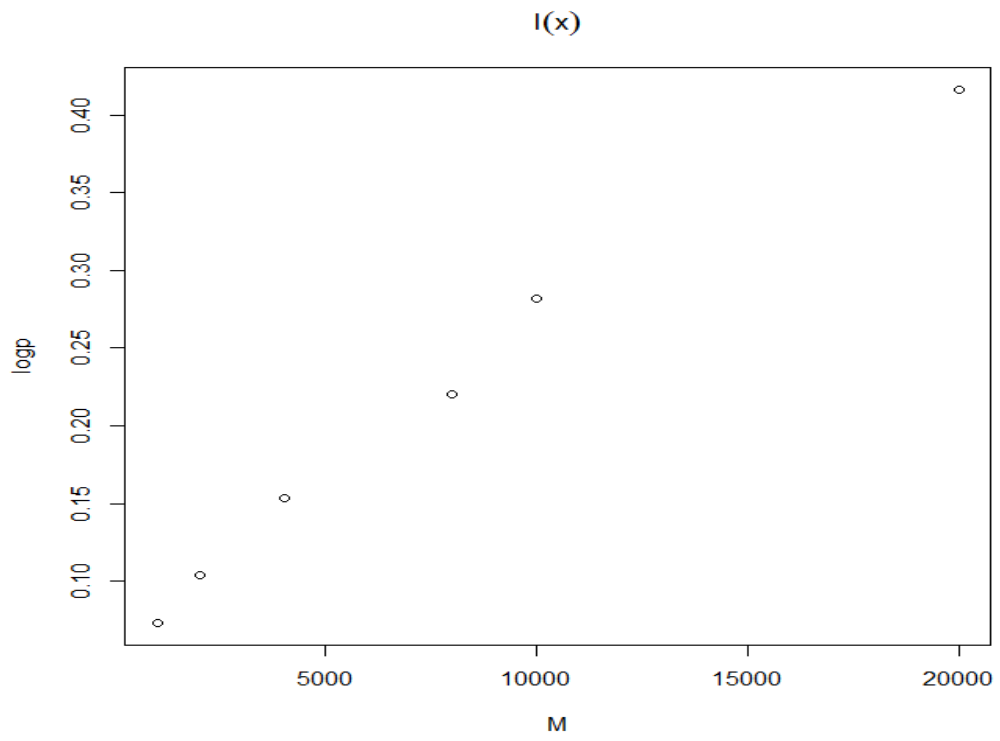
```
lines(M,logp)
```

Για τις έξι παραπάνω περιπτώσεις με σταθερό το x .

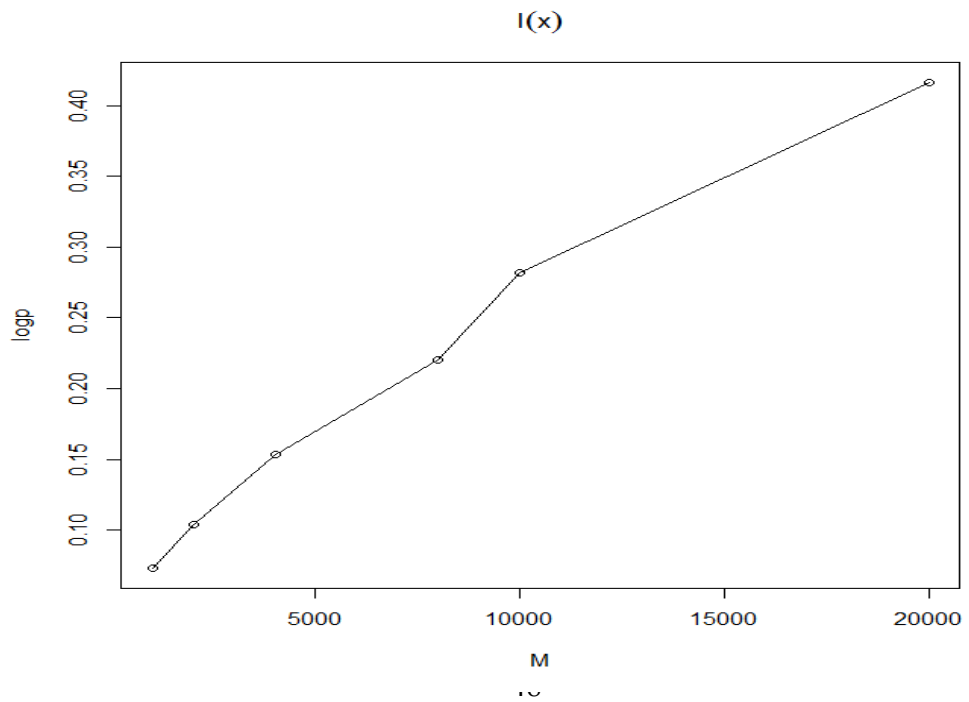
Στην επόμενη σελίδα έχουμε τα δύο γραφήματα που μας δείχνουν ουσιαστικά μια ευθεία όπου η κλίση αυτής της ευθείας είναι η συνάρτηση $I(x)$. Αρχικά κάνουμε ένα `plot(M,logp)` που φαίνονται μόνο τα σημεία και στην συνέχεια δίνουμε την εντολή `lines(M,logp)` έτσι ώστε να σχηματιστεί η ευθεία που

ζητάμε. Η κλίση $I(x)$ μας βγαίνει περίπου $I(x) \approx \left(\frac{1}{3}\right) 10^{-5}$

plot(M,logp)



lines(M,logp)



Ακόμα θα δείξουμε το MC σφάλμα για την εκτίμηση του π . Θα κάνουμε 100 εκτιμήσεις της ίδιας ποσότητας θα υπολογίσουμε τη διασπορά αυτών των αποτελεσμάτων και θα κάνουμε και το ιστόγραμμα τους.

Ο κώδικας που θα χρησιμοποιήσουμε είναι ο παρακάτω

```
p=function(n,M)
{
  dian=rep(1,n)
  for(i in 1:n)
  {
    k=0
    for (j in 1:M) {
      x <- 2*runif(1)-1 ; y <- 2*runif(1)-1;

      if (x^2+y^2<1)
      {
        k=k+1
      }
      dian[i]=4*k/M
    }
  }
  return(dian)
}
```

