Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Εφαρμοσμενων Μαθηματικων και Φυσικων Επιστημων

ΘΕΩΡΙΑ ΜΕΓΑΛΩΝ ΑΠΟΚΛΙΣΕΩΝ & ΕΦΑΡΜΟΓΕΣ

Εκπόνηση Διπλωματικής : Αντύπας Ευάγγελος

Επιβλέπων Καθηγητής : Βασίλειος Παπανικολάου
Μέλος Επιτροπής : Μιχαήλ Λουλάκης
Μέλος Επιτροπής : Γεώργιος Σμυρλής

Αθήνα 2021

Η σελίδα αφήνεται εσκεμμένα κενή.

Ευχαριστίες,

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της διπλωματικής κύριο Παπανικολάου για την πολύτιμη καθοδήγηση και τη συνολική βοήθεια που προσέφερε καθ΄ όλη τη διάρκεια της εκπόνησης της εργασίας και κατά τη διαδικασία των μεταπτυχιακών μου αιτήσεων.

Θα ήθελα επίσης να ευχαριστήσω τους συναδέλφους Νικήτα Σταθάτο και Νίκο Γεωργουδιό για τη βοήθεια τους καθ΄ όλη τη διάρκεια των σπουδών μου, αλλά και για τις συμβουλές τους για τη βελτίωση της διπλωματικής εργασίας.

Τέλος, το μεγαλύτερο ευχαριστώ το οφείλω στην οικογενεία μου για την αμέριστη στήριξη που μου παρείχε καθ΄ όλη τη διάρκεια των προπτυχιακών σπουδών μου.

Περίληψη

Η θεωρία των μεγάλων αποκλίσεων, ως κλάδος της θεωρίας πιθανοτήτων ασχολείται ουσιαστικά με τη μελέτη σπανίων ενδεχομένων. Μας ενδιαφέρουν ενδεχομένα τα οποία παρουσιάζουν σημαντική απόκλιση από την τυπική συμπεριφορά σε έναν δεδομένο χώρο πιθανότητας. Ένα τέτοιο ενδεχόμενο θα μπορούσε για παράδειγμα να είναι η κατάρρευση του χρηματιστηρίου ή η εμφάνιση μιας παγκόσμιας πανδημίας. Είναι εμφανές ότι τέτοια ενδεχόμενα εάν και σπάνια φέρουν μεγάλη βαρύτητα και συνεπώς είναι επιτακτική η ανάγκη να έχουμε μια εκτίμηση για την πιθανότητά τους. Ο σκοπός μας σε αυτήν τη διπλωματική εργασία θα είναι να αναπτύξουμε τα βασικά αποτελέσματα της θεωρίας των μεγάλων αποκλίσεων προκειμένου να μπορέσουμε να παράξουμε εκτιμήσεις για την πιθανότητα να εμφανιστεί ένα σπάνιο ενδεχόμενο. Η εργασία είναι ουσιαστικά δομημένη σε δύο μέρη, το πρώτο μέρος αφιερώνεται στη θεμελίωση της θεωρίας, ενώ το δεύτερο μέρος το οποίο αποτελείται ουσιαστικά από το τελευταίο κεφάλαιο, αφιερώνεται στη μελέτη μεγάλων αποκλίσεων για μαρκοβιανές αλυσίδες με πεπερασμένο χώρο καταστάσεων.

Abstract

The theory of large deviations concerns itself with the study of unlikely events. We are interested in events that present a significant deviation from the typical behaviour in a given probability space. Such events could be, for example a stock market crash or a global pandemic. It is obvious that such events as rare as they may be are of great importance to society, so it is an imperative need to get an estimate of their probability. In this thesis our goal is to lay the groundwork for the most important theoretical results in large deviation theory, in order to be able to estimate the probability of rare events. The thesis is essentially structured in two parts, the first part is dedicated to developing the foundations of large deviation theory, while the second part, which essentially consists of the last chapter is dedicated to developing large deviation results for finite state Markov chains.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή στις Μεγάλες Αποκλίσεις

1.1 Εισαγωγή

Η θεωρία των μεγάλων αποκλίσεων αποτελεί ένα κλασικό παράδειγμα θεωρίας η οποία γεννήθηκε μέσω των εφαρμοσμένων μαθηματικών και διαχρονικά εξελίχθηκε σε μια πολύ κομψή μαθηματική θεωρία. Ιστορικά οι μεγάλες αποκλίσεις αναπτύχθηκαν προκειμένου να απαντηθούν ερωτήματα σε δύο διαφορετικές περιοχές των εφαρμοσμένων μαθηματικών. Αρχικά με τον Cramér στην περιοχή των ασφαλιστικών μαθηματικών και στη συνέχεια στην περιοχή της μαθηματικής στατιστικής κυρίως με τους Chernoff και Bahadur.

Ο Cramér ο οποίος θεωρείται ο βασικός αρχιτέκτονας της θεωρίας των μεγάλων αποκλίσεων, ενδιαφέρθηκε για το εξής ερώτημα, ας υποθέσουμε ότι μας ανήκει μια ασφαλιστική εταιρία με έναν μεγάλο αριθμό πελατών έστω n και οι ασφαλιστικές αποζημιώσεις για κάθε πελάτη είναι ανεξάρτητες και ισόνομες πραγματικές τυχαίες μεταβλητές, δηλαδή $X_1, ..., X_n \sim \mu$, όπου μ είναι η κατανομή των X_i. Αυτό σημαίνει ότι $\forall i \in \{1, ..., n\} \quad \mathbb{P}(X_i \in B) = \mu(B) \quad \forall B \subset \mathbb{R}$ σύνολο Borel. Επιπλέον υποθέτουμε ότι οι X_i έχουν πεπερασμένη μέση τιμή m και ότι το ασφάλιστρο που χρεώνει η εταιρία μας είναι κάποιος πραγματικός αριθμός $a > m$, τότε η πιθανότητα να παρουσιάσουμε οικονομική απώλεια θα είναι:

$$\mathbb{P}(X_1 + ... + X_n > na) = \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} X_i > a) = \mathbb{P}(\bar{X}_n > a) := p_n.$$

Εφόσον $a > m$ γνωρίζουμε από τον Ασθενή Νόμο των Μεγάλων Αριθμών ότι θα ισχύει,

$$p_n \to 0, \quad \text{καθώς} \quad n \to \infty.$$

Το ερώτημα που ενδιέφερε τον Cramér είναι με ποια ταχύτητα συμβαίνει η σύγκλιση $p_n \to 0$. Με άλλα λόγια κατα πόσο είναι δυνατόν να εκτιμήσουμε την πιθανότητα χρεοκοπίας της ασφαλιστικής μας εταιρίας για μεγάλα n. Την απάντηση στο ερώτημα αυτό θα την δώσουμε παρακάτω με τη διατύπωση του θεωρήματος Cramér.

Από την άλλη πλευρά στατιστικοί όπως οι Chernoff και Bahadur ενδιαφέρθηκαν για το ίδιο ερώτημα ερχόμενοι όμως από την περιοχή των ελέγχων στατιστικών υποθέσεων. Ας υποθέσουμε ότι έχουμε ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, ..., X_n \sim N(m, 1)$ και το m είναι άγνωστο, είναι γνωστό όμως ότι ή $m = 4$ είτε $m = 6$. Το ερώτημα είναι τότε, πως μπορούμε να προσδιορίσουμε εάν $m = 4$ ή $m = 6$. Με άλλα λόγια πως μπορούμε να κάνουμε τον έλεγχο:

- $H_0 : m = 6$
- $H_1 : m = 4$

Μπορούμε να προσεγγίσουμε το πρόβλημα ως εξής, εάν $\bar{X}_n < 5$ απορρίπτουμε την H_0, αλλιώς εάν $\bar{X}_n > 5$ αποδεχόμαστε την H_0 (εάν $\bar{X}_n = 5$ απορρίπτουμε την H_0 με πιθανότητα $\frac{1}{2}$ και αντίστοιχα την αποδεχόμαστε με πιθανότητα $\frac{1}{2}$). Σε αυτήν την περίπτωση, εάν η πραγματική κατανομή είναι η $N(4, 1)$ η πιθανότητα να προβούμε σε λανθασμένη κρίση είναι η $\mathbb{P}(\bar{X}_n > 5)$ υπολογισμένη υπό την υπόθεση ότι $X_1, ..., X_n \sim N(4, 1)$, πάλι από τον Ασθενή Νόμο των Μεγάλων Αριθμών θα έχουμε ότι:

$$p_n = \mathbb{P}_{m=4}(\bar{X}_n > 5) \to 0, \quad \text{καθώς} \quad n \to \infty.$$

Και εδώ το ερώτημα που μας ενδιαφέρει είναι με ποια ταχύτητα συμβαίνει η σύγκλιση $p_n \to 0$. Με άλλα λόγια μας ενδιαφέρει με ποια ταχύτητα η ισχύς αυτού του ελέγχου τείνει στη μονάδα, δηλαδή $\mathbb{P}_{H_1}(\text{Απόρριψη της } H_0) = 1 - p_n \to 1, \quad \text{καθώς} \quad n \to \infty$. Χάριν ευκολίας επιλέξαμε την κατανομή του πληθυσμού να είναι κανονική, θα μπορούσαμε όμως να επαναλάβουμε το ερώτημα για οποιαδήποτε κατανομή με πεπερασμένη μέση τιμή m και διασπορά $\sigma^2 = 1$.

Στο σημείο αυτό θα πρέπει να σχολιάσουμε ότι τα δύο αυτά ερωτήματα που προέρχονται από διαφορετικές περιοχές των μαθηματικών είναι εν τέλει το ίδιο ερώτημα. Στην ουσία θέλουμε να υπολογίσουμε *μια μεγάλη απόκλιση* του δειγματικού μέσου από τον πληθυσμιακό μέσο. Αυτά τα ερωτήματα οδήγησαν στην ανάπτυξη της θεωρίας των μεγάλων αποκλίσεων. Ο Cramér βρήκε έναν τρόπο να υπολογίσει την πιθανότητα αυτής της μεγάλης απόκλισης για οποιαδήποτε κατανομή, κάτω από λογικές υποθέσεις. Από την άλλη πλευρά οι Chernoff και Bahadur παρήγαγαν ασυμπτωτικές εκτιμήσεις γι' αυτήν την πιθανότητα. Εάν και οι κεντρικές ιδέες της θεωρίας είναι αρκετά παλιές και πηγαίνουν πίσω έως και το 1900, οι πιο σημαντικές εξελίξεις έχουν σημειωθεί κατά το δεύτερο μισό του 20ου αιώνα με τους Donsker, Varadhan, Freidlin, Wentzel. Με τον Varadhan να είναι ο κύριος υπεύθυνος για την ενοποίηση των μεγάλων αποκλίσεων ως ενιαίας μαθηματικής περιοχής. Η θεωρία βρίσκει εφαρμογή σε ποικίλα πεδία όπως είναι η στατιστική, τα χρηματοοικονομικά και ασφαλιστικά μαθηματικά, η επιστήμη υπολογιστών και η στατιστική μηχανική.

1.2 Ένα Κινητήριο Παράδειγμα

Έστω ότι εργαζόμαστε σε έναν τυπικό χώρο πιθανότητας $(\Omega, \mathcal{F}, \mathbb{P})$ και έχουμε τις ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, ..., X_n \sim N(0,1)$. Από τον Ασθενή Νόμο των Μεγάλων Αριθμών γνωρίζουμε ότι $\forall \delta > 0 \quad p_n = \mathbb{P}(\bar{X}_n > \delta) \to 0$, καθώς $n \to \infty$. Μας ενδιαφέρει να δώσουμε μια εκτίμηση για την ταχύτητα αυτής της σύγκλισης.

Από το Κεντρικό Οριακό Θεώρημα γνωρίζουμε ότι $\sqrt{n}\bar{X}_n \sim N(0,1)$, συνεπώς θα έχουμε

$$p_n = \mathbb{P}(\bar{X}_n > \delta) = \mathbb{P}(\sqrt{n}\bar{X}_n > \sqrt{n}\delta) = 1 - \Phi(\sqrt{n}\delta), \tag{1.1}$$

όπου Φ είναι η αθροιστική συνάρτηση κατανομής και $\bar{X}_n$ ο δειγματικός μέσος. Για τη συνέχεια θα μας χρειαστούν τα εξής :

Ορισμός 1. *Έστω τ.μ. $X \sim N(0,1)$ ορίζουμε ως συνάρτηση πιθανότητας ουράς (tail distribution function) τη συνάρτηση*

$$Q(x) = \mathbb{P}(X > x) = 1 - \Phi(x).$$

Πρόταση 1. *Για τη συνάρτηση Q ισχύουν τα εξής φράγματα*

$$\frac{x}{1+x^2}\phi(x) < Q(x) \le \frac{\phi(x)}{x}, \quad x > 0, \tag{1.2}$$

όπου $\phi(x)$ είναι η συνάρτηση πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής.

Απόδειξη. Για το άνω φράγμα χρησιμοποιώντας την αλλαγή μεταβλητής $v = \frac{u^2}{2}$ θα έχουμε

$$Q(x) = \int_x^\infty \phi(u)du \le \int_x^\infty \frac{u}{x}\phi(u)du = \int_{\frac{x^2}{2}}^\infty \frac{e^{-v}}{x\sqrt{2\pi}}dv = -\frac{e^{-v}}{x\sqrt{2\pi}}\Bigg|_{\frac{x^2}{2}}^\infty = \frac{\phi(x)}{x}.$$

Για το κάτω φράγμα χρησιμοποιώντας την ταυτότητα $\phi'(u) + u\phi(u) = 0$ έχουμε

$$\left(1 + \frac{1}{x^2}\right)Q(x) = \int_x^\infty \left(1 + \frac{1}{x^2}\right)\phi(u)du > \int_x^\infty \left(1 + \frac{1}{u^2}\right)\phi(u)du =$$

$$= \int_x^\infty \left(1 + \frac{1}{u^2}\right)\frac{-\phi'(u)}{u}du = -\frac{\phi(u)}{u}\Bigg|_x^\infty = \frac{\phi(x)}{x},$$

που αποδεικνύει το ζητούμενο κάτω φράγμα.

□

Ορισμός 2. *Έστω f, g συναρτήσεις, τότε θα γράφουμε $f \sim g$ εάν*

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1.$$

Οπότε τώρα είμαστε σε θέση να διατυπώσουμε την εξής πρόταση:

Πρόταση 2. *Ισχύει $\frac{Q(x)}{\phi(x)} \sim \frac{1}{x}$ (Mill's ratio).*

Απόδειξη. Η απόδειξη είναι εύκολη, αρκεί να πολλαπλασιάσουμε με $x > 0$ και να διαιρέσουμε με $\phi(x)$ τη σχέση (1.2) και στη συνέχεια να εφαρμόσουμε το κριτήριο παρεμβολής. □

Επομένως από τη σχέση (1.1) θα έχουμε

$$p_n \sim \frac{1}{\delta\sqrt{2\pi n}} e^{-\frac{\delta^2 n}{2}}, \quad \text{καθώς} \quad n \to \infty.$$

Αυτό σημαίνει ότι η τυπική τιμή του $\bar{X}_n$ είναι της τάξεως του $\frac{1}{\sqrt{n}}$, εφόσον για κάθε $a < b$ έχουμε ότι $\mathbb{P}(\sqrt{n}\bar{X}_n \in (a, b]) = \int_a^b \phi(x)dx$, αλλά με πολύ μικρή πιθανότητα της τάξεως του $\frac{1}{\sqrt{n}} e^{-\frac{n\delta^2}{2}}$ είναι $\bar{X}_n > \delta$.

Σχόλιο 1. *Για μεγάλο n η συμβολή της πολύ μικρής αυτής πιθανότητας προέρχεται ουσιαστικά από τον όρο $e^{-\frac{n\delta^2}{2}}$, επομένως ασυμπτωτικά μπορούμε να αγνοήσουμε τον πολυωνυμικό όρο $\frac{1}{\sqrt{n}}$ και να πούμε ότι ο $\bar{X}_n$ ξεπερνά την τιμή δ με πιθανότητα της τάξεως $e^{-\frac{n\delta^2}{2}}$. Αυτή η εκθετική μείωση της μεγάλης απόκλισης οφείλεται στο γεγονός ότι η πιθανότητας ουράς της κανονικής κατανομής μειώνεται επίσης εκθετικά.*

Για να γίνει πιο κατανοητό τι θα πει ότι μπορούμε να ξεφορτωθούμε τον πολυωνυμικό όρο αρκεί να δούμε την παραπάνω έκφραση υπό λίγο διαφορετική σκοπιά, δηλαδή αντί να γράφουμε $p_n \sim (\text{σταθ.})\frac{1}{\sqrt{n}} e^{-\frac{n\delta^2}{2}}$ εξετάζουμε τον όρο

$$\frac{1}{n}\log p_n = \frac{\log((\text{σταθ.})) - \frac{1}{2}\log n - \frac{\delta^2}{2}}{n} \to -\frac{\delta^2}{2}, \quad \text{καθώς} \quad n \to \infty.$$

Αυτή είναι μια κλασική εκτίμηση στο πλαίσιο της θεωρίας των μεγάλων αποκλίσεων. Συνεπώς με την έως τώρα μελέτη μας έχουμε αποδείξει το εξής αποτέλεσμα:

Πρόταση 3. *Έστω τ.μ. $X_1, ..., X_n \sim N(0,1)$, τότε $\forall \delta > 0$ έχουμε*

$$\frac{1}{n} \log \mathbb{P}(\bar{X}_n > \delta) \to -\frac{\delta^2}{2}, \quad \textit{καθώς} \quad n \to \infty.$$

1.3 Γενικεύοντας το Παράδειγμα

Το παράδειγμα που μελετήσαμε ήταν εστιασμένο στην τυπική κανονική κατανομή, τι γίνεται όμως στην περίπτωση που $X_1, ..., X_n \sim \mu$, όπου μ είναι μια αυθαίρετη κατανομή τέτοια ώστε $\int_{\mathbb{R}} |x| d\mu(x) < \infty$, άρα η μέση τιμή $\mathbb{E}(X_i)$ είναι πεπερασμένη.

Σχόλιο 2. *Από εδώ και στο εξής με το ακρωνύμιο ΑΝΜΑ θα αναφερόμαστε στον Ασθενή Νόμο των Μεγάλων Αριθμών.*

Έστω $\delta > 0$ τότε ΑΝΜΑ $\Rightarrow p_n := \mathbb{P}(\bar{X}_n > m + \delta) \to 0$ καθώς $n \to \infty$. Όπως και πριν το ερώτημα είναι με τι ταχύτητα συμβαίνει αυτή η σύγκλιση.

Εάν υποθέσουμε ότι η μ έχει αρκούντως ελαφριά ουρά, δηλαδή η πιθανότητα ουράς μειώνεται εκθετικά, τότε είναι λογικό να αναμένουμε ότι όπως και στην περίπτωση της κανονικής κατανομής η p_n πηγαίνει στο 0 εκθετικά γρήγορα. Επομένως, ίσως να μπορούμε να βρούμε μια σχέση της μορφής

$$p_n \sim A n^b e^{-cn}, \quad n \to \infty, \tag{1.3}$$

για $A, b \in \mathbb{R}$ και κάποια σταθερά $c > 0$ η οποία εξαρτάται από το δ και την κατανομή μ.

Εάν ισχύει η σχέση (1.3) τότε μπορούμε όπως και στην κανονική περίπτωση να βρούμε c τέτοιο ώστε,

$$\lim_{n\to\infty} \frac{1}{n} \log p_n = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n > \delta + m) = -c.$$

Το ερώτημα τώρα επομένως γίνεται, πως υπολογίζουμε ένα τέτοιο $c = c(\delta, \mu)$ για κατανομές που δεν είναι η κανονική. Η απάντηση στο ερώτημα αυτό δεν είναι απλή υπόθεση και θα μας οδηγήσει στη διατύπωση του θεωρήματος Cramér. Προς το παρόν θα κάνουμε την εξής προσέγγιση.

Έστω $a > m$ γνωρίζουμε ότι ΑΝΜΑ $\Rightarrow p_n := \mathbb{P}(\bar{X}_n > a) \to 0, \quad n \to \infty$. Θέλουμε να εκτιμήσουμε την p_n για κάθε $a > m$, η εκτίμηση αυτή θα μας δώσει ένα άνω κι ένα κάτω φράγμα. Ας ξεκινήσουμε με το άνω φράγμα, η αρχική μας ιδέα είναι να χρησιμοποιήσουμε την ανισότητα Markov και θα έχουμε

$$p_n := \mathbb{P}(\bar{X}_n > a) \leq \frac{\mathbb{E}(\bar{X}_n)}{a} = \frac{m}{a}.$$

Αυτό που ελπίζουμε να πετύχουμε είναι ένα άνω φράγμα το οποίο να πηγαίνει εκθετικά στο 0, αντ΄ αυτού το φράγμα που έχουμε δεν συγκλίνει καν. Για να λύσουμε αυτό το πρόβλημα χρησιμοποιούμε την εκθετική ανισότητα Markov. Υποθέτουμε ότι η ροπογεννήτρια συνάρτηση $M_{X_1}(\theta) = \mathbb{E}(e^{\theta X_1}) = \int_{\mathbb{R}} e^{\theta x} d\mu(x) < \infty, \quad \forall \theta > 0$ και έχουμε,

$$p_n = \mathbb{P}(\sum_{i=1}^{n} X_i > na) = \mathbb{P}(e^{\theta \sum_{i=1}^{n} X_i} > e^{n\theta a}) \leq \frac{\mathbb{E}(e^{\theta \sum_{i=1}^{n} X_i})}{e^{n\theta a}} = \frac{M_{X_1}(\theta)^n}{e^{n\theta a}},$$

εφόσον οι $X_1, ..., X_n$ είναι ανεξάρτητες και ισόνομες έχουν όλες ροπογεννήτρια την $M_{X_1}(\theta)$. Είμαστε τώρα σε θέση να ορίσουμε τη *λογαριθμική ροπογεννήτρια*,

$$\Lambda(\theta) = \log M_{X_1}(\theta) = \log \int_{\mathbb{R}} e^{\theta x} d\mu(x), \quad \theta \in \mathbb{R}.$$

Σχόλιο 3. *Εν γένει θα μπορούσαμε να έχουμε $\Lambda(\theta) = \infty$ για αυτό το λόγο κάνουμε την αρχική μας υπόθεση που μας εξασφαλίζει ότι $\Lambda(\theta) < \infty, \quad \forall \theta > 0$. Επιπλέον αξίζει να παρατηρήσουμε ότι $\Lambda(0) = 0$.*

Με αυτήν την υπόθεση έχουμε πετύχει

$$p_n \leq \frac{e^{n\Lambda(\theta)}}{e^{n\theta a}} = e^{-n(\theta a - \Lambda(\theta))}.$$

Επομένως εάν καταφέρουμε να βρούμε ένα $\theta \geq 0$ τέτοιο ώστε $-n(\theta a - \Lambda(\theta)) < 0$ θα έχουμε ένα άνω φράγμα το οποίο συγκλίνει στο 0 εκθετικά γρήγορα. Πιο συγκεκριμένα το πιο σφιχτό (βέλτιστο) τέτοιο άνω φράγμα θα προκύψει λύνοντας το παρακάτω πρόβλημα βελτιστοποίησης

$$p_n \leq \inf_{\theta \geq 0}(e^{-n(\theta a - \Lambda(\theta))}) = e^{-n[\sup\limits_{\theta \geq 0}(\theta a - \Lambda(\theta))]}.$$

Εφόσον $a > m$ μπορεί να αποδειχθεί (θα το αποδείξουμε στην **Ιδιότητα 10**) ότι

$$\sup_{\theta \geq 0}(\theta a - \Lambda(\theta)) = \sup_{\theta \in \mathbb{R}}(\theta a - \Lambda(\theta)). \tag{1.4}$$

Επομένως έχουμε κάνει πλέον σημαντική πρόοδο, έχοντας μετατρέψει το προβλημά μας στο παραπάνω πρόβλημα βελτιστοποίησης, το οποίο με τη σειρά του μας οδηγεί στον εξής ορισμό.

Ορισμός 3. *Έστω $\Lambda(\theta) = \log \int_{\mathbb{R}} e^{\theta x} d\mu(x) = \log \mathbb{E}(e^{\theta X_1}), \quad \theta \in \mathbb{R}$. Τότε ορίζουμε το μετασχηματισμό Fenchel-Legendre της Λ να είναι*

$$\Lambda^*(a) \coloneqq \sup_{\theta \in \mathbb{R}} (\theta a - \Lambda(\theta)), \quad a \in \mathbb{R}.$$

Θα μελετήσουμε τις ιδιότητες της Λ^* αναλυτικά στη συνέχεια, προς το παρόν όμως ας ανακεφαλαιώσουμε τι έχουμε δείξει, υποθέτοντας ότι ισχύει η ισότητα στη σχέση (1.4) έχουμε

$$p_n \leq e^{-n\Lambda^*(a)}, \forall a > m. \tag{1.5}$$

Σχόλιο 4. *Ισχύει ότι,*

$$\forall a \in \mathbb{R}, \Lambda^*(a) = \sup_{\theta \in \mathbb{R}} (\theta a - \Lambda(\theta)) \geq 0 - \Lambda(0) = 0.$$

Επομένως έχουμε ότι $p_n \to 0$ εκθετικά γρήγορα, με άλλα λόγια ισοδύναμα με την (1.5) έχουμε,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n > a) \leq -\Lambda^*(a).$$

Επιπλέον, χρησιμοποιώντας ένα επιχείρημα αλλαγής μέτρου (change of measure) θα δείξουμε παρακάτω ότι και το κάτω φράγμα που αναζητάμε έχει παρόμοια μορφή δηλαδή,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n > a) \geq -\Lambda^*(a).$$

Για την ακρίβεια θα δείξουμε ένα γενικότερο αποτέλεσμα, από το οποίο το κάτω φράγμα θα προκύψει ως ειδική συνέπεια.

1.4 Το Θεώρημα Cramér στο $\mathbb{R}$

Ένα πιο Γενικό Ερώτημα. *Χτίζοντας πάνω στη λογική που έχουμε χρησιμοποιήσει έως τώρα, ας υποθέσουμε ότι $B \subset \mathbb{R}$ είναι σύνολο Borel του $\mathbb{R}$ τέτοιο ώστε να είναι μακριά από τη μέση τιμή m, δηλαδή $m \notin B$, τότε θα ισχύει και πάλι ότι $\mathbb{P}(\bar{X}_n \in B) \to 0$ και ενδιαφερόμαστε για την ταχύτητα σύγκλισης. Κατ΄ αναλογία με την προηγούμενη παράγραφο μας ενδιαφέρει να εξετάσουμε κατά πόσο μπορούμε να υπολογίσουμε την ποσότητα*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in B). \tag{1.6}$$

Γενικά το ερώτημα αυτό δεν έχει απάντηση για οποιοδήποτε σύνολο Borel, αλλά θα προσδιορίσουμε κάποιες καλές συνθήκες κάτω από τις οποίες μπορούμε να παράξουμε αποτελέσματα.

Σημειώνουμε ότι εν γένει ισχύει,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in B) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in \bar{B}),$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in B) \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in B^o),$$

όπου $\bar{B}$ είναι η κλειστότητα του B και B^o το εσωτερικό του B αντίστοιχα. Επομένως μας είναι αρκετό να δώσουμε ένα άνω φράγμα για τον όρο $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in \bar{B})$ και ένα κάτω φράγμα για τον όρο $\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in B^o)$ αντίστοιχα. Στις καλές περιπτώσεις που αυτά τα δύο φράγματα συμπίπτουν θα μπορούμε να υπολογίσουμε το όριο της σχέσης (1.6).

Προς στιγμήν φαίνεται σαν να περιπλέξαμε την κατάσταση περισσότερο με το να είσαγουμε τα $\bar{B}, B^o$, αλλά η παρατήρηση κλειδί εδώ είναι ότι αρκεί να δώσουμε τα αντίστοιχα φράγματα για

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in F), \quad \text{όπου} \quad F \subset \mathbb{R} \quad \text{κλειστό},$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in U), \quad \text{όπου} \quad U \subset \mathbb{R} \quad \text{ανοιχτό}.$$

Γεγονός που μας οδηγεί στο επόμενο θεώρημα.

Θεώρημα 1 (Cramér). *Έστω μ κατανομή πιθανότητας στο $\mathbb{R}$ και οι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, ..., X_n \sim \mu$, ορίζουμε $\Lambda : \mathbb{R} \to (-\infty, \infty]$ με $\Lambda(\theta) = \log \int_{\mathbb{R}} e^{\theta x} d\mu(x)$ και $\Lambda^*(x) := \sup_{\theta \in \mathbb{R}}(\theta x - \Lambda(\theta)), x \in \mathbb{R}$ (μπορεί να είναι και $+\infty$) να είναι ο μετασχηματισμός Fenchel-Legendre της Λ, τότε έχουμε τα εξής φράγματα:*

(α) Εάν $F \subset \mathbb{R}$ κλειστό, τότε

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in F) \leq - \inf_{x\in F} \Lambda^*(x).$$

(β) Εάν $U \subset \mathbb{R}$ ανοικτό, τότε

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in U) \geq - \inf_{x\in U} \Lambda^*(x).$$

Σχόλιο 5. *Στο σημείο αυτό αξίζει να παρατηρήσουμε ότι το θεώρημα Cramér δεν υποθέτει ότι η ποσότητα $\mathbb{E}(e^{\theta X_1})$ είναι πεπερασμένη, για την ακρίβεια δεν υποθέτει καν ότι η m είναι πεπερασμένη.*

1.4.1 Ιδιότητες των Λ και Λ^*

Προτού περάσουμε στην απόδειξη πρέπει πρώτα να μελετήσουμε και να αποδείξουμε κάποιες βασικές ιδιότητες των Λ και Λ^*. Θα ξεκινήσουμε κάνοντας μερικές συμβάσεις:

(α) Θα λέμε ότι η X_1 (ή ισοδύναμα η μ) έχει πεπερασμένη μέση τιμή εάν $\int_{x>0} x d\mu(x) < \infty$ και $\int_{x<0} x d\mu(x) < \infty$ (ή ισοδύναμα η $\int_{x\in\mathbb{R}} |x| d\mu(x) < \infty$). Σε αυτήν την περίπτωση θα έχουμε $m := \int_{x\in\mathbb{R}} x d\mu(x) = \mathbb{E}(X_1)$.

(β) Θα λέμε ότι η μέση τιμή m υπάρχει εάν τουλάχιστον ένα από τα ολοκληρώματα $\int_{x>0} x d\mu(x)$ ή $\int_{x<0} x d\mu(x)$ είναι πεπερασμένο. Φυσικά σε αυτήν την περίπτωση η m μπορεί να είναι και άπειρη.

(γ) Εάν και τα δύο ολοκληρώματα είναι άπειρα θα λέμε ότι η μέση τιμή m δεν υπάρχει.