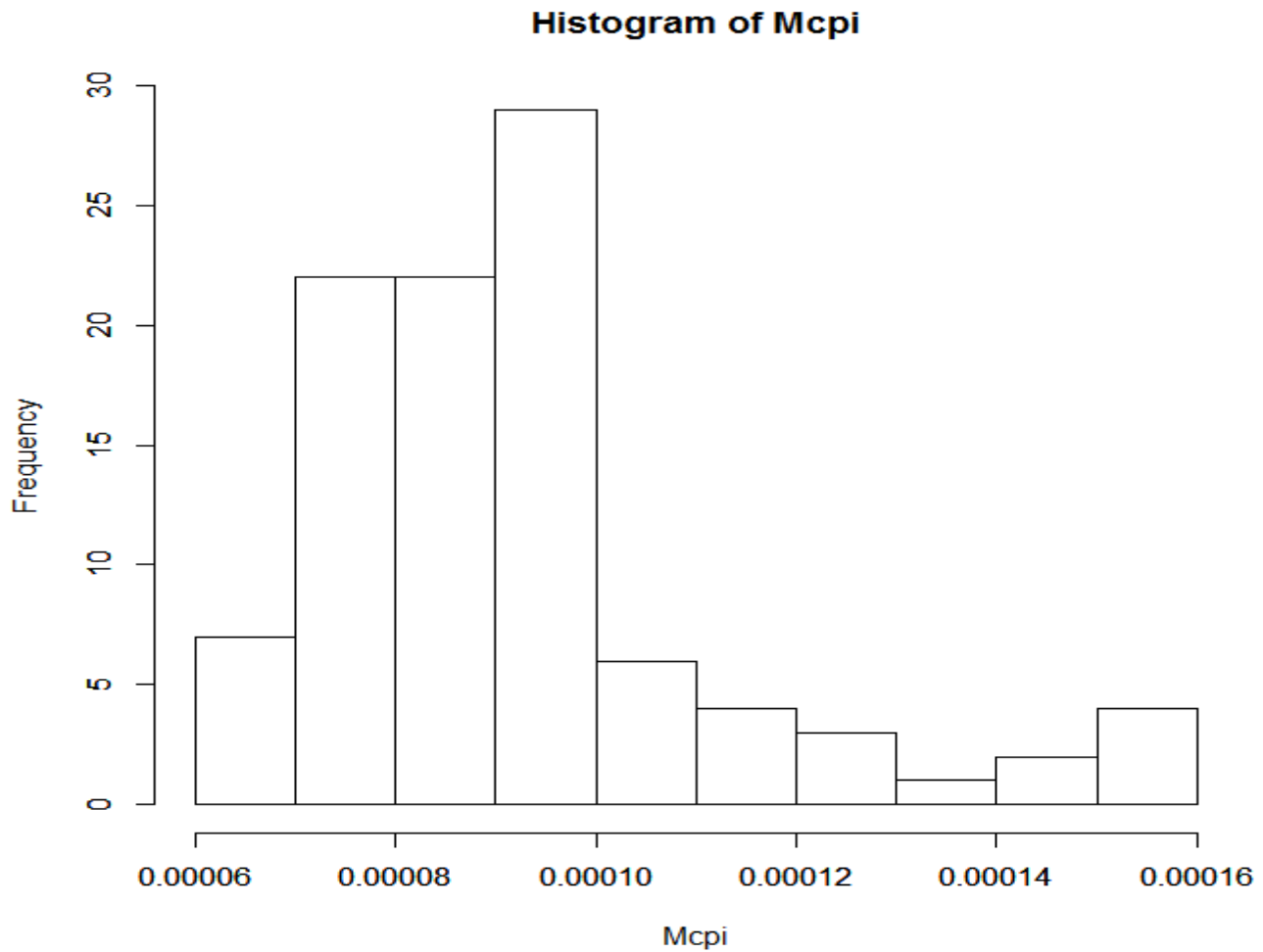
και για τον υπολογισμό του MC σφάλματος

```
mc=function(r)
{
  dian1=rep(1,r)
  for(e in 1:r)
  {
    p(1000,1000)
    x=p(1000,1000)
    j=0
    m=rep(NA,40)
    l=rep(NA,25)
    for(i in 1:40)
    {
      l=x[j:(0+i*25)]
      j=j+25
      m[i]<-mean(l)
    }
  }
}
```

```

dia1[e]=var(m)
}
return(dian1)
}
Mcp1=mc(100)
var(Mcp1)
[1] 4.084299e-10
hist(Mcp1)

```



Θα κάνουμε και μια γραφική παράσταση για να δούμε πώς συμπεριφέρεται ακριβώς η συνάρτηση I σύμφωνα με τη θεωρία και μετά σύμφωνα με τη μέθοδο Monte Carlo. Αυτή τη γραφική παράσταση θα την κάνουμε προφανώς για x στα οποία να μπορεί να πάρει τιμές η μέθοδος Monte Carlo δηλαδή περίπου για μέχρι το x να φτάνει στο 0.2

```
x<-seq(0,0.2,0.0001)
```

```
I<-ifelse(x<0,((pi+x)/4)*(log((pi+x)/(pi)))+(1-((pi+x)/4))*log((4-(pi+x))/(4-pi)),((pi-x)/4)*(log((pi-x)/(pi)))+(1-((pi-x)/4))*log((4-(pi-x))/(4-pi)))
```

```
plot(x,I)
```

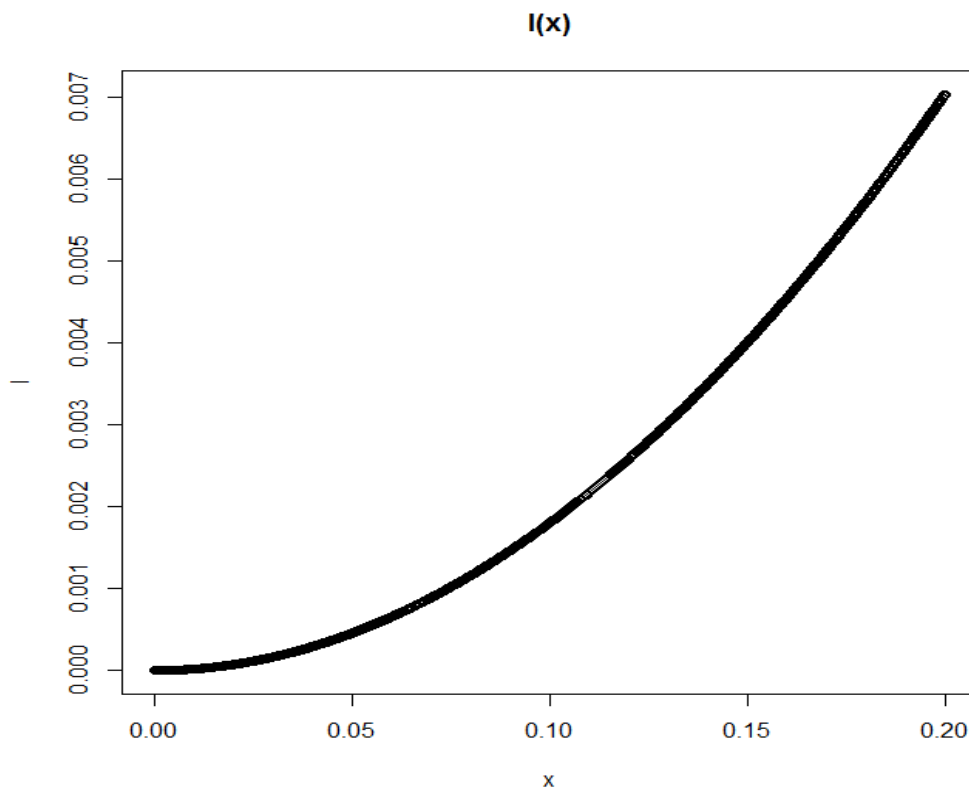
```
par(new=TRUE)
```

```
x<-c(0.02,0.04,0.06,0.08,0.10,0.12,0.14)
```

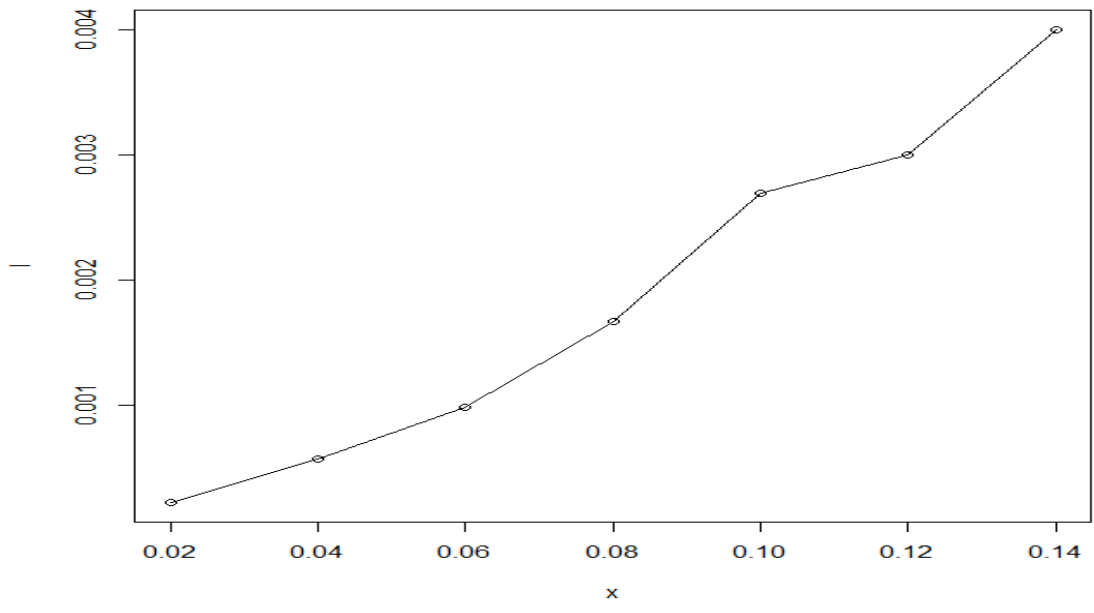
```
I<-c(0.00022,0.00057,0.00098,0.00167,0.00269,0.0030,0.0040)
```

```
lines(x,I2)
```