Ιδιότητα 1. *Για θετικές τυχαίες μεταβλητές η μέση τιμή υπάρχει πάντα και είναι θετική. Πιο συγκεκριμένα η συνάρτηση $\Lambda(\theta)$ είναι καλώς ορισμένη $\forall \theta \in \mathbb{R}$ και παίρνει τιμές στο διάστημα $(-\infty, +\infty]$.*

Ιδιότητα 2. *Η Λ είναι κυρτή.*

Απόδειξη. Έστω $c \in (0, 1)$ και $\theta, \psi \in \mathbb{R}$, τότε,

$$\Lambda(c\theta + (1-c)\psi) = \log \int_{\mathbb{R}} e^{(c\theta+(1-c)\psi)x} d\mu(x). \tag{1.7}$$

Επιπλέον, χρησιμοποιώντας την ανισότητα Hölder, δηλαδή $\|fg\| \le \|f\|_p \|g\|_q$, όπου $\frac{1}{p}+\frac{1}{q}=1$ συζυγείς εκθέτες, για $p=\frac{1}{c}$ και $q=\frac{1}{1-c}$ θα έχουμε,

$$\int_{\mathbb{R}} e^{(c\theta+(1-c)\psi)x} d\mu(x) = \int_{\mathbb{R}} e^{c\theta x} e^{(1-c)\psi x} d\mu(x) \le \left[\int_{\mathbb{R}} e^{\theta x} d\mu(x)\right]^c \left[\int_{\mathbb{R}} e^{\psi x} d\mu(x)\right]^{1-c}.$$

Επομενως λογαριθμίζοντας και τα δύο μέλη θα πάρουμε

$$\log \int_{\mathbb{R}} e^{(c\theta+(1-c)\psi)x} d\mu(x) \le c \log\left[\int_{\mathbb{R}} e^{\theta x} d\mu(x)\right] + (1-c)\log\left[\int_{\mathbb{R}} e^{\psi x} d\mu(x)\right]$$

Συνεπώς η σχέση (1.7) θα μας δώσει

$$\Lambda(c\theta+(1-c)\psi) \le c\Lambda(\theta)+(1-c)\Lambda(\psi),$$

άρα η Λ είναι κυρτή. □

Ιδιότητα 3. *Ισχύει* $\Lambda^* \ge 0$.

Απόδειξη.

$$\forall x \in \mathbb{R} \quad \Lambda^*(x) = \sup_{\theta\in\mathbb{R}}[\theta x - \Lambda(\theta)] \ge 0x - \Lambda(0) = 0 - \log 1 = 0.$$

□

Ιδιότητα 4. *Η* Λ^* *είναι κυρτή.*

Απόδειξη.

$$\forall x \in \mathbb{R} \quad \Lambda^*(x) = \sup\{\theta x - \Lambda(\theta), \theta \in \mathbb{R}\} = \sup\{\theta x - \Lambda(\theta), \theta \in \mathbb{R} | \Lambda(\theta) < \infty\}.$$

Εφόσον $\Lambda(0)=0$ οι τιμές του ϑ για τις οποίες $\Lambda(\theta)=\infty$ δε συνεισφέρουν στο supremum. Επομένως ορίζουμε $L=\{\theta\in\mathbb{R}|\Lambda(\theta)<\infty\}$, προφανώς το L δεν είναι κενό γιατί έχουμε ήδη αποδείξει ότι $0\in L$. Είναι προφανές ότι με αυτή τη σύμβαση $\Lambda^*(x)=\sup_{\theta\in L}(\theta x-\Lambda(\theta))$.

Έστω $x, y \in \mathbb{R}, \quad c \in (0,1), \quad \theta_0 \in L$, τότε

$$\theta_0(cx+(1-c)y) - \Lambda(\theta_0) = c(\theta_0 x - \Lambda(\theta_0)) + (1-c)(\theta_0 y - \Lambda(\theta_0)) \le$$

$$\le c\sup_{\theta\in L}(\theta_0 x - \Lambda(\theta_0)) + (1-c)\sup_{\theta\in L}(\theta_0 y - \Lambda(\theta_0)) = c\Lambda^*(x) + (1-c)\Lambda^*(y).$$

Τώρα παίρνοντας το supremum για $\theta_0 \in L$ θα έχουμε,

$$\Lambda^*(cx + (1-c)y) \leq c\Lambda^* + (1-c)\Lambda^*(y),$$

άρα η Λ^* είναι κυρτή. □

Σχόλιο 6. *Για κάθε $\theta \in L$ η συνάρτηση $a_\theta(x) = \theta x - \Lambda(\theta), \quad x \in \mathbb{R}$ είναι αφινική συνάρτηση (affine function). Η Λ^* είναι το κατά σημείο supremum των a_θ και επομένως είναι κυρτή. Εν γένει το κατά σημείο supremum οποιασδήποτε συλλογής αφινικών συναρτήσεων μπορεί να αποδειχθεί ότι είναι κυρτή συνάρτηση, χρησιμοποιώντας το ίδιο επιχείρημα που χρησιμοποιήθηκε στην απόδειξη της ιδιότητας (4).*

Ιδιότητα 5. *Η Λ^* είναι κάτω ημισυνεχής, δηλαδή $\forall a \in \mathbb{R}$ το σύνολο $\{x \in \mathbb{R} | \Lambda^*(x) > a\}$ είναι ανοικτό.*

Απόδειξη. Όπως είπαμε $\Lambda^*(x) = \sup_{\theta \in L} a_\theta(x), \quad x \in \mathbb{R}$, με την κάθε a_θ να είναι αφινική και επομένως συνεχής συνάρτηση. Συνεπώς για κάθε $a \in \mathbb{R}$,

$$\{x \in \mathbb{R} | \Lambda^*(x) > a\} = \bigcup_{\theta \in L} (a_\theta > a)$$

είναι ανοικτό. □

Ιδιότητα 6. *Εάν $\Lambda(\theta) < \infty$ για κάποιο $\theta > 0$, τότε m υπάρχει και $-\infty \leq m < \infty$.*

Απόδειξη. Για κάποιο $\theta > 0$,

$$\Lambda(\theta) < \infty \Rightarrow \int_{\mathbb{R}} e^{\theta x} d\mu(x) < \infty \Rightarrow \int_{x>0} e^{\theta x} d\mu(x) < \infty. \tag{1.8}$$

Επιπλέον για κάθε $x > 0$,

$$\theta x < e^{\theta x} \Rightarrow x < \frac{1}{\theta} e^{\theta x}.$$

Άρα από τη σχέση (1.8) θα έχουμε

$$\int_{x>0} x d\mu(x) < \int_{x>0} \frac{1}{\theta} e^{\theta x} d\mu(x) < \infty.$$

Συνεπώς, έχουμε

$$m = \int_{\mathbb{R}} x d\mu(x) = \int_{x>0} x d\mu(x) + \int_{x<0} x d\mu(x) \in [-\infty, \infty).$$

διότι μπορεί να είναι $\int_{x<0} x d\mu(x) = -\infty$. □

Ιδιότητα 7. *Εάν $m = +\infty$, τότε $\Lambda(\theta) = +\infty$, για κάθε $\theta > 0$.*

Απόδειξη. Η απόδειξη είναι άμεση με εις άτοπον απαγωγή από την ιδιότητα (6). $\square$

Ιδιότητα 8. *Εάν $\Lambda(\theta) < \infty$ για κάποιο $\theta < 0$, τότε m υπάρχει και $-\infty < m \leq \infty$.*

Απόδειξη. Η απόδειξη είναι όμοια με την ιδιότητα (6) με αντίθετα πρόσημα. $\square$

Ιδιότητα 9. *Εάν $m = -\infty$, τότε $\Lambda(\theta) = +\infty$, για κάθε $\theta > 0$.*

Απόδειξη. Η απόδειξη είναι άμεση με εις άτοπον απαγωγή από την ιδιότητα (8). $\square$

Ιδιότητα 10. *Έστω ότι m υπάρχει και $x \geq m$, τότε $\Lambda^*(x) = \sup_{\theta \geq 0}(\theta x - \Lambda(\theta))$, πιο συγκεκριμένα αυτή η ιδιότητα είναι η υπόθεση που κάναμε στη σχέση* (1.4).

Απόδειξη. Εάν $m = +\infty$, τότε δεν έχουμε κάτι να αποδείξουμε. Εάν $m = -\infty$, τότε από την ιδιότητα (9) έχουμε $\Lambda(\theta) = \infty, \quad \theta > 0$, οπότε $\theta x - \Lambda(\theta) = -\infty, \quad \theta < 0$, επομένως

$$\Lambda^*(x) = \sup_{\theta \in \mathbb{R}}(\theta x - \Lambda(\theta)) = \sup_{\theta \geq 0}(\theta x - \Lambda(\theta)).$$

Συνεπώς μένει να ελέξγουμε την περίπτωση $-\infty < m < +\infty$, τότε για κάθε $\theta \in \mathbb{R}$ από την ανισότητα Jensen, δηλαδή $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$, όπου f είναι κυρτή συνάρτηση, επειδή ο λογάριθμος είναι κοίλη συνάρτηση θα αλλάξει η φορά της ανισότητας και επομένως θα έχουμε

$$\Lambda(\theta) \geq \int_{\mathbb{R}} \log e^{\theta x} d\mu(x) = \int_{\mathbb{R}} \theta x d\mu(x) = \theta m.$$

Επομένως για κάθε $\theta \in \mathbb{R}$, έχουμε

$$\theta m - \Lambda(\theta) \leq 0 \Rightarrow \Lambda^*(m) = \sup_{\theta \in \mathbb{R}}(\theta m - \Lambda(\theta)) = 0$$

Από την άλλη $\forall x \geq m$ και $\forall \theta < 0$,

$$\theta x - \Lambda(\theta) \leq \theta m - \Lambda(\theta) \leq 0,$$

ενώ για $\theta = 0$, $\theta x - \Lambda(\theta) = 0x - \Lambda(0) = 0$. Επομένως

$$\forall x \geq m \quad \Lambda^*(x) = \sup_{\theta \geq 0}(\theta x - \Lambda(\theta)),$$

που ήταν το ζητούμενο. $\square$

Σχόλιο 7. *Ουσιαστικά, έχουμε αποδείξει ότι όποτε το m υπάρχει και είναι πεπερασμένο ισχύει $\Lambda^*(m) = 0$.*

Ιδιότητα 11. *Εάν m υπάρχει, τότε η Λ^* είναι αύξουσα στο διάστημα $[m, +\infty)$.*

Απόδειξη. Έστω $y \geq x \geq m$, τότε $\forall \theta \geq 0$,

$$\theta y - \Lambda(\theta) \geq \theta x - \Lambda(\theta) \Rightarrow \sup_{\theta \geq 0}(\theta y - \Lambda(\theta)) \geq \sup_{\theta \geq 0}(\theta x - \Lambda(\theta)) \Rightarrow \Lambda^*(y) \geq \Lambda^*(x),$$

που ήταν και το ζητούμενο. □

Ιδιότητα 12. *Έστω ότι m υπάρχει και $x \leq m$, τότε $\Lambda^*(x) = \sup_{\theta \leq 0}(\theta x - \Lambda(\theta))$.*

Απόδειξη. Η απόδειξη είναι όμοια με την ιδιότητα (10). □

Ιδιότητα 13. *Εάν m υπάρχει, τότε η Λ^* είναι φθίνουσα στο διάστημα $(-\infty, m]$.*

Απόδειξη. Η απόδειξη είναι όμοια με την ιδιότητα (11). □

Σχόλιο 8. *(α) Έχουμε αποδείξει ότι όταν $m \in (-\infty, \infty)$ η Λ^* είναι φθίνουσα στο διάστημα $(-\infty, m]$ και αύξουσα στο διάστημα $[m, \infty)$ με το m να είναι σημείο ελαχίστου. Επιπλέον έχουμε αποδείξει ότι $\Lambda^*(m) = 0$, επομένως $\inf_{x \in \mathbb{R}} \Lambda^*(x) = \sup_{x \in \mathbb{R}} \Lambda^*(x) = \Lambda^*(m) = 0$.*

(β) $m = -\infty \Rightarrow \Lambda^*(a) \searrow 0$ καθώς $a \searrow 0$.

Απόδειξη. Για κάθε $\theta \geq 0$ και $x \geq a$ έχουμε

$$\theta(x - a) \geq 0 \Rightarrow e^{\theta(x-a)} \geq 1.$$

Επομένως,

$$\mu[a, \infty) = \int_{x \geq a} d\mu(x) \leq \int_{x \geq a} e^{\theta(x-a)} d\mu(x) \leq \int_{x \in \mathbb{R}} e^{\theta(x-a)} d\mu(x).$$

Συνεπώς,

$$\Lambda^*(a) = \sup_{\theta \geq 0}(\theta a - \Lambda(\theta)) = -\inf_{\theta \geq 0}(\Lambda(\theta) - \theta a) = -\inf_{\theta \geq 0}[\log \int_{\mathbb{R}} e^{\theta x} d\mu(x) - \theta a] =$$

$$= -\inf_{\theta \geq 0}[\log(\int_{\mathbb{R}} e^{\theta x} d\mu(x) - e^{\theta a})] = -\inf_{\theta \geq 0}[\log(\int_{\mathbb{R}} e^{\theta(x-a)} d\mu(x)] \leq -\mu[a, \infty).$$

Παίρνοντας όριο καθώς $a \to -\infty$ έχουμε $\Lambda^*(a) \searrow 0$. $\square$

(γ) $m = +\infty \Rightarrow \Lambda^*(a) \searrow 0$ *καθώς* $a \nearrow 0$.

Απόδειξη. Η απόδειξη είναι όμοια με το (β). $\square$

Συνεχίζουμε τώρα με την παρουσίαση των ιδιοτήτων.

Ιδιότητα 14. $\inf_{x\in\mathbb{R}} \Lambda^*(x) = 0$.

Απόδειξη. Εφόσον $\Lambda^* \geq 0$ στις περιπτώσεις $m \in (-\infty, \infty)$, $\quad m = -\infty$, $\quad m = +\infty$ το ζητούμενο προκύπτει από τα (α),(β),(γ) του παραπάνω σχολίου αντίστοιχα. Εάν το m δεν υπάρχει, τότε

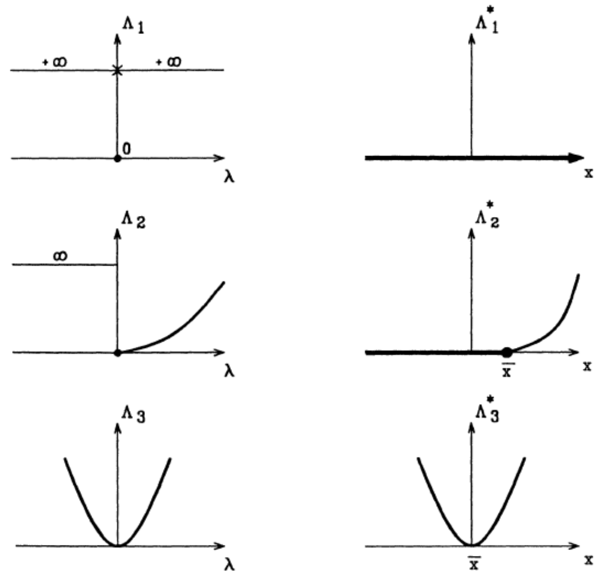
$$\Lambda(\theta) = \infty, \quad \forall \theta \neq 0.$$

Επομένως σε αυτήν την περίπτωση για κάθε $x \in \mathbb{R}$,

$$\Lambda^*(x) = \sup_{\theta \geq 0}(\theta x - \Lambda(\theta)) = 0 - \Lambda(0),$$

άρα το ζητούμενο ισχύει κατά τετριμμένο τρόπο. $\square$

Γραφικά αυτό που έχουμε δείξει ως τώρα,



Σχόλιο 9. *Όταν το m υπάρχει μπορούμε να σκεφτόμαστε το Λ^* ως ένα μέτρο του πόσο μακριά είναι το x από το m. Επομένως όσο πιο μακριά είναι το $x \geq m$ από το m, τόσο πιο γρήγορα συμβαίνει η σύγκλιση*

$$\mathbb{P}(\bar{X}_n > x) \leq e^{-n\Lambda^*(x)} \to 0, \quad n \to \infty.$$

Πιο γενικά, εάν $m \in (-\infty, \infty)$ και $m \notin \bar{B}$, όπου $B \subset \mathbb{R}$ σύνολο Borel, τέτοιο ώστε

$$\inf_{x \in \bar{B}} \Lambda^*(x) = \inf_{x \in B^o} \Lambda^*(x) = \inf_{x \in B} \Lambda^*(x),$$

τότε η εκτίμηση του θεωρήματος Cramér δίνει,

$$\mathbb{P}(\bar{X}_n \in B) \approx e^{-n \inf_{x\in B} \Lambda^*(x)},$$

αγνοώντας ενδεχομένως έναν πολυωνυμικό όρο, δηλαδή το $\inf_{x\in B} \Lambda^*(x)$ είναι ένα μέτρο του πόσο μακριά είναι το x από το B, επομένως όσο πιο μεγάλη είναι αυτή η απόσταση τόσο μεγαλύτερος είναι αυτός ο όρος και κατά συνέπεια τόσο πιο γρήγορα συγκλίνει η πιθανότητα $\mathbb{P}(\bar{X}_n \in B) \to 0$. Υπό αυτή τη σκοπιά το απότελεσμα του θεωρήματος Cramér είναι εξαιρετικά λογικό.