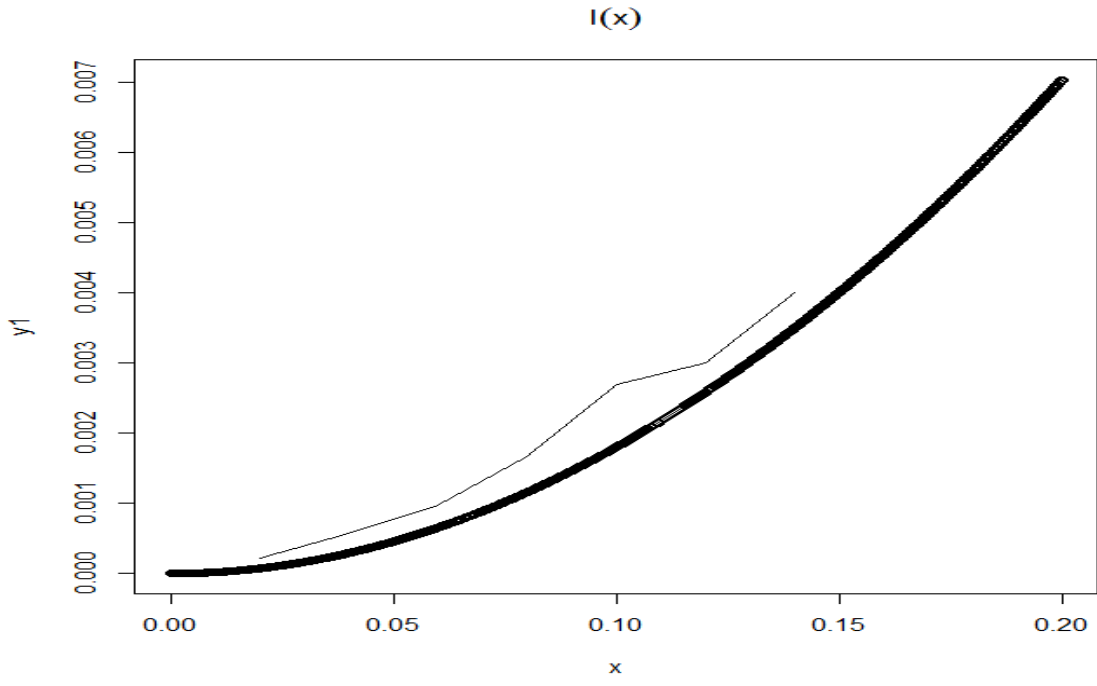
plot(x,I) είναι η γραφική παράσταση της θεωρητικής συνάρτησης ταχύτητας



lines(x,I2) η γραφική παράσταση για τις εκτιμήσεις που κάναμε με τη μέθοδο Monte Carlo



Και εδώ είναι η συγκριτική γραφική παράσταση της μεθόδου Monte Carlo με τη θεωρητική συνάρτηση ταχύτητας



Τέλος θα δείξουμε το MC σφάλμα για την εκτίμηση του $I(x)$. Θα κάνουμε 100 εκτιμήσεις της ίδιας ποσότητας θα υπολογίσουμε τη διασπορά αυτών των αποτελεσμάτων και θα κάνουμε και το ιστόγραμμα τους όπως ακριβώς κάναμε και στην περίπτωση του π .

Ο κώδικας είναι ο παρακάτω

```
b=function(O,n,M,t)
{
dian1=rep(1,O)
for(z in 1:O)
{
dian=rep(1,n)
for(i in 1:n)
{
k=0
for (j in 1:M) {
x <- 2*runif(1)-1 ; y <- 2*runif(1)-1;

if (x^2+y^2<1)
{
k=k+1
}
dian[i]=4*k/M
}
}
er=rep(1,n)
for(i in 1:n)
{
er[i]=abs(dian[i]-pi)
}
v=0
for(j in 1:n)
{
```

```

if(er[j]>t)
{
v=v+1
}
}
dian1[z]=-(1/n)*log(v/n)
}
return(dian1)
}

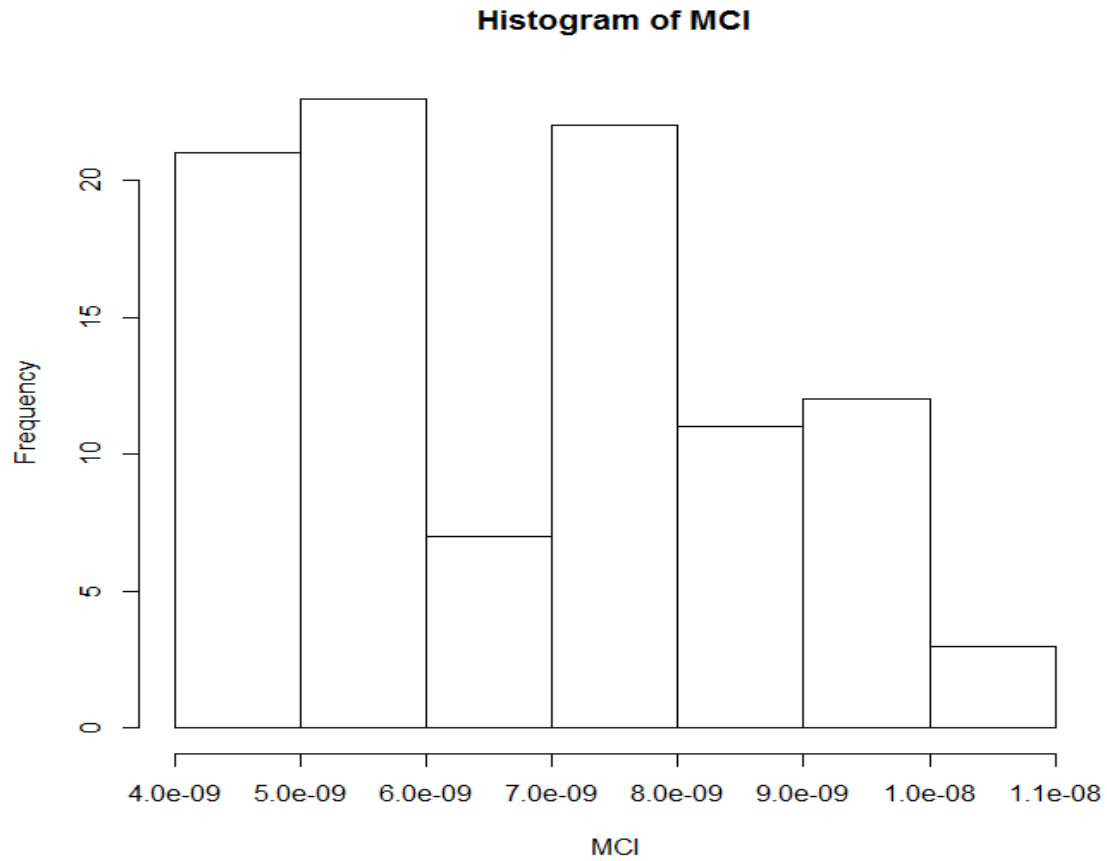
```

```

mc=function(r)
{
dian1=rep(1,r)
for(e in 1:r)
{
b(1000,1000)
x<-b(1000,1000)
j<-0
m<-rep(NA,40)
l<-rep(NA,25)
for(i in 1:40)
{
l<-x[j:(0+i*25)]
j<-j+25
m[i]<-mean(l)
}
dia1[e]=var(m)
}
return(dian1)
}

```

```
MCI=mc(100)
var(MCI)
[1] 3.571023e-18
hist(MCI)
```



Το παραπάνω διάγραμμα είναι το διάγραμμα συχνότητων για τις τιμές του MCI που πέρνουμε από το πρόγραμμα που φτιάξαμε παραπάνω.