Ιδιότητα 15. *Έστω $D_\Lambda = L = \{\theta \in \mathbb{R} : \Lambda(\theta) < \infty\}$ και έστω $D_\Lambda^o \neq \emptyset$, τότε η Λ είναι διαφορίσιμη πάνω στο D_Λ^o και*

$$\Lambda'(\theta) = e^{-\Lambda(\theta)} \int_{\mathbb{R}} x e^{\theta x} d\mu(x), \quad \theta \in D_\Lambda^o$$

Απόδειξη. Έχουμε

$$\Lambda(\theta) = \log \int_{\mathbb{R}} e^{\theta x} d\mu(x) \Rightarrow \Lambda'(\theta) = \frac{\partial}{\partial \theta} \log \int_{\mathbb{R}} e^{\theta x} d\mu(x) =$$

$$= \frac{1}{\int_{\mathbb{R}} e^{\theta x} d\mu(x)} \log \frac{\partial}{\partial \theta} \int_{\mathbb{R}} e^{\theta x} d\mu(x) = e^{-\Lambda(\theta)} \log \frac{\partial}{\partial \theta} \int_{\mathbb{R}} e^{\theta x} d\mu(x). \tag{1.9}$$

Επομένως για την ολοκλήρωση της απόδειξης μένει να δείξουμε ότι είναι ορθό να εναλλάξουμε τη σειρά παραγώγισης και ολοκλήρωσης στη σχέση (1.9). Πράγματι, έστω η οικογένεια συναρτήσεων

$$f_\epsilon(x) = \frac{e^{(\theta+\epsilon)x} - e^{\theta x}}{\epsilon}, \quad \epsilon \in \mathbb{R},$$

η οποία συγκλίνει κατά σημείο

$$f_\epsilon(x) = \frac{e^{(\theta+\epsilon)x} - e^{\theta x}}{\epsilon} \to x e^{-\theta x}, \quad \epsilon \to 0.$$

Επιπλέον, έστω $\delta > 0$ κατάλληλα μικρό, ορίζουμε τη συνάρτηση

$$g(x) = \frac{e^{\theta x}(e^{\delta |x|} - 1)}{\delta},$$

τότε για $\epsilon \in (-\delta, \delta)$

$$|f_\epsilon| \leq g \quad \& \quad \int_{\mathbb{R}} g(x) d\mu(x) < \infty,$$

συνεπώς λόγω του θεωρήματος κυριαρχημένης σύγκλισης η σχέση (1.9) θα μας δώσει

$$\Lambda'(\theta) = e^{-\Lambda(\theta)} \log \int_{\mathbb{R}} \frac{\partial}{\partial \theta} e^{\theta x} d\mu(x) = e^{-\Lambda(\theta)} \int_{\mathbb{R}} x e^{\theta} d\mu(x).$$

□

Ιδιότητα 16. *Έστω $D_\Lambda \neq \{0\}$, εάν $\eta \in D_\Lambda^o$ και $\Lambda'(\eta) = y \in \mathbb{R}$, τότε $\Lambda^*(y) = \eta y - \Lambda(\eta)$, δηλαδή το supremum πιάνεται στο σημείο η.*

Απόδειξη. Έστω η συνάρτηση

$$g(\theta) = \theta y - \Lambda(\theta), \quad \theta \in D_\Lambda^o \quad \Rightarrow -g(\theta) = \Lambda(\theta) - \theta y.$$

Θα δείξουμε ότι η $-g$ είναι κυρτή και συνεπώς η g είναι κοίλη. Πράγματι, έστω $c \in [0,1]$ και $\theta, \psi \in D_\Lambda^o$, επιπλέον εάν λάβουμε υπ΄ όψιν ότι η Λ είναι κυρτή θα έχουμε

$$-g(c\theta + (1-c)\psi) \leq c\Lambda(\theta) + (1-c)\Lambda(\psi) - c\theta y - (1-c)\psi y =$$

$$= c(-g)(\theta) + (1-c)(-g)(\psi),$$

επομένως η g είναι κοίλη. Επιπλέον,

$$g'(\eta) = y - \Lambda'(\eta) = 0,$$

οπότε τώρα έχουμε,

$$\Lambda^*(y) = \sup_{\theta \in \mathbb{R}} (\theta y - \Lambda(\theta)) = \max_{\theta \in D_\Lambda^o} g(\theta) = g(\eta).$$

□

Πλέον, έχουμε ολοκληρώσει την παρουσίαση και την απόδειξη των ιδιοτήτων (1)-(16) και είμαστε σε θέση να τις χρησιμοποιήσουμε προκειμένου να αποδείξουμε το θεώρημα Cramér.

1.4.2 Η Απόδειξη του Θεωρήματος Cramér

Απόδειξη. (θεωρήματος Cramér) (α) Έστω $F \subset \mathbb{R}$ μη κενό κλειστό σύνολο. Το ζητούμενο ισχύει τετριμμένα όταν $\inf_{x\in F} \Lambda^*(x) = 0$. Επομένως εξετάζουμε την περίπτωση που $\inf_{x\in F} \Lambda^*(x) > 0$. Πιο συγκεκριμένα, αυτό σημαίνει ότι το m υπάρχει, (διαφορετικά από την ιδιότητα (15) θα είχαμε ότι $\Lambda^* \equiv 0$).

Επιπλέον, από το (β) του Σχολίου της σελίδας 15,

$$m = -\infty \Rightarrow F \quad \text{έχει} \quad \min,$$

έστω β το ελάχιστο του F.

Ομοίως, από το (γ) του Σχολίου της σελίδας 16,

$$m = +\infty \Rightarrow F \quad \text{έχει} \quad \max,$$

έστω α το μέγιστο του F.

Στην περίπτωση που

$$-\infty < m < +\infty \Rightarrow \Lambda^*(m) = 0 \Rightarrow m \notin F,$$

τότε ορίζουμε τα σύνολα

$$\mathcal{A} = \{x \in F, x < m\} \quad \& \quad \mathcal{B} = \{x \in F, x > m\}$$

και παίρνουμε τα α,β να είναι

$$a = \sup \mathcal{A} = \max \mathcal{A} \quad \& \quad \beta = \inf \mathcal{B} = \min \mathcal{B}.$$

Θα μελετήσουμε τώρα όλες τις δυνατές περιπτώσεις που μπορούν να εμφανιστούν, ξεκινώντας από την απλούστερη, δηλαδή $F = [x, +\infty), \quad x > m$ τότε για κάθε $\theta \geq 0$ με εφαρμογή της ανισότητας Chebyshev έχουμε

$$\mathbb{P}(\bar{X}_n \in F) = \mathbb{E}[\mathbb{1}_{(\bar{X}_n - x \geq 0)}] \leq \mathbb{E}[e^{n\theta(\bar{X}_n - x)}\mathbb{1}_{(\bar{X}_n - x \geq 0)}] \leq \mathbb{E}[e^{n\theta(\bar{X}_n - x)}] =$$

$$= e^{-n\theta x} \prod_{i=1}^{n} \mathbb{E}[e^{\theta X_i}] = e^{-n(\theta x - \Lambda(\theta))} \tag{1.10}$$

Επομένως από τη σχέση (1.10) και την ιδιότητα (10) έχουμε

$$\mathbb{P}(\bar{X}_n \in [x, \infty)) \leq \inf_{\theta \geq 0} e^{-n(\theta x - \Lambda(\theta))} = e^{-n \sup_{\theta \geq 0}(\theta x - \Lambda(\theta))} = e^{-n\Lambda^*(x)}, \tag{1.11}$$

που είναι το ζητούμενο. Η απόδειξη για τη συμμετρική περίπτωση, δηλαδή για την περίπτωση $F = (-\infty, x], x < m$ είναι όμοια και μας δίνει,

$$\mathbb{P}(\bar{X}_n \in (-\infty, x]) = e^{-n\Lambda^*(x)}. \tag{1.12}$$

Ας υποθέσουμε τώρα ότι το F είναι οποιοδήποτε κλειστό σύνολο, έχουμε τις εξής περιπτώσεις.

Περίπτωση I. *Για $m = +\infty$ έχουμε $a \in (-\infty, \infty)$ και επομένως $F \subset (-\infty, a]$ με $a < m$ και η Λ^* είναι φθίνουσα στο διάστημα $(\infty, m]$ συνεπώς,*

$$- \inf_{y \in F} \Lambda^*(y) = -\Lambda^*(a) \Rightarrow \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in F) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in (-\infty, a]) \leq$$

$$\leq -\Lambda^*(a) = - \inf_{y \in F} \Lambda^*(y),$$

η απόδειξη για τη συμμετρική περίπτωση $m = -\infty$ είναι όμοια.

Περίπτωση II. *Για $-\infty < m < \infty$, εφόσον το m είναι πεπερασμένο θα έχουμε $\Lambda^*(m) = 0$ και εξ΄ υποθέσεως έχουμε $\inf_{x \in F} \Lambda(x) = 0$, άρα αναγκαστικά θα έχουμε $m \in F^c$.*

Έστω (x_-, x_+) να είναι η ένωση όλων των ανοιχτών διαστημάτων $(\gamma, \delta) \subset F^c$ τα οποία περιέχουν το m. Τότε δεν είναι δύσκολο να δούμε ότι θα πρέπει $x_- = \alpha, x_+ = \beta$ και σίγουρα ένα από τα x_-, x_+ είναι πεπερασμένα γιατί το F είναι μη κενό.

Εάν x_- είναι πεπερασμένο, τότε $x_- \in F$, συνεπώς $\Lambda^*(x_-) \leq \inf_{x \in F} \Lambda^*(x)$. Ομοίως $\Lambda^*(x_+) \leq \inf_{x \in F} \Lambda^*(x)$ όταν x_+ είναι πεπερασμένο. Εφαρμόζοντας τη σχέση (1.11) για $x = x_+$ και αντίστοιχα τη σχέση (1.12) $x = x_-$ και λαμβάνοντας υπ΄όψιν την ανισότητα Boole έχουμε,

$$\mathbb{P}(\bar{X}_n \in F) \leq \mathbb{P}(\bar{X}_n \in (-\infty, x_-)) + \mathbb{P}(\bar{X}_n \in (x_+, +\infty)) \leq 2e^{-n \inf_{x \in F} \Lambda^*(x)},$$

και το ζητούμενο άνω φράγμα προκύπτει άμεσα εάν θεωρήσουμε το κανονικοποιημένο λογαριθμικό όριο στην παραπάνω σχέση.

(β) Έστω $U \subset \mathbb{R}$ μη κενό ανοιχτό σύνολο. Το επιχείρημα που θα χρησιμοποιήσουμε για την απόδειξη του κάτω φράγματος χρησιμοποιεί την τεχνική της αλλαγής μέτρου. Πιο συγκεκριμένα για κάθε μέτρο ν στο $\mathbb{R}$ θα συμβολίζουμε με $\mathbb{P}_\nu$ το μέτρο πιθανότητας κάτω από το οποίο οι τυχαίες μεταβλητές $X_1, ..., X_n \sim \nu$, (για ότι έχουμε αποδείξει έως τώρα έχουμε χρησιμοποιήσει τη σύμβαση $\mathbb{P} = \mathbb{P}_\mu$). Αυτό σημαίνει ότι για κάθε σύνολο Borel $B \subset \mathbb{R}$, έχουμε

$$\nu(B) = \mathbb{P}_\nu(X_1 \in B).$$

Αντιστοίχως τις συναρτήσεις Λ, Λ^* θα τις συμβολίζουμε με $\Lambda_\nu, \Lambda^*_\nu$. Αυτό σημαίνει ότι

$$\Lambda_\nu(\theta) = \log \int_{\mathbb{R}} e^{\theta x} d\nu(x), \theta \in \mathbb{R} \quad \& \quad \Lambda^*_\nu(x) = \sup_{x \in \mathbb{R}} (\theta x - \Lambda_\nu(\theta)), x \in \mathbb{R}.$$

Έστω $M(\mathbb{R}) :=$ ο χώρος όλων των μέτρων πιθανότητας στο $\mathbb{R}$, θα αποδείξουμε ότι για κάθε $\mu \in M(\mathbb{R}), \delta > 0$ ισχύει

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_\mu(-\delta < \bar{X}_n < \delta) \geq -\Lambda^*_\mu(0). \tag{1.13}$$

Πράγματι, έστω $x \in U$, εφόσον το U είναι ανοιχτό $\exists \delta > 0$ τέτοιο ώστε $(x-\delta, x+\delta) \subset U$. Επιπλέον για $B \subset \mathbb{R}$ σύνολο Borel, ορίζουμε το σύνολο $x+B := \{x+b, b \in B\}$ και το μέτρο τ πάνω στο $\mathbb{R}$ να είναι τέτοιο ώστε για $B \subset \mathbb{R}$ σύνολο Borel

$$\tau(B) = \mu(x+B).$$

Σχόλιο 10. *Το μέτρο τ είναι η κατανομή της τυχαίας μεταβλήτης $X - x$ όταν η $X \sim \mu$.*

Παρατηρούμε τώρα ότι εάν εφαρμόσουμε την ανισότητα (1.13) για το μέτρο τ θα έχουμε,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\tau(-\delta < \bar{X}_n < \delta) \geq -\Lambda_\tau^*(0), \tag{1.14}$$

επιπλέον ισχύει

$$\mathbb{P}_\tau(-\delta < \bar{X}_n < \delta) = \mathbb{P}_\mu(-\delta + x < \bar{X}_n < \delta + x),$$

ενώ από την άλλη $\forall \theta \in \mathbb{R}$, έχουμε

$$\Lambda_\tau(\theta) = \log \mathbb{E}_\mu(e^{\theta(X-x)}) = -\theta x + \log \mathbb{E}_\mu(e^{\theta x}) = \Lambda_\mu(\theta) - \theta x.$$

Επομένως, θα ισχύει

$$\Lambda_\tau^*(0) = \sup_{\theta\in\mathbb{R}}(0 - \Lambda_\tau(\theta)) = \sup_{\theta\in\mathbb{R}}(\theta x - \Lambda_\mu(\theta)) = \Lambda_\mu^*(x).$$

Άρα από τη σχέση (1.14), αντικαθιστούμε και παίρνουμε,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(x - \delta < \bar{X}_n < x + \delta) \geq -\Lambda_\mu^*(x),$$

συνεπώς

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in U) \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(x - \delta < \bar{X}_n < x + \delta) \geq -\Lambda_\mu^*(x),$$

άρα

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in U) \geq \sup_{x\in U}(-\Lambda_\mu^*(x)) \geq -\inf_{x\in U} \Lambda_\mu^*(x),$$

το οποίο είναι και το ζητούμενο κάτω φράγμα, άρα για την ολοκλήρωση της απόδειξης μας απομένει να δείξουμε την ορθότητα της σχέσης (1.14).