Τελικό Συμπέρασμα

Από όλα τα γραφήματα βλέπουμε ότι τα σημεία σχηματίζουν περίπου μια γραφική παράσταση κυρτής συνάρτησης κάτι που περιμένουμε σίγουρα αφού ο θεωρητικός μας τύπος αυτό λέει, βέβαια στο τελευταίο γράφημα για την εκτίμηση του $I(x)$ που πέρνουμε δεν είναι καθόλου καλή και μας δείχνει ότι η μέθοδος Monte Carlo δεν δουλεύει τόσο καλά στο προγράμμα μας. Το συμπέρασμα λοιπόν είναι ότι η μέθοδος Monte Carlo δεν μας δίνει και τις καλύτερες εκτιμήσεις για την ταχύτητα σύγκλισης και αυτό το γεγονός γίνεται όλο και πιο φανερό όσο προχωράμε σε ενδεχόμενα πάρα πολύ σπάνια δηλαδή που έχουν πάρα πολύ μικρή πιθανότητα να συμβούν. Για παράδειγμα μετά από $x=0.14$ η μέθοδος δεν μπορεί να υπολογίσει την πιθανότητα του ενδεχομένου και την βγάζει 0, με αποτέλεσμα η $I(x)$ να πηγαίνει στο άπειρο. Οπότε θα χρησιμοποιήσουμε τη τεχνική της δειγματοληψίας κατά σπουδαιότητα (Importance Sampling)

3.2 Τεχνική της δειγματοληψίας κατα σπουδαιότητας (Μέθοδος Importance Sampling)

Όπως είδαμε και προηγουμένως τα προβλήματα της μεθόδου Monte Carlo είναι αρκετά οπότε θα χρησιμοποιήσουμε τώρα μια άλλη μέθοδο που έχει αναπτυχθεί τη μέθοδο Importance Sampling.

Η μέθοδος Importance Sampling είναι μια εκτίμηση όπως και η μέθοδος Monte Carlo σε ένα δείγμα x_1, x_2, \dots, x_n που από μια κατανομή πάμε σε μια άλλη.

f η αρχική κατανομή g η καινούργια

$$E_f[h(X)] \approx \frac{1}{v} \sum_{i=1}^v \frac{(f(x_i)h(x_i))}{(g(x_i))} \quad (1)$$

$$E_f[h(X)] = \int_x h(x) \frac{(f(x))}{(g(x))} g(x) dx = E_g\left[h(X) \frac{(f(X))}{(g(X))}\right]$$

Η εκτιμήτρια που δίνεται από την (1) συγκλίνει για τον ίδιο ακριβώς λόγο που συγκλίνει η εκτιμήτρια Monte Carlo.

Ουσιαστικά αυτό που θέλουμε να κάνουμε στη μέθοδο Importance Sampling είναι να αλλάξουμε την πιθανότητα μετάβασης $\tilde{p}(x, y)$ για παράδειγμα αν έχουμε ότι

$$E[f(x)] \rightarrow 0$$

και πρέπει να υπολογίσουμε τη πιθανότητα

$$P\left[\frac{(f(x_1), \dots, f(x_n))}{n} > 0.2\right] = ;$$

Αρχικά με τη βοήθεια του Κεντρικού Οριακού θεωρήματος θα κάνουμε αυτή την μετατροπή

$$\frac{(X_1 + X_2 + \dots + X_n - np)}{n} \rightarrow N\left[0, \frac{\sigma^2}{n}\right] \text{ και το νέο } X_i = 1$$

αν $\frac{(f(x_1), \dots, f(x_n))}{n} > 0.2$ και $X_i = 0$ διαφορετικά

Συνεπώς εμείς τώρα έχουμε φτιάξει μία μεταβλητή για την οποία ισχύει ότι

$$\frac{(f(x_1), \dots, f(x_n))}{n} \rightarrow 0.2 \text{ άρα } \int f d\hat{\pi} = 0.2$$

Ακόμα ισχύει ότι $P[A] = \int_A dP = \int_A f d\tilde{P}$

Η καινούργια p υπολογίζεται ότι είναι $\tilde{p}(x, y) = \frac{e^{\lambda(f(x) - f(y))}}{z_\lambda} p(x, y)$

Ακόμα $\tilde{P}[X_1 = x_1, \dots, X_n = x_n] = \tilde{\pi}_0(x_1) \tilde{p}(x_1, x_2) \dots \tilde{p}(x_{n-1}, x_n)$

$$P[A] = \int_A dP = \int_A \pi(x_1, \dots, x_n) f(x_1, \dots, x_n) dP(x_1, \dots, x_n)$$

$$\frac{(d \tilde{p})}{(d p)} = \frac{\prod_{i=1}^{n-1} (\tilde{p}(x_i, x_{i+1}))}{(p(x_i, x_{i+1}))}$$

Τώρα θα κάνουμε ένα παράδειγμα Importance sampling με χρήση του στατιστικού περιβάλλοντος της R

Θα υπολογίσουμε την πιθανότητα $I = P(Z > 3)$ ενώ το $Z \sim N(0,1)$

$$I = \int_3^{\infty} f(x) dx \approx 0.0013$$

$$I = \int_{-\infty}^{\infty} h(x) f(x) dx \quad \text{για } h(x) = 0 \text{ } x \leq 3 \text{ αλλιώς } h(x) = 1$$

Importance sampling

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N h(x_i) \frac{f(x_i)}{g(x_i)} \quad \text{με την } f(x) \text{ να την ξέρουμε και πρέπει να επιλέξουμε την } g.$$

Έστω $g(x) \sim N(4,1)$

$$b(x) = \frac{f(x)}{g(x)} = e^{8-4x}$$

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N h(x_i) b(x_i)$$

Ο κώδικας είναι ο εξής

```
for (j in 1:100){
```

```
N = 100;
```

```
h = rnorm (N,1) + 4;
```

```
for (i in 1:N){
```

```
h >= 3;
```

```
b = exp(8-4*h);
```

```
w = h*b;
```

```
}
```

```

I = sum(w)/N;
}
MEAN = mean(I)
VAR = var(I)
MEAN
[1] 0.001393818

```

πολύ κοντά στο πραγματικό αποτέλεσμα που είναι 0.0013 εάν σε αυτό το πρόβλημα χρησιμοποιούσαμε Monte Carlo τότε θα χρησιμοποιούσαμε τον παρακάτω κώδικα

```

for (i in 1:100){
a= sum(rnorm(100,1)>=3)/100;
}
meanMC = mean(a)
varMC = var(a)
> meanMC
[1] 0.03

```

πολύ χειρότερη δηλαδή από την Importance sampling.

Μέθοδος Importance Sampling για τον υπολογισμό του π

Τώρα θα προσπαθήσουμε να χρησιμοποιήσουμε τη μέθοδο Importance sampling και στο παράδειγμα του 4.1. Αρχικά θα πάρουμε μια πιθανότητα που είναι πολύ σπάνια για να δούμε πως συμπεριφέρεται δηλαδή την πιθανότητα η εκτιμήτρια μας για το π να συγκλίνει στο $\pi+x$ δηλαδή. $\hat{\pi} \rightarrow \pi + x$ Αυτή η πιθανότητα με τη μέθοδο Monte Carlo όσες επαναλήψεις και να κάναμε μας έβγαине 0 άρα θα προσπαθήσουμε να την εκτιμήσουμε με την μέθοδο Importance sampling και να δούμε πώς συμπεριφέρεται. Αρχικά θα αλλάξουμε την κατανομή που θα ακολουθεί η εκτιμητριά μας αφού θα πρέπει αντι για να συγκλίνει στο π να συγκλίνει στο $\pi+x$ για να γίνει αυτό πρέπει να ακολουθηθεί η παρακάτω διαδικασία

Αρχικά για εμάς ο χώρος καταστασεών μας είναι ο εξής $\Omega = \{0,1\}^n$ και επίσης ισχύει οτι $x \in A \approx 1_A X(\omega)$. Για ένα διακριτό χώρο πιθανότητας όπως ο δικός μας έχουμε ότι

$$P[w] = \left(\frac{\pi}{4}\right)^{\sum w_i} \left(1 - \left(\frac{\pi}{4}\right)\right)^{(n - \sum w_i)} \quad \text{για χάρην ευκολίας έστω}$$

$|w| = \sum w_i$ άρα

$$P[w] = \left(\frac{\pi}{4}\right)^{|w|} \left(1 - \left(\frac{\pi}{4}\right)^{(n-|w|)}\right)$$

τόρα για το ενδεχόμενο x έχουμε $\{x \in A\} = \{|\hat{\pi} - \pi| \geq x\} = \{|\left(4 \sum \frac{w_i}{n}\right) - \pi| \geq x\} = \left\{\left|\left(\frac{|w|}{n}\right) - \left(\frac{\pi}{4}\right)\right| \geq \left(\frac{x}{4}\right)\right\}$

Οπότε για το νέο μέτρο πιθανότητας θα έχουμε ότι

$$Q[w] = \left(\frac{\pi+x}{4}\right)^{|w|} \left(1 - \left(\frac{\pi+x}{4}\right)^{(n-|w|)}\right)$$