Πρέπει να αποδείξουμε ότι

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(-\delta < \bar{X}_n < \delta) \geq -\Lambda_\mu^*(0) = -\inf_{\theta\in\mathbb{R}} \Lambda_\mu(\theta),$$

πιο συγκεκριμένα πρέπει να παρακολουθήσουμε τη συμπεριφορά του infimum καθώς το θ διατρέχει τους πραγματικούς αριθμούς. Η απόδειξη είναι μακροσκελής και για αυτό θα την κάνουμε κατά βήματα ξεκινώντας από την απλούστερη περίπτωση και προχωρόντας προς τις πιο σύνθετες.

Περίπτωση I. *Το μ έχει στήριγμα το διάστημα* $[0,\infty)$*, δηλαδή* $\mu((-\infty,0)) = 0$.

Στην περίπτωση αυτή

$$\Lambda(\theta) = \log \int_{[0,+\infty)} e^{\theta x} d\mu(x), \quad \theta \in \mathbb{R},$$

τότε δεν είναι δύσκολο να δούμε ότι η Λ είναι αύξουσα στο $\mathbb{R}$ και κατά συνέπεια μπορούμε να βρούμε το infimum παίρνοντας $\theta \to -\infty$. Θα πρέπει να σημειώσουμε ότι για $\theta \to -\infty$

$$e^{\theta x} = \begin{cases} 0, & \text{για} \quad x = 0 \\ 1, & \text{για} \quad x > 0 \end{cases}.$$

Με βάση αυτό από το θεώρημα κυριαρχημένης σύγκλισης προκύπτει ότι

$$\Lambda_\mu(\theta) \searrow \log \mu(\{0\}), \theta \to -\infty \Rightarrow \inf_{\theta\in\mathbb{R}} \Lambda_\mu(\theta) = \log \mu(\{0\}).$$

Τώρα παρατηρούμε ότι το ενδεχόμενο $(X_1 = X_2 = ... = X_n = 0)$ συνεπάγεται κατά τετριμμένο τρόπο το ενδεχόμενο $\bar{X}_n \in (-\delta, \delta)$, συνεπώς θα έχουμε

$$\mathbb{P}_\mu(-\delta < \bar{X}_n < \delta) \geq \mathbb{P}_\mu(X_1 = ... = X_n = 0) = (\mu(\{0\}))^n \Rightarrow$$

$$\Rightarrow \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(-\delta < \bar{X}_n < \delta) \geq \mu(\{0\}) = \inf_{\theta\in\mathbb{R}} \Lambda_\mu(\theta),$$

που είναι και το ζητούμενο. Η απόδειξη για την περίπτωση που το μ έχει στήριγμα το διάστημα $(-\infty, 0]$ είναι όμοια.

Περίπτωση II. *Έχουμε* $\mu((-\infty,0)) > 0, \mu((0,+\infty)) > 0$ *και το μ έχει συμπαγές στήριγμα.*

Σε αυτήν την περίπτωση η Λ_μ θα είναι πεπερασμένη $\forall\theta \in \mathbb{R}$, άρα $D_\Lambda^o = \mathbb{R}$ οπότε από την ιδιότητα (15) η Λ_μ θα είναι διαφορίσιμη και κατ΄ επέκταση συνεχής σε όλο το $\mathbb{R}$. Επιπλέον η m πρέπει να υπάρχει γιατί το μ έχει συμπαγές στήριγμα και άρα είναι φραγμένο, οπότε

$$\Lambda_\mu(\theta) = \log \mathbb{E}(e^{\theta X_1}) \geq \mathbb{E}(\log e^{\theta X_1}) = \theta m \Rightarrow$$

$$\Lambda_\mu(\theta) \to +\infty, \quad \text{καθώς} \quad |\theta| \to +\infty.$$

Επιπλέον η Λ_μ είναι κυρτή, άρα $\exists \eta \in \mathbb{R}$ τέτοιο ώστε,

$$\Lambda_\mu(\eta) = \inf_{\theta \in \mathbb{R}} \Lambda_\mu(\theta) \quad \& \quad \Lambda'_\mu(\eta) = 0.$$

Σταθεροποιούμε ένα τέτοιο η και από την ιδιότητα (15) θα έχουμε

$$0 = \Lambda'_\mu(\eta) = e^{-\Lambda_\mu(\eta)} \int_{\mathbb{R}} x e^{\eta x} d\mu(x) \Rightarrow$$

$$\int_{\mathbb{R}} x e^{(\eta x - \Lambda_\mu(\eta))} d\mu(x) = 0. \tag{1.15}$$

Σχόλιο 11. *Το μέτρο $\tilde{\mu}$ που ορίζεται από τη σχέση $d\tilde{\mu} = e^{(\eta x - \Lambda_\mu(x))} d\mu(x)$, δηλαδή*

$$\forall B \subset \mathbb{R} \quad Borel \quad \tilde{\mu}(B) = \int_B e^{(\eta x - \Lambda_\mu(x))} d\mu(x),$$

είναι μέτρο πιθανότητας στο $\mathbb{R}$ διότι

$$\tilde{\mu}(\mathbb{R}) = e^{-\Lambda_\mu(\eta)} e^{\Lambda_\mu(\eta)} = 1.$$

Από τη σχέση (1.15) θα έχουμε ότι το μέτρο $\tilde{\mu}$ έχει μέση τιμή 0. Έστω τώρα ότι διαλέγουμε ένα $\epsilon \in (0, \delta)$, όπου το δ είναι αυτό που έχουμε σταθεροποιήσει από την αρχή της απόδειξης, τότε παρατηρούμε ότι

$$\mathbb{P}_\mu(\bar{X}_n \in (-\epsilon, \epsilon)) = \int_S \prod_{i=1}^n d\mu(x_i), \quad \text{όπου} \quad S = \{(x_1, ..., x_n) \in \mathbb{R}^n : \left| \sum_{i=1}^n x_i \right| < n\epsilon\}.$$

Επιπλέον παρατηρούμε ότι

$$(x_1, ..., x_n) \in S \Rightarrow \left| \sum_{i=1}^n x_i \right| < n\epsilon \Rightarrow n \sum_{i=1}^n x_i \leq \left| n \sum_{i=1}^n x_i \right| < n\epsilon|n| \Rightarrow$$

$$\Rightarrow -n\epsilon|n| + n \sum_{i=1}^n x_i \leq 0 \Rightarrow e^{(-n\epsilon|n| + n\sum_{i=1}^n x_i)} \leq 1.$$

Επομένως

$$\mathbb{P}_\mu(\bar{X}_n \in (-\epsilon, \epsilon)) \leq \int_S e^{(-n\epsilon|n| + n\sum_{i=1}^n x_i)} \prod_{i=1}^n d\mu(x_i) = e^{-n\epsilon|n|} \int_S e^{n\sum_{i=1}^n} \prod_{i=1}^n d\mu(x_i). \tag{1.16}$$

Τώρα έχουμε

$$d\tilde{\mu} = e^{\eta x - \Lambda_\mu(\eta)} d\mu \Rightarrow e^{\Lambda_\mu(\eta)} d\tilde{\mu} = e^{\eta x} d\mu \Rightarrow e^{n\sum_{i=1}^n x_i} \prod_{i=1}^n d\mu(x_i) = e^{\eta \Lambda_\mu(n)} \prod_{i=1}^n d\tilde{\mu}(x_i).$$

Άρα, υπό αυτό το πρίσμα η σχέση (1.16) θα μας δώσει

$$\mathbb{P}_\mu(X_n \in (-\epsilon, \epsilon)) \geq e^{-n(\epsilon|\eta|) - \Lambda_\mu(\eta))} \int_S \prod_{i=1}^n d\tilde{\mu}(x_i) = e^{-n(\epsilon|\eta|) - \Lambda_\mu(\eta))} \mathbb{P}_\mu(X_n \in (-\epsilon, \epsilon)).$$

Εφόσον $\delta > \epsilon$, θα έχουμε

$$\frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in (-\delta, \delta)) \geq \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in (-\epsilon, \epsilon)) \geq$$

$$\geq -\epsilon|\eta| + \Lambda_\mu(\eta) + \frac{1}{n} \log \mathbb{P}_{\tilde{\mu}}(\bar{X}_n \in (-\epsilon, \epsilon)). \tag{1.17}$$

Επιπλέον το $\tilde{\mu}$ έχει μέση τιμή 0, επομένως ΑΝΜΑ $\Rightarrow \bar{X}_n \to 0, n \to \infty$ υπό το $\mathbb{P}_{\tilde{\mu}}$, συνεπώς

$$\mathbb{P}_{\tilde{\mu}}(\bar{X}_n \in (-\epsilon, \epsilon)) \to 1, \quad n \to \infty$$

Άρα, λαμβάνοντας το $\liminf$ για $n \to \infty$ στη σχέση (1.17) θα έχουμε

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in (-\delta, \delta)) \geq -\epsilon|\eta| + \Lambda_\mu(\eta),$$

και εφόσον το ϵ που επιλέξαμε ήταν αυθαίρετο έπεται ότι,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in (-\delta, \delta)) \geq \Lambda_\mu(\eta) = \inf_{\theta \in \mathbb{R}} \Lambda(\theta),$$

που αποδεικνύει το ζητούμενο.

Περίπτωση III. *Έχουμε $\mu((-\infty, 0)) > 0, \mu((-\infty, 0)) > 0$ και το στήριγμα του μ είναι μη φραγμένο.*

Έστω ένα αρκετά μεγάλο $M \in \mathbb{N}$, τέτοιο ώστε $\mu((-M, 0)) > 0, \mu((M, 0)) > 0$, η ύπαρξη του M προκύπτει άμεσα από τις υποθέσεις που έχουμε κάνει. Ορίζουμε τώρα το μέτρο

$$\mu_M(B) = \frac{\mu(B \cap [-M, M])}{C_M}, \text{όπου} \quad C_M = \mu([-M, M]).$$

Σχόλιο 12. *Ουσιαστικά το μ_M είναι ο περιορισμός του μ πάνω στο διάστημα [-M,M] με μια επιπλέον κανονικοποίηση ώστε να αποτελεί μέτρο πιθανότητας.*

Παρατηρούμε τώρα ότι πάνω στο σύνολο [-M,M] θα έχουμε

$$\mu_M = \frac{1}{C_M}\mu \Rightarrow \frac{d\mu_M}{d\mu} = \frac{1}{C_M} \Rightarrow d\mu = C_M d\mu_M.$$

Ορίζουμε τώρα το σύνολο

$$S' = \{(x_1, ..., x_n) \in \mathbb{R}^n : \left|\sum_{i=1}^{n} x_i\right| < n\delta\}$$

και θα έχουμε

$$\mathbb{P}_\mu[\bar{X}_n \in (-\delta,\delta)] = \int_{S'} \prod_{i=1}^{n} d\mu(x_i) \geq \int_{S'\cap[-M,M]^n} \prod_{i=1}^{n} d\mu(x_i) =$$

$$= \int_{S'\cap[-M,M]^n} (C_M)^n \prod_{i=1}^{n} d\mu_M(x_i) = (C_M)^n \mathbb{P}_{\mu_M}[\bar{X}_n \in (\delta,\delta)].$$

Ορίζουμε $\Lambda_M(\theta) = \log\int_{\mathbb{R}} e^{\theta x} d\mu_M(x), \theta \in \mathbb{R}$, τότε έχουμε

$$\forall\theta \in \mathbb{R}, \quad \Lambda_M(\theta) = \log\int_{[-M,M]} e^{\theta x} d\mu_M(x) = \log[\frac{1}{C_M}\int_{[-M,M]} e^{\theta x} d\mu(x)] - \log C_M.$$

Επιπλέον, εάν ορίσουμε $\tilde{\Lambda}_M(\theta) := \log[\frac{1}{C_M}\int_{[-M,M]} e^{\theta x} d\mu(x)]$, η παραπάνω σχέση θα μας δώσει,

$$\tilde{\Lambda}_M(\theta) = \Lambda_M(\theta) + \log(M), \theta \in \mathbb{R}.$$

Ουσιαστικά η $\tilde{\Lambda}_M$ είναι η *΄μη κανονικοποιημένη έκδοση΄* της Λ_M. Στη συνέχεια θα έχουμε

$$\liminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}_\mu[\bar{X}_n \in (-\delta,\delta)] \geq \log C_M + \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_{\mu_M}[\bar{X}_n \in (-\delta,\delta)] \geq$$

$$\geq \log C_M + \inf_{\theta\in\mathbb{R}} \Lambda_M(\theta) = \inf_{\theta\in\mathbb{R}} \tilde{\Lambda}_M(\theta). \tag{1.18}$$

Τώρα εφόσον η $\tilde{\Lambda}_M$ θα αυξάνει καθώς $M \to \infty$ η $\{\inf_{\theta\in\mathbb{R}} \tilde{\Lambda}_M(\theta)\}$ θα είναι μια αύξουσα ακολουθία. Έστω, ότι

$$\{\inf_{\theta\in\mathbb{R}} \tilde{\Lambda}_M(\theta)\} \nearrow \alpha.$$

Επομένως εάν πάρουμε $M \to \infty$ στη σχέση (1.18) θα έχουμε

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu[\bar{X}_n \in (-\delta, \delta)] \geq \alpha,$$

για την ολοκλήρωση της απόδειξης τώρα θα πρέπει να δείξουμε ότι $\alpha \geq \inf_{\theta\in\mathbb{R}} \Lambda_\mu(\theta)$, για την απόδειξη αυτού θα πρέπει να δείξουμε ότι υπάρχει ένα $\theta^* \in \mathbb{R}$ τέτοιο ώστε $\alpha \geq \Lambda_\mu(\theta^*)$, θα το εξασφαλίσουμε αυτό με τη βοήθεια ενός επιχειρήματος διαγωνοποίησης (Cantor).

Προτού προχωρήσουμε παρακάτω είναι άξιο παρατήρησης το γεγονός ότι το α θα είναι αναγκαστικά πεπερασμένο. Για να το δούμε αυτό παρατηρούμε το εξής,

$$\lim_{|\theta|\to\infty} \tilde{\Lambda}_M(\theta) = \infty \Rightarrow \inf_{\theta\in\mathbb{R}} \tilde{\Lambda}_M(\theta) \in (-\infty, \infty) \Rightarrow \alpha > -\infty.$$

Επιπλέον, θα ισχύει

$$\tilde{\Lambda}_\mu(0) = \log(\mu([-M, M])) \leq 0 \Rightarrow \inf_{\theta\in\mathbb{R}} \tilde{\Lambda}_M(\theta) \leq 0, \quad \forall M,$$

συνεπώς παίρνοντας $M \to \infty$, έχουμε

$$\alpha < 0 \Rightarrow \alpha < \infty,$$

που είναι και το ζητούμενο.

Θα δείξουμε τώρα ότι υπάρχει ένα $\theta^* \in \mathbb{R}$ τέτοιο ώστε

$$\alpha \geq \Lambda_\mu(\theta^*) = \log \int_{\mathbb{R}} e^{\theta^* x} d\mu(x).$$

Προς αυτό το σκοπό ορίζουμε το σύνολο

$$K_M := \{\theta \in \mathbb{R} : \log \int_{[-M,M]} e^{\theta x} d\mu(x)\},$$

για κάθε $M \in \mathbb{N}$ αρκετά μεγάλο, ας πούμε $M \geq M_0$ τέτοιο ώστε $\mu((0, M]), \mu([-M, 0)) > 0$.

Παρατηρούμε τώρα ότι αρκεί να δείξουμε ότι

$$\bigcap_{M\geq M_0} K_M \neq \emptyset, \tag{1.19}$$

διότι τότε μπορούμε να πάρουμε, $\theta^* \in \cap_{M\geq M_0} K_M$ το οποίο συνεπάγεται

$$\log \int_{[-M,M]} e^{\theta^* x} d\mu(x), \forall M \geq M_0,$$

και από αυτό μπορούμε να χρησιμοποιήσουμε το Θεώρημα Μονότονης Σύγκλισης για $M \to \infty$ και να λάβουμε

$$\log \int_{\mathbb{R}} e^{\theta^* x} d\mu(x) \leq \alpha,$$

το οποίο ολοκληρώνει την απόδειξη. Επομένως εάν δείξουμε την (1.19) έχουμε τελειώσει.

Παρατηρούμε ότι

$$\forall M \geq M_0, \quad \Lambda_\mu(\theta) = \log \int_{[-M,M]} e^{\theta x} d\mu(x) \to \infty, |\theta| \to \infty$$

και η Λ_M είναι συνεχής συνάρτηση και για κάθε $M \geq M_0$ το K_M είναι συμπαγές. Επιπλέον η $\{\Lambda_M\}_{M\geq M_0}$ είναι κατά σημείο αύξουσα ακολουθία συναρτήσεων και επομένως η οικογένεια $\{K_M\}_{M\geq M_0}$ είναι φθίνουσα (ως προς τη σχέση του περιέχεσθαι). Τώρα θα έχουμε,

$$\forall M \geq M_0, \quad \alpha \geq \inf_{\theta\in\mathbb{R}} \tilde{\Lambda}_M(\theta) = \min_{\theta\in\mathbb{R}} \tilde{\Lambda}_M(\theta),$$

λόγω συνέχειας, άρα $K_M \neq \emptyset \quad \forall M \geq M_0$, το οποίο από το θεώρημα Cantor συνεπάγεται

$$\bigcap_{M\geq M_0} K_M \neq \emptyset.$$

□

1.4.3 Παραδείγματα για Γνωστές Κατανομές

Στην παράγραφο αυτή θα δούμε πως υπολογίζεται ο μετασχηματισμός Fenchel-Legendre για ορισμένες γνωστές κατανομές, ο οποίος ουσιαστικά καθορίζει την ταχύτητα της σύγκλισης.

Παράδειγμα 1. *Έστω $X_1, X_2, .., X_n \sim N(m, \sigma^2)$, τότε*

$$\Lambda(\theta) = \log \mathbb{E}(e^{\theta X}) = m\theta + \frac{1}{2}\sigma^2\theta^2 \Rightarrow \Lambda^*(x) = \sup_{\theta\in\mathbb{R}}(\theta x - m\theta - \frac{1}{2}\sigma^2\theta^2),$$

και παραγωγίζοντας ως προς θ θα πάρουμε,

$$\frac{\partial}{\partial\theta}(\theta x - m\theta - \frac{1}{2}\sigma^2\theta^2) = 0 \Rightarrow \tilde{\theta} = \frac{(x-m)}{\sigma^2},$$

συνεπώς

$$\Lambda^*(x) = (\tilde{\theta}x - m\tilde{\theta} - \frac{1}{2}\sigma^2\tilde{\theta}^2) = \frac{1}{2}\frac{(x-m)^2}{\sigma^2}.$$

Άρα για $x > m$, έχουμε

$$\mathbb{P}(\bar{X}_n > nx) \approx e^{-n\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}.$$

Για $m = 0, \sigma^2 = 1$ βλέπουμε πως η εκτίμηση που παίρνουμε μέσω του μετασχηματισμού Fenchel-Legendre συμφωνεί ως προς τον εκθετικό παράγοντα με την εκτίμηση που υπολογίσαμε στην εισαγωγή.

Παράδειγμα 2. *Έστω $X_1, X_2, .., X_n \sim Exp(\frac{1}{m})$, τότε*

$$\Lambda(\theta) = \log(1 - m\theta) \Rightarrow \Lambda^*(x) = \sup_{\theta \in \mathbb{R}}(\theta x + \log(1 - m\theta)),$$

παραγωγίζοντας ως προς θ, έχουμε

$$\frac{\partial}{\partial \theta}(\theta x + \log(1 - m\theta)) = 0 \Rightarrow \tilde{\theta} = \frac{(x - m)}{xm},$$

συνεπώς

$$\Lambda^*(x) = \tilde{\theta} x + \log(1 - m\theta) = \frac{x}{m} - 1 - \log(\frac{x}{m}).$$

Άρα για $x > m$, έχουμε

$$\mathbb{P}(\bar{X}_n > nx) \approx e^{-n(\frac{x}{m} - 1 - \log(\frac{x}{m}))}.$$

Σχόλιο 13. *Βλέπουμε και στις δύο περιπτώσεις ότι επιβεβαιώνεται το θεωρητικό αποτέλεσμα που έχουμε αποδείξει πιο πάνω ότι $\Lambda^*(m) = 0$.*

Έχει ενδιαφέρον εδώ να παρατηρήσουμε ότι δεν μπορούμε πάντα να πάρουμε κάποιο χρήσιμο αποτέλεσμα για τη $\bar{X}_n$ μέσω του θεωρήματος Cramér, προς αυτό το σκοπό δίνουμε το παρακάτω παράδειγμα που $\Lambda(\theta)$ δεν υπάρχει και η θεωρία των μεγάλων αποκλίσεων δεν παράγει κάποιο χρήσιμο αποτέλεσμα.

Παράδειγμα 3. *Η κλάση των αυστηρά Levý(strict Levý) τυχαίων μεταβλητών ορίζεται από τη χαρακτηριστική εξίσωση*

$$\mathbb{E}(e^{i\theta X}) = e^{-\gamma|\theta|^\alpha}, \quad \theta \in \mathbb{R}, \gamma > 0, \alpha \in (0, 2).$$

Σε αυτήν την περίπτωση λόγω του ότι οι ουρές της συνάρτησης πυκνότητας πιθανότητας είναι της μορφής $f(x) \sim x^{-1-\alpha}, \quad |x| \to \infty$ (power law) η $\mathbb{E}(e^{\theta X})$ δεν συγκλίνει για $\theta \neq 0$ και κατά συνέπεια έχουμε

$$\Lambda(\theta) = \begin{cases} 0, & \text{για} \quad \theta = 0 \\ \infty, & \text{αλλιώς} \end{cases}$$

οπότε σε αυτήν την περίπτωση, εάν επιχειρήσουμε να κάνουμε κάποιον υπολογισμό θα έχουμε $\Lambda^* \equiv 0$ που δεν παράγει κάποια χρήσιμη εκτίμηση.

Κεφάλαιο 2

Μεγάλες Αποκλίσεις σε Χώρους Πεπερασμένης Διάστασης

Στο κεφάλαιο αυτό ο σκοπός θα είναι να γενικεύσουμε το μαθηματικό φορμαλισμό που έχουμε χρησιμοποιήσει έως τώρα, εκμεταλλευόμενοι αυτό το γεγονός θα παρουσιάσουμε και θα αποδείξουμε το θεώρημα Cramér στον $\mathbb{R}^d$, το θεώρημα Gärtner-Ellis, το λήμμα Varadhan και τέλος θα εξετάσουμε πως προκύπτει το θεώρημα Sanov.

2.1 Η Αρχή των Μεγάλων Αποκλίσεων

Προκειμένου να εκμεταλλευτούμε τη γενικότητα του συμβολισμού, θεωρούμε έναν πολωνικό χώρο $\mathbb{S}$, δηλαδή έναν πλήρη και διαχωρίσιμο μετρικό χώρο και $\mathcal{B}$ να είναι η σ-άλγεβρα Borel του. Επομένως το ζεύγος $(\mathbb{S}, \mathcal{B})$ είναι ένας μετρήσιμος χώρος, συνεπώς μπορούμε να ορίσουμε πάνω του μια οικογένεια μέτρων πιθανότητας $\{\mu_\epsilon\}$, όπου το ϵ μπορεί να ανήκει σε ένα οποιοδήποτε σύνολο δεικτών (είτε αριθμήσιμο είτε υπεραριθμήσιμο).

Σχόλιο 14. *Εν γένει ο $\mathbb{S}$ μπορεί να είναι απειροδιάστατος, για τους σκοπούς αυτής της εργασίας όμως θα περιοριστούμε σε χώρους πεπερασμένης διάστασης. Για το λόγο αυτό διατυπώνουμε τα προαπαιτούμενα για τη γενική περίπτωση σε αυτήν την εισαγωγική παράγραφο και στις επόμενες παραγράφους θα τα χρησιμοποιήσουμε συγκεκριμένα για τον $\mathbb{R}^d$.*

Η αρχή των μεγάλων αποκλίσεων μας παρέχει έναν χαρακτηρισμό της οριακής συμπεριφοράς της $\{\mu_\epsilon\}$ για $\epsilon \longrightarrow 0$ ως προς μια συνάρτηση ταχύτητας I. Θα δούμε ότι αυτός χαρακτηρισμός στην ουσία γίνεται μέσω ασυμπτωτικών εκθετικών φραγμάτων, ακριβώς όπως στην περίπτωση του θεωρήματος Cramér που έχουμε ήδη μελετήσει.

Πριν περάσουμε στη διατύπωση της αρχής των μεγάλων αποκλίσεων θα δώσουμε επίσημα πλέον τους ορισμούς για ορισμένες έννοιες που χρησιμοποιήσαμε στο προηγούμενο κεφάλαιο.

Ορισμός 4. *Μια απεικόνιση I: $\mathbb{S} \to [0,\infty]$ θα λέγεται συνάρτηση ταχύτητας εάν είναι κάτω ημισυνεχής (lower semicontinuous), δηλαδή $\forall \alpha > 0$ το σύνολο $\Psi_I(\alpha) = \{x : I(x) \leq \alpha\}$ είναι κλειστό υποσύνολο του $\mathbb{S}$. Θα ονομάζουμε τη I καλή εάν $\forall a > 0$ το σύνολο $\Psi_I(\alpha) = \{x : I(x) \leq \alpha\}$ είναι συμπαγές.*

Ορισμός 5 (Large Deviation Principle)**.** *Θα λέμε ότι η οικογένεια $\{\mu_\epsilon\}$ ικανοποιεί την Αρχή Μεγάλων Αποκλίσεων με συνάρτηση ταχύτητας I εάν και μόνο εάν $\forall B \in \mathcal{B}$ ισχύει:*

$$-\inf_{x \in B^o} I(x) \leq \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \leq \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \leq -\inf_{x \in \bar{B}} I(x). \tag{2.1}$$

Σχόλιο 15. *Από εδώ και στο εξής με τη συντομογραφία LDP θα αναφερόμαστε στην Αρχή των Μεγάλων Αποκλίσεων (Large Deviation Principle).*

Είναι προφανές ότι εάν $\{\mu_\epsilon\}$ ικανοποιεί την LDP και $B \in \mathcal{B}$ τέτοιο ώστε:

$$\inf_{x \in B^o} I(x) = \inf_{x \in \bar{B}} I(x) := I_B,$$

τότε

$$\lim_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) = -I_B. \tag{2.2}$$

Ένα σύνολο B που ικανοποιεί την (2.2) καλείται σύνολο συνέχειας, εν γένει η LDP μας δίνει ακριβή όρια μόνο για σύνολα συνέχειας.

Έχοντας παρακολουθήσει και την απόδειξη του θεωρήματος Cramér στο $\mathbb{R}$ είμαστε σε θέση να κατανοήσουμε λίγο καλύτερα γιατί ο ορισμός της LDP έχει αυτή τη συγκεκριμένη μορφή. Αρχικά στην περίπτωση που έχουμε να κάνουμε με μη ατομικά μέτρα θα ισχύει $\mu_\epsilon(\{x\}) = 0 \quad \forall x \in \mathbb{S}$, συνεπώς εάν στον ορισμό παίρναμε το κάτω όριο πάνω από το B αντί του B^o θα έπρεπε να καταλήξουμε ότι $I(x) \equiv \infty$, το οποίο είναι προφανώς άτοπο γιατί τα μ_ϵ είναι μέτρα πιθανότητας. Συνεπώς βλέπουμε ότι το να επιβάλλουμε κάποιους τοπολογικούς περιορισμούς είναι φυσιολογικό και αναγκαίο. Ο ορισμός της LDP κωδικοποιεί με έναν πολύ βολικό τρόπο το γεγονός ότι τα ασυμπτωτικά φράγματα, από τη μία πλευρά είναι αρκετά ακριβή ώστε να είναι χρήσιμα και από την άλλη είναι αρκετά γενικά ώστε να είναι ορθά.

Από την άλλη εφόσον $\mu_\epsilon(\mathbb{S}) = 1 \quad \forall \epsilon > 0$ είναι αναγκαίο να ισχύει $\inf_{x \in \mathbb{S}} I(x) = 0$ προκειμένου το άνω φράγμα του ορισμού να έχει νόημα. Επιπλέον όταν η I είναι καλή αυτό σημαίνει ότι θα υπάρχει μοναδικό σημείο x τέτοιο ώστε $I(x) = 0$. Αξίζει ακόμη να παρατηρήσουμε ότι το άνω όριο ισχύει τετριμμένα όταν $\inf_{x \in \bar{B}} I(x) = 0$, ενώ το κάτω όριο ισχύει τετριμμένα όταν $\inf_{x \in B^o} I(x) = \infty$. Αυτή η παρατήρηση μας οδηγεί στην ισοδύναμη διατύπωση της LDP:

Άνω φράγμα: $\forall \alpha < 0$ και κάθε μετρήσιμο σύνολο B με $\bar{B} \subset \Psi_I(a)^c$ έχουμε:

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \leq -\alpha. \tag{2.3}$$

Κάτω φράγμα: $\forall x \in D_I = \{x : I(x) < \infty\}$ και κάθε μετρήσιμο σύνολο B με $x \in B^o$ έχουμε:

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(B) \geq -I(x). \tag{2.4}$$

Σχόλιο 16. *Εφόσον έχουμε υποθέσει ότι εργαζόμαστε σε μετρικό χώρο η LDP μπορεί να ελεγχθεί για ακολουθίες και τότε η σχέση (2.1) παίρνει την πιο γνώριμη σε εμάς μορφή*

$$-\inf_{x\in B^o} I(x) \leq \liminf_{n\to\infty} \alpha_n \log \mu_n(B) \leq \limsup_{n\to\infty} \alpha_n \log \mu_n(B) \leq -\inf_{x\in \bar{B}} I(x),$$

όπου για την ακολουθία α_n ισχύει $\alpha_n \to 0$ καθώς $n \to \infty$.

Προφανώς όταν όλα είναι ′ *καλά* ′ θα πάρουμε μια ισότητα της μορφής:

$$\lim_{n\to\infty} \alpha_n \log \mu_n(B) = -I_B.$$

Ορισμός 6. *Θα λέμε ότι η οικογένεια των μέτρων πιθανότητας $\{\mu_\epsilon\}$ (ή αντιστοίχως $\{\mu_n\}$) είναι εκθετικά σφιχτή (exponentially tight) στο $\mathbb{S}$, εάν για κάθε $\alpha < \infty$ υπάρχει συμπαγές σύνολο $K_\alpha \subset \mathbb{S}$ τέτοιο ώστε*

$$\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(K_\alpha^c) < -\alpha.$$

Σχόλιο 17. *Στο σημείο αυτό θα πρέπει να είμαστε προσεκτικοί σχετικά με τη σχέση της εκθετικής σφιχτότητας και του κατά πόσο είναι καλή η συνάρτηση ταχύτητας, πιο συγκεκριμένα σε έναν γενικό χώρο δεν είναι απαραίτητο η $\{\mu_\epsilon\}$ να είναι εκθετικά σφιχτή προκειμένου να ικανοποιεί την LDP με καλή συνάρτηση ταχύτητας, ωστόσο σε πολωνικούς χώρους μια καλή συνάρτηση ταχύτητας συνεπάγεται ότι η $\{\mu_\epsilon\}$ είναι εκθετικά σφιχτή.*

Η έννοια της εκθετικής σφιχτότητας θα μας φανεί χρήσιμη για την απόδειξη του θεωρήματος Cramér στο $\mathbb{R}^d$, πριν περάσουμε στη διατύπωση του θεωρήματος δίνουμε έναν ακόμα χρήσιμο ορισμό.