Επειδή είμαστε σε διακριτό χώρο έχουμε ότι το $\frac{dP}{dQ}$ είναι απλά το πηλίκο των μέτρων πιθανότητας

για το ενδεχόμενο x αντιστοίχως δηλαδή $\frac{dP}{dQ} = \left(\frac{\pi}{\pi+x}\right)^{|w|} \left(\frac{4-\pi}{4-\pi-x}\right)^{(n-|w|)}$ αυτό είναι το

ενδεχόμενο για $P[|\hat{\pi}(\omega) - \pi| > x] = P[\hat{\pi}(\omega) > \pi + x]$ και θα πρέπει να πάρουμε αντίστοιχα και το

ενδεχόμενο για $P[|\hat{\pi}(\omega) - \pi| > x] = P[\hat{\pi}(\omega) < \pi - x]$ όπου εκεί το αντίστοιχο $\frac{dP}{dQ}$ είναι

$$\frac{dP}{dQ} = \left(\frac{\pi}{\pi-x}\right)^{|w|} \left(\frac{4-\pi}{4-\pi+x}\right)^{(n-|w|)}$$

p=function(n,M,t)

{

v=0

for(i in 1:n)

{

k=0

for (j in 1:M) {

x <- 2*runif(1)-1 ; y <- 2*runif(1)-1;

if ((x^2+y^2<1) || ((-1 < x)&& (x<(-1+sqrt(t)/2)) && (-1 < y)&& (y<(-1+sqrt(t)/2))) || ((-1 < x) && (x<(-1+sqrt(t)/2)) && ((1-sqrt(t)/2) < y)&&(y < 1)) || (((1-sqrt(t)/2) < x) && (x<1) && ((1-sqrt(t)/2) < y)&& (y < 1)) || (((1-sqrt(t)/2) < x) && (x < 1) && (-1/2 < y) && (y<(-1+sqrt(t)/2)))))

{

k=k+1

}

}

if(k>((pi+t)*M)/4)

```

{
v=v+((pi/(pi+t))^k)*((4-pi)/(4-pi-t)^(M-k))
}
l=0
for (a in 1:M) {
x <- 2*runif(1)-1 ; y <- 2*runif(1)-1;
if((t/pi<x^2+y^2) && (x^2+y^2<1))
{
l=l+1
}
}
if(l<((pi-t)*M)/4)
{
v=v+((pi/(pi-t))^l)*((4-pi)/(4-pi+t)^(M-l))

}
}
return (v/n)
}

```

Θα παραθέσουμε και ένα γράφημα που να δείχνει την μέθοδο Importance Sampling και τα αποτελέσματα της σε σχέση με τη μέθοδο Monte Carlo αλλά και την θεωρητική γραφική παράσταση.

```

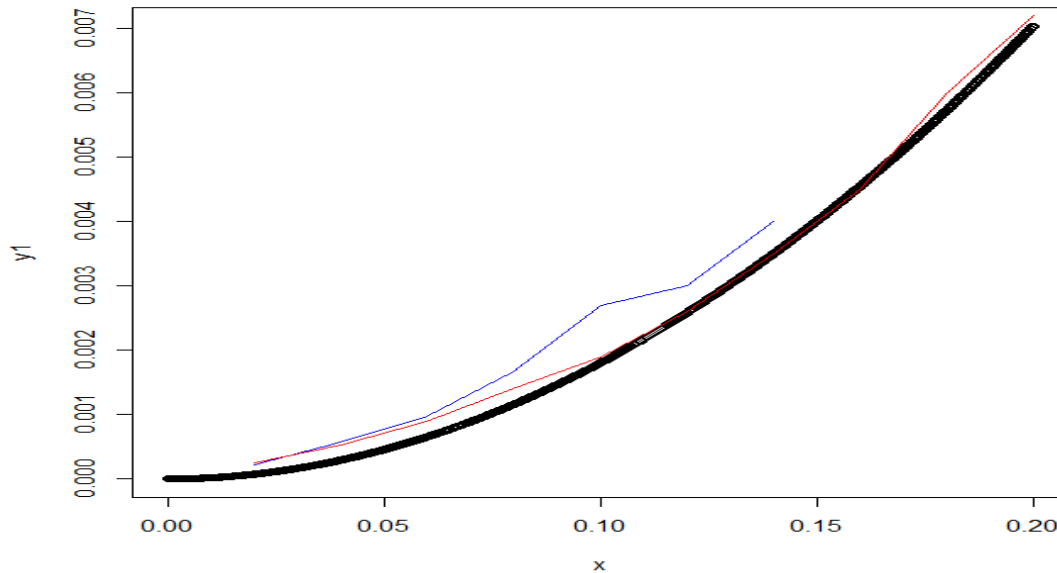
x<-seq(0,0.2,0.0001)
y1<-ifelse(x<0,((pi+x)/4)*(log((pi+x)/(pi)))+(1-((pi+x)/4))*log((4-(pi+x))/(4-pi)),((pi-x)/4)*(log((pi-x)/(pi)))+(1-((pi-x)/4))*log((4-(pi-x))/(4-pi)))
plot(x,y1)
par(new=TRUE)
x<-c(0.02,0.04,0.06,0.08,0.10,0.12,0.14)
I<-c(0.00022,0.00057,0.00098,0.00167,0.00269,0.0030,0.0040)
lines(x,I,col="blue")
par(new=TRUE)

```

```
x<-c(0.02,0.04,0.06,0.08,0.10,0.12,0.14,0.16,0.18,0.20)
```

```
I2<-c(0.00025,0.00052,0.0009,0.0014,0.0019,0.0026,0.0035,0.0045,0.0060,0.0072)
```

```
lines(x,I2,col="red")
```



Στο γράφημα με μαύρο χρωμα έχουμε την θεωρητική τιμή της $I(x)$ με μπλε χρώμα τη μέθοδο Monte Carlo και με κόκκινο χρώμα την Importance Sampling . Στο παραπάνω γράφημα φαίνεται οτι η μέθοδο Importance Sampling είναι πολύ καλύτερη αφού είναι κυρτή, είναι πιο κοντά στις θεωρητικές τιμές της $I(x)$ και μάλιστα το γραφημά της μοιάζει εξαιρετικά με το γράφημα της θεωρητικής $I(x)$. Ακόμα η μέθοδος Monte Carlo αποτυγχάνει να περιγράψει την συνάρτηση ταχύτητας για ποιά σπάνια ενδεχόμενα ενώ η μέθοδος Importance Sampling καταφέρνει να την περιγράψει και μάλιστα με πολύ μεγάλη ακρίβεια.

Κεφάλαιο 4

Θεώρημα Cramer για Μαρκοβιανές Αλυσίδες

4.1 Μαρκοβιανές Αλυσίδες

Μαρκοβιανές Αλυσίδες

Είναι ένα μαθηματικό σύστημα που μεταβάλλεται από μια κατάσταση σε μια άλλη, ανάμεσα σε ένα πεπερασμένο αριθμό καταστάσεων. Είναι μια τυχαία διαδικασία που δε διατηρεί μνήμη για τις προηγούμενες μεταβολές: Η επόμενη κατάσταση εξαρτάται μόνο από την τωρινή κατάσταση και σε καμιά περίπτωση από αυτές που προηγήθηκαν. Αυτή η συγκεκριμένη ιδιότητα ονομάζεται μαρκοβιανή. Οι Μαρκοβιανές Αλυσίδες έχουν πολλές εφαρμογές ως στατιστικά μοντέλα καθημερινών διαδικασιών .

Όταν μιλάμε για Μαρκοβιανές Αλυσίδες συνήθως χρησιμοποιούμε διακριτό χρόνο αν και σε κάποιες λιγότερες περιπτώσεις την βρίσκουμε και σε συνεχές χώρο καταστάσεων.