Ορισμός 7. *Για μια συνάρτηση ταχύτητας I και για οποιδήποτε $\delta > 0$ η συνάρτηση δ-ταχύτητας (δ-rate function) ορίζεται ως εξής :*

$$I^\delta(x) := \min\{I(x) - \delta, \frac{1}{\delta}\}.$$

Σχόλιο 18. *Παρόλο που εν γένει η $I^\delta(x)$ δεν είναι συνάρτηση ταχύτητας η χρησιμότητα του ορισμού έγκειται στο γεγονός ότι για οποιοδήποτε σύνολο B ισχύει,*

$$\lim_{\delta\to 0} \inf_{x\in B} I^\delta(x) = \inf_{x\in B} I(x).$$

Όταν ο χώρος $(\mathbb{S}, \mathcal{B})$ είναι πλήρης χώρος μέτρου, η LDP είναι ισοδύναμη με,

(α) Για κάθε κλειστό σύνολο $F \subset \mathbb{S}$,

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(F) \leq -\inf_{x \in F} I(x). \tag{2.5}$$

(β) Για κάθε ανοιχτό σύνολο $G \subset \mathbb{S}$,

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G) \geq -\inf_{x \in G} I(x). \tag{2.6}$$

Λήμμα 1. *Έστω $\{\mu_\epsilon\}$ μια εκθετικά σφιχτή οικογένεια μέτρων πιθανότητας, τότε ισχύουν τα εξής:*

(α) Εάν ισχύει το άνω φράγμα της σχέσης (2.3) για κάποιο $\alpha < \infty$ και για όλα τα συμπαγή σύνολα $\Psi_I^c(\alpha)$, τότε ισχύει και για όλα τα μετρήσιμα σύνολα Γ για τα οποία $\bar{\Gamma} \subset \Psi_I^c(\alpha)$. Επιπλέον, εάν ο χώρος $(\mathbb{S}, \mathcal{B})$ είναι πλήρης χώρος μέτρου και η (2.5) ισχύει για όλα τα συμπαγή θα ισχύει και για όλα τα κλειστά.

(β) Εάν ισχύει το κάτω φράγμα της σχέσης (2.4) για κάθε μετρήσιμο σύνολο (αντίστοιχα για την περίπτωση που ο χώρος $(\mathbb{S}, \mathcal{B})$ είναι πλήρης χώρος μέτρου το κάτω φράγμα της σχέσης (2.6) για όλα τα ανοιχτά), τότε η I είναι μια καλή συνάρτηση ταχύτητας.

Απόδειξη. Η απόδειξη μπορεί να βρεθεί στο [1]. $\square$

2.2 Το Θεώρημα Cramér στο $\mathbb{R}^d$

Το θεώρημα Cramér στο $\mathbb{R}^d$ και το θεώρημα Gärtner-Ellis που θα μελετήσουμε στην επόμενη παράγραφο αποτελούν άμεσες γενικεύσεις του μονοδιάστατου θεωρήματος Cramér, το μεν κατά προφανή τρόπο το δε αφαιρώντας την υπόθεση της ανεξαρτησίας των X_n.

Έστω ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, ..., X_n \sim \mu$, οι οποίες παίρνουν τιμές στο $\mathbb{R}^d$, τότε ορίζουμε κατά τα γνωστά

$$\Lambda(\theta) = \log \int_{\mathbb{R}^d} e^{\theta x} d\mu(x), \theta \in \mathbb{R}^d \quad \& \quad \Lambda^*(x) = \sup_{\theta \in \mathbb{R}^d} (\theta x - \Lambda(\theta)).$$

Επιπλέον συμβολίζουμε με μ_n την κατανομή της τυχαίας μεταβλητής $\bar{X}_n, n \in \mathbb{N}$ και για τη διευκόλυνση των αποδείξεων παίρνουμε $D_\Lambda = \mathbb{R}^d$, δηλαδή $\Lambda(\theta) < \infty$ για κάθε θ, πιο συγκεκριμένα παίρνουμε $\mathbb{E}(|X_1^2|) < \infty$ και $\bar{X}_n \xrightarrow[\text{κατά πιθ.}]{n \to \infty} m$.

Θεώρημα 2 (Cramér στο $\mathbb{R}^d$)**.** *Έστω $D_\Lambda = \mathbb{R}^d$,τότε η ακολουθία μέτρων πιθανότητας $\{\mu_n\}$ ακολουθεί την LDP με συνάρτηση ταχύτητας την κυρτή Λ^* και $\{\alpha_n\} = \frac{1}{n}, \quad n \in \mathbb{N}$. Επιπλέον η Λ^* είναι καλή.*

Το επόμενο λήμμα συνοψίζει τις ιδιότητες των Λ και Λ^* που χρειαζόμαστε για την απόδειξη.

Λήμμα 2. *(α) Η Λ είναι κυρτή και διαφορισίμη παντού και η Λ^* είναι καλή και κυρτή συνάρτηση ταχύτητας.*

(β) $y = \nabla\Lambda(\eta) \Rightarrow \Lambda^*(y) = \langle \eta, y \rangle - \Lambda(\eta)$, όπου με $\langle \cdot, \cdot \rangle$ συμβολίζουμε το σύνηθες εσωτερικό γινόμενο του $\mathbb{R}^d$.

Απόδειξη. (α) Θα σκιαγραφήσουμε σύντομα την απόδειξη του λήμματος γιατί τα επιχειρήματα που θα χρησιμοποιήσουμε είναι παρόμοια επιχερήματα με τη μονοδιάστατη περίπτωση. Η κυρτότητα της Λ προκύπτει άμεσα από την ανισότητα Hölder, ενώ η διαφορισιμότητα της προκύπτει από το θεώρημα κυριαρχημένης σύγκλισης. Για τη Λ^* η κυρτότητα και η ιδιότητα της κάτω ημισυνέχειας προκύπτουν αλλάζοντας κατάλληλα τις ιδιότητες 4-5 του προηγούμενου κεφαλαίου. Επιπλέον, εφόσον $\Lambda(0) = 0$, η Λ^* θα είναι μη αρνητική και κάτω ημισυνεχής και άρα είναι συνάρτηση ταχύτητας. Επιπλέον παρατηρούμε ότι για κάθε $\kappa > 0$

$$\Lambda^*(x) \geq \kappa|x| - \sup_{|\theta|=\kappa} \{\Lambda(\theta)\}.$$

Πιο συγκεκριμένα όλα τα level-sets της Λ^* είναι φραγμένα και άρα η Λ^* είναι καλή.

(β) Έστω $y = \nabla\Lambda(\eta)$, σταθεροποιούμε ένα σημείο $\theta \in \mathbb{R}^d$ και έστω

$$g(\alpha) := \alpha\langle \theta - \eta, y \rangle - \Lambda(\eta + \alpha(\theta - \eta)) + \langle \eta, y \rangle, a \in [0, 1].$$

Εφόσον η Λ είναι κυρτή και το $\Lambda(\eta) < \infty$, η $g(\cdot)$ θα είναι κοίλη και $|g(0)| < \infty$. Συνεπώς,

$$g(1) - g(0) \leq \liminf_{\alpha \to 0} \frac{g(\alpha) - g(0)}{\alpha} = \langle \theta - \eta, y - \nabla\theta(\eta) \rangle = 0.$$

Συνεπώς για κάθε θ θα έχουμε,

$$g(1) = [\langle \theta, y \rangle - \Lambda(\theta)] \leq g(0) = [\langle \eta, y \rangle - \Lambda(\theta)] \leq \Lambda^*(y).$$

Το ζητούμενο προκύπτει τώρα εάν πάρουμε το supremum για κάθε θ. □

2.2.1 Απόδειξη Θεωρήματος Cramér στο $\mathbb{R}^d$

Απόδειξη. Θα ξεκινήσουμε την απόδειξη αποδεικνύοντας το άνω φράγμα της LDP, ακολουθώντας εδώ μια ελαφρώς διαφορετική προσέγγιση απ ότι στο $\mathbb{R}$ και ο λόγος είναι ότι στο $\mathbb{R}^d$ χάνουμε τη μονοτονία της Λ^*, οπότε η μέθοδος να κλείσουμε κάθε κλειστό σύνολο F μέσα σε δύο ημιχώρους δεν είναι το ίδιο αποδοτική. Αντ΄ αυτού για το άνω φράγμα των πιθανοτήτων που η ακολουθία $\{\mu_n\}$ αποδίδει στις μπάλες του $\mathbb{R}^d$ θα χρησιμοποιήσουμε την ανισότητα Chebyshev. Στη συνέχεια μπορούμε να καλύψουμε τα συμπαγή σύνολα με μια πεπερασμένη συλλογή κατάλληλα μικρών μπαλών και το άνω φράγμα για συμπαγή σύνολα προκύπτει από την ανισότητα Boole.

Όπως αναφέραμε στην εισαγωγή αυτού του κεφαλαίου, το να δείξουμε την ισχύ του άνω φράγματος είναι ισοδύναμο με το να αποδείξουμε ότι για κάθε $\delta > 0$ και κάθε κλειστό σύνολο $F \subset \mathbb{R}^d$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(F) \leq \delta - \inf_{x \in F} I^\delta(x),$$

όπου η I^δ είναι η δ-συνάρτηση ταχύτητας που παράγεται από τη Λ^*.

Έστω ότι σταθεροποιούμε ένα συμπαγές σύνολο $\Gamma \subset \mathbb{R}^d$. Για κάθε $\gamma \in \Gamma$ διαλέγουμε $\theta_\gamma \in \mathbb{R}^d$ τέτοιο ώστε,

$$\langle \theta_\gamma, \gamma \rangle - \Lambda(\theta_\gamma) \geq I^\delta(\gamma).$$

Για κάθε γ, διαλέγουμε $\rho_\gamma > 0$ τέτοιο ώστε $\rho_\gamma |\theta_\gamma| \leq \delta$ και έστω $B_{\gamma,\rho_\gamma} = \{x : |x - \gamma| < \rho_\gamma\}$ να είναι η μπάλα με κέντρο γ και ακτίνα ρ_γ. Παρατηρούμε τώρα ότι για κάθε $n \in \mathbb{N}, \theta \in \mathbb{R}^d$ και κάθε μετρήσιμο σύνολο $G \subset \mathbb{R}^d$ έχουμε,

$$\mu_n(G) = \mathbb{E}[\mathbb{1}_{\bar{X}_n \in G}] \leq \mathbb{E}[e^{\langle \theta, \bar{X}_n \rangle - \inf_{x \in G}\{\langle \theta, x \rangle\}}].$$

Πιο συγκεκριμένα, για κάθε n και για κάθε $\gamma \in \Gamma$ έχουμε,

$$\mu_n(B_{\gamma,\rho_\gamma}) \leq \mathbb{E}[e^{n\langle \theta_\gamma, \bar{X}_n \rangle}] e^{-\inf_{x \in B_{\gamma,\rho_\gamma}}\{n\langle \theta_\gamma, x \rangle\}}.$$

Επίσης για κάθε $\gamma \in \Gamma$ έχουμε,

$$-\inf_{x\in B_{\gamma,\rho_\gamma}} \langle \theta_\gamma, x\rangle \leq \rho_\gamma |\theta_\gamma| \leq \delta - \langle \theta_\gamma, q\rangle,$$

και επομένως

$$\frac{1}{n}\log \mu_n(B_{\gamma,\rho_\gamma}) \leq -inf_{x\in B_{\gamma,\rho_\gamma}}\{\langle \theta_\gamma, x\rangle + \Lambda(\theta_\gamma)\} \leq \delta - \langle \theta_\gamma, \gamma\rangle + \Lambda(\theta_\gamma).$$

Εφόσον το Γ είναι συμπαγές, εξ' ορισμού μπορούμε από το ανοιχτό κάλυμμα $\bigcup_{\gamma\in\Gamma} B_{\gamma,\rho_\gamma}$ του Γ να εξάγουμε ένα πεπερασμένο υποκάλυμμα που να αποτελείται από $N = N(\Gamma,\delta) < \infty$ μπάλες, με κέντρα $\gamma_1, ..., \gamma_N \in \Gamma$, οπότε από την ανισότητα Boole και την ανισότητα της προηγούμενης σχέσης έχουμε

$$\frac{1}{n}\log \mu_n(\Gamma) \leq \frac{1}{n}\log N + \delta - \min_{i=1,\dots,N}\{\langle \theta_{\gamma_i}, \gamma_i\rangle - \Lambda(\theta_{\gamma_i})\}.$$

Συνεπώς από την επιλογή του θ_γ,

$$\limsup_{n\to\infty} \frac{1}{n}\log \mu_n(\Gamma) \leq \delta - \min_{i=1,\dots,N} I^\delta(\gamma_i).$$

Εφόσον το $\gamma_i \in \Gamma$, αυτό που καταφέραμε είναι να αποδείξουμε το άνω φράγμα για όλα τα συμπαγή σύνολα. Τώρα μπορούμε να επεκτείνουμε το άνω φράγμα των μεγάλων αποκλίσεων από τα συμπαγή στα κλειστά υποσύνολα του $\mathbb{R}^d$ που είναι και το ζητούμενο, δείχνοντας ότι η $\{\mu_n\}$ είναι μια εκθετικά σφιχτή οικογένεια μέτρων πιθανότητας και εφαρμόζοντας το Λήμμα 1 στην εισαγωγή αυτού του κεφαλαίου.

Έστω $H_\rho \coloneqq [-\rho,\rho]^d$. Εφόσον $H_\rho^c = \bigcup_{j=1}^d \{x : |x^j| > \rho\}$, η ανισότητα Boole θα μας δώσει

$$\mu_n(H_\rho^c) \leq \sum_{j=1}^d \mu_n^j([\rho,\infty]) + \sum_{j=1}^d \mu_n^j((-\infty,-\rho]) \tag{2.7}$$

όπου $\mu_n^j, j = 1, ..., d$ είναι οι κατανομές των συντεταγμένων του τυχαίου διανύσματος $\bar{X}_n$, δηλαδή των $\frac{1}{n}\sum_{i=1}^n X_i^j$. Τώρα χρησιμοποιούμε τις σχέσεις,

$$\mu_n([x,\infty)) \leq e^{-n\Lambda^*(x)} \quad \& \quad \mu_n(-\infty,x]) \leq e^{-n\Lambda^*(x)}, \quad x < m,$$

(είναι ουσιαστικά η περίπτωση I του μονοδιάστατου αντίστοιχου) έχουμε ότι για κάθε $\rho \geq |m|$,

$$\mu_n^j((-\infty,-\rho]) \leq e^{-n\Lambda_j^*(-\rho)}, \mu_n^j([\rho,\infty)) \leq e^{-n\Lambda_j^*(\rho)},$$

όπου με Λ_j^* συμβολίζουμε το μετασχηματισμό Fenchel-Legendre της $\log \mathbb{E}[e^{\theta X_1^j}], j = 1, ..., m$.

Θα χρειαστούμε τώρα το εξής λήμμα.

Λήμμα 3. *Εάν $0 \in D_\Lambda{}^o$, τότε η Λ^* είναι καλή συνάρτηση ταχύτητας. Επιπλέον εάν $D_\Lambda = \mathbb{R}$, τότε*

$$\lim_{|x|\to\infty} \frac{\Lambda^*(x)}{|x|} = \infty. \tag{2.8}$$

Απόδειξη. Εφόσον $0 \in D_\Lambda{}^o$, υπάρχει $\theta_- < 0$ και $\theta_+ > 0$ τα οποία περιέχονται και τα δύο στο D_Λ. Εφόσον για κάθε $\theta \in \mathbb{R}$,

$$\frac{\Lambda^*(x)}{|x|} \geq \theta \mathrm{sign(x)} - \frac{\Lambda(\theta)}{|x|},$$

έπεται ότι,

$$\liminf_{|x|\to\infty} \frac{\Lambda^*(x)}{|x|} \min\{\theta_-, \theta_+\} > 0.$$

Πιο συγκεκριμένα, $\Lambda^*(x) \longrightarrow \infty$ καθώς $|x| \longrightarrow \infty$, και τα σύνολα $\Psi_{\Lambda^*}(\alpha), \alpha \in [0,\infty)$(level sets), είναι κλειστά και φραγμένα και άρα είναι συμπαγή. Συνεπώς η Λ^* είναι καλή, η περίπτωση για $D_\Lambda = \mathbb{R}$ προκύπτει άμεσα εάν θεωρήσουμε $\theta_- = \theta_+ = \infty$. Αυτό ολοκληρώνει την απόδειξη. □

Επομένως, από το παραπάνω λήμμα προκύπτει ότι $\Lambda_j^*(x) \to \infty$ καθώς $|x| \to \infty$. Συνεπώς συνδυάζοντας τα προηγούμενα άνω φράγματα με τη σχέση (2.3) και παίρνοντας το όριο αρχικά $n \to \infty$ και $\rho \to \infty$ παίρνουμε τη σχέση,

$$\lim_{\rho\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(H_\rho^c) = -\infty.$$

Συνεπώς, η οικογένεια $\{\mu_n\}$ είναι εκθετικά σφιχτή, εφόσον οι υπερκύβοι H_ρ είναι συμπαγείς.