Το πιο γνωστό παράδειγμα Μαρκοβιανής Αλυσίδας είναι ίσως ο τυχαίος περίπατος όπου ένας άνθρωπος κάνει μια τυχαία διαδρομή ξεκινώντας από την θέση 0 και με ίδια πιθανότητα $1/2$ κάθε φορά μπορεί να κινηθεί είτε μπροστά είτε πίσω. Αυτές οι πιθανότητες λοιπόν για το αν θα πάει $+1$ ή -1 είναι ανεξάρτητες από το που βρισκόταν νωρίτερα ο άνθρωπος. Μια σειρά ανεξάρτητων γεγονότων (για παράδειγμα, μια σειρά από στριψίματα νομίσματος) ικανοποιεί τον επίσημο ορισμό της Μαρκοβιανής αλυσίδας. Παρ' όλα αυτά, η θεωρία συνήθως εφαρμόζεται μόνο όταν η πιθανότητα κατανομής του επομένου βήματος εξαρτάται αμελητέα από το τι συναίβει πριν.

Ορισμός Μαρκοβιανής Αλυσίδας

Μια Μαρκοβιανή Αλυσίδα αποτελεί μια ακολουθία τυχαίων μεταβλητών με την Μαρκοβιανή ιδιότητα, δηλαδή με δεδομένη την τωρινή κατάσταση οι παρελθοντικές και μελλοντικές καταστάσεις είναι ανεξάρτητες.

$$P(X_{n+1}=x|X_1=x_1, X_2=x_2, \dots, X_n=x_n) = P(X_{n+1}=x|X_n=x_n)$$

Οι Μαρκοβιανες Αλυσιδες συχνα περιγραφονται από ένα κατευθυνόμενο γράφημα που οι ακρες του επιγραφουν τις πιθανοτητες μετάβασης από τη μια κατάσταση στις άλλες και ολος αυτός ο χώρος ονομάζεται χώρος καταστάσεων.

Η κάθε κατάσταση στην οποία μπορεί να βρεθεί μια Μαρκοβιανή Αλυσίδα μπορεί να έχει περίοδο τότε αυτή η κατάσταση ονομάζεται περιοδική. Αν οι επιστροφές σε αυτή την κατάσταση είναι μηδέν τότε αυτή η κατάσταση ονομάζεται απεριοδική. Μια Μαρκοβιανή Αλυσίδα είναι απεριοδική αν κάθε κατάσταση της είναι απεριοδική. Μια αμείωτη Μαρκοβιανη Αλυσίδα χρειάζεται μόνο μια απεριοδική κατάσταση ώστε να πούμε ότι όλες οι καταστάσεις είναι απεριοδικές.

Όταν λοιπόν μια κατάσταση δεν είναι περιοδική και οπως πριν την ονομάσαμε απεριοδική θα την ονομάζουμε μεταβατική κάτι που είναι λογικό γιατί αυτό σημαίνει οτι κάποια στιγμή μέσα στην Μαρκοβιανή Αλυσίδα μπορούμε να μεταβούμε σε αυτή την κατάσταση. Μια κατάσταση i λέμε ότι είναι μεταβατική όταν, αν ξεκινήσουμε από την κατάσταση i , υπάρχει μια μη-μηδενική πιθανότητα να μη γυρίσουμε ποτέ στην κατάσταση. Τέλος υπάρχει και άλλο ενος είδος κατάστασης η απορροφητική. Μια κατάσταση i ονομάζεται απορροφητική αν είναι αδύνατο να φύγουμε από αυτή την κατάσταση. Αν κάθε κατάσταση μπορεί να φτάσει σε μια απορροφητική κατάσταση, τότε λέμε ότι η αλυσίδα μας είναι Απορροφητική Μαρκοβιανη Αλυσίδα.

4.2 Εφαρμογες Μαρκοβιανών Αλυσίδων στις Σύγχρονες Επιστήμες

Οι Μαρκοβιανές Αλυσίδες εμφανίζονται σε πάρα πολλές διαφορετικές επιστήμες όπως

Χημεία

Η χημεία είναι συχνά ένας τομέας όπου οι Μαρκοβιανές αλυσίδες και οι Μαρκοβιανές καταστάσεις συνεχούς χρόνου είναι ιδιαίτερα χρήσιμες επειδή αυτά τα απλά φυσικά συστήματα τείνουν να ικανοποιούν αρκετά καλά τη Μαρκοβιανή ιδιότητα. Το κλασικό μοντέλο της ενζυμικής δραστηριότητας, η κινητική Michaelis-Menten, μπορεί να θεωρηθεί ως μια Μαρκοβιανή αλυσίδα, όπου σε κάθε βήμα η αντίδραση προχωράει προς κάποια κατεύθυνση. Ενώ η κινητική Michaelis-Menten είναι αρκετά απλή, πολύ περισσότερο πολύπλοκα δίκτυα αντιδράσεων μπορούν επίσης να μοντελοποιηθούν με Μαρκοβιανές αλυσίδες. Ένας αλγόριθμος που βασίζεται σε Μαρκοβιανή αλυσίδα χρησιμοποιήθηκε επίσης για να εστιάσει τη βασιζόμενη σε θραύσματα ανάπτυξη χημικών *in silico* προς μια επιθυμητή τάξη ενώσεων όπως φάρμακα ή φυσικά προϊόντα. Καθώς ένα στοιχείο αναπτύσσεται, ένα θραύσμα επιλέγεται από το εν τη γενέσει στοιχείο ως η "παρούσα" κατάσταση. Δεν έχει επίγνωση του παρελθόντος του (δε γνωρίζει τι βρίσκεται ήδη δεσμευμένο σε αυτό). Στη συνέχεια, μεταβαίνει στην επόμενη κατάσταση όταν ένα θραύσμα προσκολλάται σε αυτό. Οι πιθανότητες μετάβασης διευθύνονται από βάσεις δεδομένων αυθεντικών τάξεων ενώσεων.

Επίσης, η ανάπτυξη (και σύνθεση) συμπολυμερών μπορεί να μοντελοποιηθεί χρησιμοποιώντας Μαρκοβιανές αλυσίδες. Με βάση το κλάσμα αντίδρασης των μονομερών που αποτελούν την αναπτυσσόμενη πολυμερική αλυσίδα, η σύνθεση της αλυσίδας μπορεί να υπολογισθεί (πχ αν τα μονομερή τείνουν να προστίθεται με εναλασσόμενο τρόπο ή με μακριές σειρές του ίδιου μονομερούς).

Παρόμοια, έχει προταθεί ότι η κρυσταλλοποίηση και ανάπτυξη κάποιων επιταξιακών υπερπλεγμάτων οξειδίων μπορούν να περιγραφούν με ακρίβεια με Μαρκοβιανές αλυσίδες.

Φυσική

Τα Μαρκοβιανά συστήματα εμφανίζονται εκτενώς στη θερμοδυναμική και στη στατική μηχανική, όποτε πιθανότητες χρησιμοποιούνται για να αναπαραστήσουν άγνωστες ή μη μοντελοποιημένες λεπτομέρειες του συστήματος, αν μπορεί να υποθεθεί ότι η δυναμική δεν μεταβάλλεται με το χρόνο και ότι καμία σχετική ιστορία που να μην περιλαμβάνεται ήδη στην περιγραφή κατάστασης δε χρειάζεται να θεωρηθεί.

Επιστήμη πληροφοριών

Οι Μαρκοβιανές αλυσίδες χρησιμοποιούνται ευρέως στην επεξεργασία πληροφοριών. Η διάσημη εργασία του Claude Shannon από το 1948 "Μια μαθηματική θεωρία επικοινωνίας", η οποία σε ένα βήμα δημιούργησε το πεδίο της θεωρίας πληροφοριών, ξεκινάει παρουσιάζοντας την έννοια της εντροπίας μέσω Μαρκοβιανής μοντελοποίησης της Αγγλικής γλώσσας. Τέτοια ιδανικά μοντέλα μπορούν να συλλάβουν πολλές από τις στατιστικές ανωμαλίες των συστημάτων. Ακόμη και χωρίς περιγραφή της πλήρους δομής του συστήματος, τέτοια μοντέλα σημάτων μπορούν να κάνουν δυνατή την πολύ αποτελεσματική συμπίεση δεδομένων μέσω τεχνικών κωδικοποίησης εντροπίας όπως η αριθμητική κωδικοποίηση. Επίσης επιτρέπουν αποτελεσματική εκτίμηση καταστάσεων και αναγνώριση προτύπων. Οι Μαρκοβιανές αλυσίδες παίζουν επίσης σημαντικό ρόλο στην ενισχυτική μάθηση.

Οι Μαρκοβιανές αλυσίδες είναι επιπλέον η βάση για Κρυμμένα Μαρκοβιανά μοντέλα, που αποτελούν σημαντικό εργαλείο σε ποικίλα πεδία όπως τηλεφωνικά δίκτυα (που χρησιμοποιούν τον αλγόριθμο Viterbi για διόρθωση σφαλμάτων), αναγνώριση λόγου και βιοπληροφορική.

Εφαρμογές Διαδικτύου

Η κατάταξη μιας ιστοσελίδας με βάση τον αλγόριθμο PageRank, όπως χρησιμοποιείται από τη Google, ορίζεται από μια Μαρκοβιανή αλυσίδα. Είναι η πιθανότητα να βρισकेσαι στη σελίδα i στη στατική κατανομή της ακόλουθης Μαρκοβιανής αλυσίδας για όλες τις (γνωστές) ιστοσελίδες. Αν N είναι ο αριθμός των γνωστών ιστοσελίδων και μια σελίδα i έχει k_i συνδέσμους σε αυτή, τότε έχει

πιθανότητα μετάβασης $\left(\frac{a}{k_i} + \frac{(1-a)}{N}\right)$ για όλες της σελίδες που συνδέονται με αυτήν και

$\frac{(1-a)}{N}$ για όλες τις σελίδες που δε συνδέονται με αυτήν. Η παράμετρος a θεωρείται περίπου 0,85.

Μαρκοβιανά μοντέλα έχουν επίσης χρησιμοποιηθεί για την ανάλυση της συμπεριφοράς περιήγησης των χρηστών στο διαδίκτυο. Η μετάβαση διαδικτυακών συνδέσμων ενός χρήστη σε μια συγκεκριμένη ιστοσελίδα μπορεί να μοντελοποιηθεί χρησιμοποιώντας Μαρκοβιανά μοντέλα πρώτης ή δεύτερης τάξης και μπορεί να χρησιμοποιηθεί για να προβλέψει μελλοντική περιήγηση και να εξατομικεύσει την ιστοσελίδα για ένα συγκεκριμένο χρήστη.

Οικονομία και Οικονομικά

Οι Μαρκοβιανές αλυσίδες χρησιμοποιούνται στα Χρηματοοικονομικά για να μοντελοποιήσουν μια ποικιλία διαφορετικών φαινομένων, συμπεριλαμβανομένων τιμών κεφαλαίων και πτώσεις αγορών. Το πρώτο χρηματοοικονομικό μοντέλο που χρησιμοποίησε Μαρκοβιανή αλυσίδα ήταν του Prasad και συνεργατών το 1974. Ένα άλλο ήταν το μοντέλο αλλαγής καθεστώτος του James D. Hamilton (1989), στο οποίο μια Μαρκοβιανή αλυσίδα χρησιμοποιείται για να μοντελοποιήσει αλλαγές μεταξύ περιόδων υψηλής μεταβλητότητας και χαμηλής μεταβλητότητας επιστροφής κεφαλαίων. Ένα πιο πρόσφατο παράδειγμα είναι το Πολυκλασματικό Μαρκοβιανό μοντέλο Αλλαγής τιμολόγησης κεφαλαίων, που βασίζεται στην ευκολία των προηγούμενων μοντέλων αλλαγής καθεστώτων. Χρησιμοποιεί μια τυχαία μακριά Μαρκοβιανή αλυσίδα για να οδηγήσει τα επίπεδα μεταβλητότητας επιστροφών κεφαλαίου. Η δυναμική μακροοικονομία ευρέως χρησιμοποιεί Μαρκοβιανές αλυσίδες. Ένα παράδειγμα είναι η χρήση τους για την εξωγενή μοντελοποίηση τιμών δικαιουσύνης (μετοχών) σε μια ρύθμιση γενικής ισορροπίας.

Κοινωνικές Επιστήμες

Οι Μαρκοβιανές αλυσίδες χρησιμοποιούνται γενικά στην περιγραφή επιχειρημάτων που εξαρτώνται από τη διαδρομή, όπου οι παρούσες δομικές διαμορφώσεις διέπουν τα μελλοντικά αποτελέσματα. Ένα παράδειγμα είναι η αναδιατύπωση της ιδέας, αρχικά λόγω του "Κεφαλαίου" του Καρλ Μαρξ, συνδέοντας την οικονομική ανάπτυξη με την άνοδο του καπιταλισμού. Στην παρούσα έρευνα, είναι συνηθισμένη η χρήση Μαρκοβιανής αλυσίδας για να μοντελοποιήσει πώς όταν μια χώρα φτάσει ένα συγκεκριμένο επίπεδο οικονομικής ανάπτυξης, η διαμόρφωση δομικών παραγόντων, όπως το μέγεθος της εμπορικής αστικής τάξης, η αναλογία της αστικής προς την αγροτική κατοικία, ο ρυθμός πολιτικής κινητοποίησης κτλ θα οδηγήσει σε υψηλότερη πιθανότητα μετάβασης από απολυταρχικό σε δημοκρατικό καθεστώς.

Βιολογία

Οι Μαρκοβιανές αλυσίδες έχουν επίσης πολλές εφαρμογές σε βιολογικά μοντέλα, κυρίως πληθυσμικές διαδικασίες, με χρησιμότητα στη μοντελοποίηση διαδικασιών που είναι (τουλάχιστον) ανάλογες με βιολογικούς πληθυσμούς. Η μήτρα Leslie είναι ένα παράδειγμα, παρόλο που μερικές από τις καταχωρίσεις δεν είναι πιθανότητες (μπορεί να είναι μεγαλύτερες από 1). Ένα άλλο παράδειγμα είναι η μοντελοποίηση του κυτταρικού σχήματος σε διαιρούμενα φύλλα επιθηλιακών κυττάρων. Ένα ακόμη παράδειγμα είναι η κατάσταση ιοντικών καναλιών στις κυτταρικές μεμβράνες. Οι Μαρκοβιανές αλυσίδες χρησιμοποιούνται επίσης σε προσομοιώσεις εγκεφαλικής λειτουργίας, όπως η προσομοίωση του νεοφλοιού των θηλαστικών.

Γενετική

Οι Μαρκοβιανές αλυσίδες έχουν χρησιμοποιηθεί στην πληθυσμιακή γενετική για να περιγράψουν την αλλαγή στις συχνότητες των γονιδίων σε μικρούς πληθυσμούς υπό την επιρροή γενετικής παρέκκλισης, για παράδειγμα στη μέθοδο εξίσωσης διάχυσης .

4.3 Εργοδικό Θεώρημα για Μαρκοβιανές Αλυσίδες

Εργοδικό Θεώρημα

Μια κατάσταση λέμε ότι είναι εργοδική αν είναι απεριοδική και θετικά επαναλαμβανομένη. Με άλλα λόγια, μια κατάσταση i είναι εργοδική αν είναι επαναληπτική, έχει περίοδο ίση με 1 και έχει πεπερασμένο μέσο χρόνο επανάληψης. Αν όλες οι καταστάσεις σε μια αμείωτη Μαρκοβιανή Αλυσίδα είναι εργοδικές τότε λέμε ότι η Αλυσίδα είναι Εργοδική.

Έχει αποδειχτεί ότι μια πεπερασμένη αμείωτη Μαρκοβιανή Αλυσίδα είναι εργοδική αν έχει μια απεριοδική κατάσταση. Ένα Μαρκοβιανό μοντέλο έχει την εργοδική ιδιότητα αν υπάρχει ένας πεπερασμένος αριθμός N τέτοιος ώστε κάθε κατάσταση μπορεί να φτάσει σε κάθε άλλη κατάσταση σε ακριβώς N βήματα. Σε περίπτωση που έχουμε ένα πλήρως συνδεδεμένο πίνακα μετάβασης όπου όλες οι μεταβάσεις έχουν μη-μηδενική πιθανότητα, αυτή η προϋπόθεση ικανοποιείται για $N=1$. Ένα μοντέλο με παραπάνω από μια καταστάσεις και με μόνο μια πιθανή μετάβαση σε κάθε κατάσταση δε μπορεί να είναι εργοδικό

Εργοδικό Θεώρημα για αλυσίδες Markov : Έστω $X = (X_n)_{n \in \mathbf{N}}$ μη αναγώγιμη αλυσίδα Markov στον αριθμήσιμο χώρο φάσης $M = \{x_i | i \in \mathbf{N}\}$. Αν υπάρχει αναλλοίωτη κατανομή μ στον M για τη X , τότε

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow{n \rightarrow \infty} \int_M f d\mu =: \bar{f} \quad \text{σχεδόν βεβαίως για κάθε φραγμένη } f : M \rightarrow \mathbb{R}$$

$$P\left[\frac{1}{n} \sum_{i=1}^n f(X_i) \rightarrow E_\pi[f(X)]\right] \xrightarrow{n \rightarrow +\infty} 1$$

Βλέπουμε ότι οι εκτιμήσεις για την ταχύτητα σύγκλισης των πιθανοτήτων σπάνιων ενδεχομένων στο 0 μας δίνουν εκτιμήσεις για το πόσο χρόνο πρέπει να αφήσουμε μια ανέλιξη Markov να τρέξει ώστε να πλησιάσει την κατάσταση ισορροπίας της μ .

4.4 Θεώρημα (Perron-Frobenius)

Θεώρημα (Perron-Frobenius)

Παίρνουμε $B = (B(i, j))_{i, j=1}^{|\Sigma|}$ που είναι μη υποβιβάσιμος πίνακας. Όπου ο B έχει μια ιδιοτιμή p καλούμενη ως ιδιοτιμή Perron-Frobenius για την οποία ισχύει ότι

(α) $p > 0$ είναι πραγματική

(β) για κάθε ιδιοτιμή λ του πίνακα B ισχύει ότι, $|\lambda| \leq p$

(γ) Υπάρχουν αριστερά και δεξιά ιδιοδιανύσματα που αντιστοιχούν στην ιδιοτιμή p και έχουν αυστηρά θετικές συντεταγμένες.

(δ) Υπάρχουν αριστερά και δεξιά ιδιοδιανύσματα που αντιστοιχούν στην ιδιοτιμή p που είναι μοναδικά

(ε) Για κάθε $i \in \Sigma$ και κάθε $\varphi = (\varphi_1, \dots, \varphi_{|\Sigma|})$ έτσι ώστε $\forall j \varphi_j > 0$

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \log \left[\sum_{j=1}^{|\Sigma|} B^n(i, j) \varphi_j \right] \right)$$

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \log \left[\sum_{j=1}^{|\Sigma|} B^n(i, j) \varphi_j \right] \right) = \lim_{n \rightarrow +\infty} \left(\frac{1}{n} \log \left[\sum_{j=1}^{|\Sigma|} \varphi_j B^n(j, i) \right] \right) = \log p$$

Απόδειξη

Τα πρώτα τέσσερα ερωτήματα δεν μας αφορούν τόσο για την απόδειξη του Θεωρήματος Cramer για Μαρκοβιανές Αλυσίδες πάμε λοιπόν στο ερώτημα (ε).

Παίρνουμε $\alpha \triangleq \sup_i \theta_i$, $\beta \triangleq \inf_i \theta_i$, $\theta_i > 0$ $\gamma \triangleq \sup_i \varphi_i$, $\delta \triangleq \inf_i \varphi_i$, $\varphi_i > 0$ με θ να είναι το δεξί ιδιοδιάνυσμα.

$$\frac{\gamma}{\beta} B^n(i, j) \theta_j \geq B^n(i, j) \varphi_j \geq \frac{\delta}{\alpha} B^n(i, j) \theta_j$$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \log \left[\sum_{j=1}^{|\Sigma|} B^n(i, j) \varphi_j \right] \right) = \lim_{n \rightarrow \infty} \left(\frac{1}{n} \log \left[\sum_{j=1}^{|\Sigma|} B^n(i, j) \theta_j \right] \right)$$

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \log (p_i^{n\theta}) \right) = \log p$$

Με παρόμοιο τρόπο καταλήγουμε στο ότι

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \log \left[\sum_{j=1}^{|\Sigma|} B^n(i, j) \varphi_j \right] \right) = \log p$$

4.5 Θεώρημα Cramer για Μαρκοβιανές Αλυσίδες

Θεώρημα Cramer για Μαρκοβιανές Αλυσίδες

Έστω μια μαρκοβιανή αλυσίδα $\{Y_k\}$ που επεξεργάζεται τον μη υποβιβασίμο πίνακα Π. Για κάθε

$$z \in \mathbb{R}^d$$

ισχύει

$$I(z) \triangleq \sup_{\lambda \in \mathbb{R}^d} (\langle \lambda, z \rangle - \log p(\Pi_\lambda)) \quad \text{όπου } \Pi_\lambda \text{ είναι ο πίνακας με στοιχεία}$$

$\pi_\lambda(x, y) = \pi(x, y) e^{\lambda f(y)}$. Ακόμα χρησιμοποιούμε την σύμβαση ότι ο Π_λ είναι και αυτός μη υποβιβασίμος.

Ο εμπειρικός μέσος Z_n ικανοποιεί τις (LDP) με μια κυρτή και “καλή” συνάρτηση I. Επιπρόσθετα για κάθε σύνολο $\Gamma \subseteq \mathbb{R}^d$ και για οποιαδήποτε αρχική κατάσταση $\sigma \in \Sigma$

$$\begin{aligned} -\inf_{z \in \Gamma^o} I(z) &\leq \liminf_{n \rightarrow \infty} \left(\frac{1}{n} \log P_\sigma^\pi(Z_n \in \Gamma) \right) \\ -\inf_{z \in \Gamma^o} I(z) &\leq \liminf_{n \rightarrow \infty} \left(\frac{1}{n} \log P_\sigma^\pi(Z_n \in \Gamma) \right) \leq -\inf_{z \in \Gamma} I(z) \end{aligned}$$

Απόδειξη

$$A_n(\lambda) \triangleq \log E_\sigma^\pi [e^{\langle \lambda, Z_n \rangle}]$$

Το Θεώρημα Gartner-Ellis αρκεί για να δούμε αν το όριο

$$A_n(\lambda) \triangleq \lim_{n \rightarrow \infty} \left(\frac{1}{n} A_n(n\lambda) \right) = \lim_{n \rightarrow \infty} \left(\frac{1}{n} \log E_\sigma^\pi [e^{n \langle \lambda, Z_n \rangle}] \right)$$

υπάρχει $\forall \lambda \in \mathbb{R}^d$, η A είναι πεπερασμένη και παντού παραγωγίσιμη στο \mathbb{R}^d

$$A_n(n\lambda) = \log E_\sigma^\pi [e^{\langle \lambda, \sum_{k=1}^n x_k \rangle}]$$

$$A_n(n\lambda) = \log \sum_{y_1, \dots, y_n} P_\sigma^\pi(Y_1 = y_1, \dots, Y_N = y_n) \prod_{k=1}^n e^{\langle \lambda, f(y_k) \rangle}$$

$$A_n(n\lambda) = \log \sum_{y_1, \dots, y_n} \pi(\sigma, y_1) e^{\langle \lambda, f(y_1) \rangle}, \dots, \pi(y_{n-1}, y_n) e^{\langle \lambda, f(y_n) \rangle}$$

$$A_n(n\lambda) = \log \sum_{y_n=1}^{|\Sigma|} (\Pi_\lambda)^n(\sigma, y_n)$$

Αφού ο πίνακας είναι ανάγωγος από το Θεώρημα Perron-Frobenius έχουμε

$$A(\lambda) = \lim_{n \rightarrow \infty} \left(\frac{1}{n} A_n(n\lambda) \right) = \log p(\Pi_\lambda)$$

Άρα δείξαμε ότι το $A(\lambda)$ είναι αυτό που ψάχναμε.

$$I(z) \triangleq \sup_{\lambda \in \mathbb{R}^d} (\langle \lambda, z \rangle - \log p(\Pi_\lambda))$$

Βιβλιογραφία

- [1] Amir Dembo, Ofer Zeitouni :Large Deviations Techniques and Applications, Second Edition
1998, Springer
- [2] Frank den Hollander :Large Deviations , American Mathematical Society
- [3] Γεώργιος Κοκολάκης και Ιωάννης Σπηλιώτης :Εισαγωγή στις Πιθανότητες, Αθήνα 2002
- [4] Μιχάλης Λουλάκης :Θεωρία Μεγάλων Αποκλίσεων, Ηράκλειο 2005
- [5] Δημήτρης Φουσκάκης :Ανάλυση Δεδομένων με τη Χρήση της R, Εκδόσεις Τσότρας
- [6] Ιωάννης Σπηλιώτης : Στοχαστικές Διαφορικές Εξισώσεις, Αθήνα 2004