Επομένως έχουμε ολοκληρώσει την απόδειξη του άνω φράγματος και προχωράμε στο κάτω φράγμα. Για το σκοπό μας, επαρκεί να αποδείξουμε ότι για κάθε $y \in D_{\Lambda^*}$ και κάθε $\delta > 0$, ισχύει

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B_{y,\delta}) \geq -\Lambda^*(y).$$

Ας υποθέσουμε αρχικά $y = \nabla\Lambda(\eta)$ για κάποιο $\eta \in \mathbb{R}^m$. Τώρα ορίζουμε το μέτρο πιθανότητας ν από τη σχέση

$$\frac{d\nu}{d\mu}(z) = e^{\langle \eta, z\rangle - \Lambda^*(y)},$$

και έστω ότι συμβολίζουμε με ν_n την κατανομή του $\bar{X}_n$, όταν οι X_i είναι ανεξάρτητες και ισόνομες και ακολουθούν την ν, τότε

$$\frac{1}{n} \log \mu_n(B_{y,\delta}) = \Lambda(\eta) - \langle \eta, y\rangle + \frac{1}{n} \log \int_{z\in B_{y,\delta}} e^{n\langle \eta, y-z\rangle} \nu_n(dz)$$

$$\geq \Lambda(\eta) - \langle \eta, y\rangle - |\eta|\delta + \frac{1}{n} \log \nu_n(B_{y,\delta}).$$

Τώρα από το θεώρημα κυριαρχημένης σύγκλισης έχουμε,

$$\mathbb{E}_\nu(X_1) = \frac{1}{\mathbb{E}(e^{\eta X_1})} \int_{\mathbb{R}^d} x e^{\langle \eta, x\rangle} d\nu = \nabla\Lambda(\eta) = y,$$

και από ΑΝΜΑ θα έχουμε, $\lim_{n\to\infty} \nu_n(B_{y,\delta}) = 1$ για κάθε $\delta > 0$. Επιπλέον, εφόσον

$$\Lambda(\eta) - \langle \eta, y\rangle \geq -\Lambda^*(y),$$

η προηγούμενη ανισότητα συνεπάγεται το κάτω φράγμα,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B_{y,\delta}) \geq -\Lambda^*(y) - |\eta|\delta.$$

Επομένως,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B_{y,\delta}) \geq \liminf_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B_{y,\delta}) \geq -\Lambda^*(y).$$

Θα θέλαμε τώρα να επεκτείνουμε το κάτω φράγμα που αποδείξαμε έτσι ώστε να συμπεριλάβουμε όλα τα $y \in D_{\Lambda^*}$ τέτοια ώστε $y \notin \{\nabla\Lambda(\lambda) : \theta \in \mathbb{R}^d\}$, τώρα κανονικοποιούμε το μ προσθέτοντας στο κάθε X_j μια κατάλληλα μικρή κανονική τυχαία μεταβλητή. Πιο συγκεκριμένα, σταθεροποιούμε ένα $M < \infty$ και έστω κ να είναι η από κοινού κατανομή των ισόνομων και ανεξάρτητων τυχαίων διανυσμάτων $Y_j = X_j + \frac{V_j}{\sqrt{M}}$, όπου $V_1, ..., V_n$ είναι τυπικές πολυμετάβλητες κανονικές τυχαίες μεταβλητές ανεξάρτητες από τις $X_1, ..., X_n$. Έστω Λ_M να είναι η λογαριθμική ροπογεννήτρια της Y_1, ενώ με κ_n συμβολίζουμε την κατανομή της $\bar{X}_n^{(M)} = \sum_{j=1}^n Y_j$.

Η λογαριθμική ροπογεννήτρια της τυπικής πολυμετάβλητης κανονικής κατανομής είναι $\frac{|\theta|^2}{2}$, συνεπώς έπεται ότι,

$$\Lambda_M(\theta) = \Lambda(\theta) + \frac{1}{2M}|\theta|^2 \geq \Lambda(\theta),$$

και επομένως,

$$\Lambda^*(y) = \sup_{\theta\in\mathbb{R}^d} \{\langle \theta, y\rangle - \Lambda(\theta)\} \geq \sup_{\theta\in\mathbb{R}^d} \{\langle \theta, y\rangle - \Lambda_M(\theta)\}.$$

Εξ' υποθέσεως έχουμε ότι, η $m = \mathbb{E}(X_1)$ είναι πεπερασμένη, επομένως η ανισότητα Jensen θα μας δώσει ότι για κάθε $\theta \in \mathbb{R}^d$ ισχύει $\Lambda(\theta) \geq \langle \theta, x\rangle$. Επομένως η πεπερασμένη και διαφορίσιμη συνάρτηση,

$$g(\theta) := \langle \theta, y\rangle - \Lambda_M(\theta)$$

ικανοποιεί τη σχέση,

$$\lim_{\rho\to\infty} \sup_{|\theta|>\rho} g(\theta) = -\infty.$$

Συνεπώς το supremum της g πάνω από το $\mathbb{R}^d$ ' *πίανεται* ' για κάποιο πεπερασμένο η για το οποίο,

$$0 = \nabla g(\eta) = y - \nabla \Lambda_M(\eta),$$

δηλαδή $y = \nabla \Lambda_M(\eta)$. Άρα, από την προηγούμενη απόδειξη, το κάτω φράγμα της LDP θα ισχύει για την $\{\kappa_n\}$ και για κάθε $\delta > 0$ έχουμε,

$$\liminf_{n\to\infty} \frac{1}{n} \log \kappa_n(B_{y,\delta}) \geq -\Lambda^*(y) > -\infty.$$

Τώρα παρατηρούμε ότι η $\bar{X}_n^M$ θα ακολουθεί την ίδια κατανομή με την $\bar{X}_n + \frac{V}{\sqrt{Mn}}$, όπου $V \sim Normal(0, I)$ και είναι ανεξάρτητη από την $\bar{X}_n$. Επομένως,

$$\mu_n(B_{y,2\delta}) \geq \kappa_n(B_{y,\delta}) - \mathbb{P}(|V| \geq \sqrt{Mn}\delta).$$

Τελικά, επειδή η V είναι τυπική πολυμετάβλητη κανονική θα ισχύει,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|V| \geq \sqrt{Mn\delta}) \leq -\frac{M\delta^2}{2}.$$

Τώρα η απόδειξη ολοκληρώνεται συνδυάζοντας τις τρεις προηγούμενες ανισότητες και παίρνοντας πρώτα $n \to \infty$ και στη συνέχεια $M \to \infty$. □

Σχόλιο 19. *Μια πιο προσεκτική επιθεώρηση της παραπάνω απόδειξης φανερώνει ότι για την απόδειξη του άνω φράγματος για τα συμπαγή σύνολα δεν χρειάστηκε να υποθέσουμε τίποτα για το σύνολο* D_Λ. *Επιπλέον η συνθήκη ότι* $0 \in D_\Lambda{}^o$ *αρκεί για να σιγουρέψουμε ότι η* Λ^* *είναι καλή συνάρτηση ταχύτητας και ότι η οικογένεια* $\{\mu_n\}$ *είναι εκθετικά σφιχτή. Αντιθέτως για την απόδειξη του κάτω φράγματος χρειάστηκε να κάνουμε την ισχυρή υπόθεση ότι* $D_\Lambda = \mathbb{R}^d$.

2.3 Το Θεώρημα Gärtner-Ellis

Η ισχύς θεωρήματος Cramér που εξετάσαμε στην προηγούμενη ενότητα περιορίζεται στην περίπτωση που οι τυχαίες μεταβλητές που μελετάμε είναι ανεξάρτητες και ισόνομες. Ωστόσο μια ματιά στην απόδειξη του θεωρήματος Cramér θα πρέπει να μας πείσει ότι μια επέκταση στην οποία χαλαρώνουμε την απαίτηση της ανεξαρτησίας είναι δυνατή. Η επέκταση αυτή είναι το αντικείμενο του θεωρήματος Gärtner-Ellis, που θα μελετήσουμε λεπτομερώς σε αυτήν την παράγραφο.

Έστω η ακολουθία τυχαίων διανυσμάτων $\{Z_n\}$, όπου $Z_n \in \mathbb{R}^d$ ακολουθεί την μ_n και έχει λογαριθμική ροπογεννήτρια,

$$\Lambda_n(\theta) := \log \mathbb{E}[e^{\langle \theta, Z_n \rangle}]. \tag{2.9}$$

Η ύπαρξη του ορίου της λογαριθμικής ροπογεννήτριας μας υποδεικνύει ότι η $\{\mu_n\}$ θα πρέπει να ικανοποιεί την LDP. Πιο συγκεκριμένα κάνουμε την εξής υπόθεση για το υπόλοιπο αυτής της παραγράφου.

Υπόθεση 1. *Για κάθε $\theta \in \mathbb{R}^d$, η συνάρτηση Λ που ορίζεται ως το όριο,*

$$\Lambda(\theta) := \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\theta)$$

υπάρχει ως επεκτεταμένος πραγματικός αριθμός (δηλαδή ανήκει στο διάστημα $[-\infty, +\infty]$) και $0 \in D_\Lambda^o$, όπου $D_\Lambda = \{\theta \in \mathbb{R}^d : \Lambda(\theta) < \infty\}$.

Στο σημείο αυτό ας παρατηρήσουμε ότι εάν η μ_n είναι η κατανομή του μέσου $\bar{X}_n$ ανεξαρτήτων και ισονόμων τυχαίων διανυσμάτων $X_i \in \mathbb{R}^d, i = 1, ..., n$, τότε για κάθε $n \in \mathbb{N}$ έχουμε,

$$\frac{1}{n}\Lambda_n(n\theta) = \Lambda(\theta) := \log \mathbb{E}[e^{\langle \lambda, X_1 \rangle}],$$

για $0 \in D_\Lambda^o$ η υπόθεση θα ισχύει.

Τώρα, έστω κατά τα γνωστά Λ^* ο μετασχηματισμός Fenchel-Legendre της Λ, με $D_{\Lambda^*} = \{x \in \mathbb{R}^d : \Lambda^*(x) < \infty\}$. Κατ΄ αναλογία με το θεώρημα Cramer ο σκοπός μας είναι να βρούμε τις συνθήκες κάτω από τις οποίες η $\{\mu_n\}$ ακολουθεί την LDP με συνάρτηση ταχύτητας την Λ^*.

Ορισμός 8. *Θα λέμε ότι ένα $y \in \mathbb{R}^d$ είναι εκτεθειμένο σημείο (exposed point) της Λ^* εάν για κάποιο $\theta \in \mathbb{R}^d$ και για κάθε $x \neq y$,*

$$\langle \theta, y \rangle - \Lambda^*(y) > \langle \theta, x \rangle - \Lambda^*(x) \tag{2.10}$$

Το θ στη σχέση (2.10) θα καλείται εκτεθειμένο υπερεπίπεδο(exposing hyperplane).

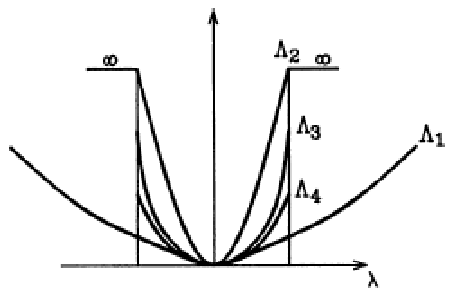
Ορισμός 9. *Θα λέμε ότι μια κυρτή συνάρτηση $\Lambda : \mathbb{R}^d \longrightarrow (-\infty, \infty]$ είναι ουσιωδώς λεία εάν:*

(α) Το D_Λ° είναι μη κενό.

(β) Η Λ είναι διαφορίσιμη σε ολόκληρο το D_Λ°.

(γ) Η Λ είναι απότομη (steep), δηλαδή $\lim_{n\to\infty} |\nabla\Lambda(\lambda_n)| = \infty$, όποτε η $\{\lambda_n\}$ είναι μια ακολουθία του D_Λ° η οποία συγκλίνει σε κάποιο συνοριακό σημείο του D_Λ°.

Ένα γραφικό παράδειγμα του τι θα πεί ότι η Λ είναι απότομη, στο παρακάτω σχήμα οι $\Lambda_1, \Lambda_2, \Lambda_3$ είναι απότομες ενώ η Λ_4 όχι.



Είμαστε τώρα σε θέση να διατυπώσουμε το θεώρημα Gärtner-Ellis.

Θεώρημα 3 (Gärtner-Ellis). *Έστω ότι ισχύει η υπόθεση 1, τότε*

(α) Για κάθε κλειστό σύνολο F,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq - \inf_{x\in F} \Lambda^*(x). \tag{2.11}$$

(β) Για κάθε ανοιχτό σύνολο G,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(F) \geq - \inf_{x\in G\cap\mathcal{F}} \Lambda^*(x) \tag{2.12}$$

όπου $\mathcal{F}$ είναι το σύνολο των εκτεθειμένων σημείων της Λ^* των οποίων τα εκτεθειμένα υπερεπίπεδα ανήκουν στο D_Λ°.

(γ) Εάν η Λ είναι ουσιωδώς λεία και κάτω ημισυνεχής, τότε ισχύει η LDP με καλή συνάρτηση ταχύτητας την Λ^*.

Σχόλιο 20. *(α) Όλα τα αποτελέσματα που αναπτύσσονται σε αυτή την παράγραφο εξακολουθούν να ισχύουν όταν η ακολουθία* $\frac{1}{n}$ *αντικατασταθεί από μία αυθαίρετη ακολουθία* $\alpha_n \to 0$*, όπως στη διατύπωση που παρουσιάσαμε στην εισαγωγή.*

(β) Τα πιο ουσιώδη επιχειρήματα για την απόδειξη των σκελών (α),(β) του θεωρήματος Gärtner-Ellis, είναι τα επιχειρήματα που έχουμε ήδη χρησιμοποιήσει για την απόδειξη του θεωρήματος Cramer, δηλαδή η χρήση της ανισότητας Chebyshev για την απόδειξη του άνω φράγματος και η εκθετική αλλαγή μέτρου για την απόδειξη του κάτω φράγματος. Ωστόσο επειδή πλέον ο νόμος των μεγάλων αριθμών δεν μας είναι διαθέσιμος, όπως θα δούμε για την απόδειξη του κάτω φράγματος θα χρησιμοποιήσουμε το άνω φράγμα.

(γ) Παρόλο που το θεώρημα είναι αρκετά γενικό, δεν καλύπτει όλες τις περιπτώσεις στο $\mathbb{R}^d$ *που ικανοποιείται μια LDP. Ένα παράδειγμα μια τέτοιας περίπτωσης είναι το εξής, έστω* $Z_n \sim Exponential(\lambda)$*. Τότε η υπόθεση που κάναμε ισχύει για* $\Lambda(\theta) = 0$ *για* $\theta < 1$ *και* $\Lambda(\theta) = \infty$ *αλλιώς. Επιπλέον η κατανομή των* Z_n *θα έχει συνάρτηση πυκνότητας πιθανότητας* $ne^{-nz}\mathbb{1}_{[0,\infty)}(z)$*, που σημαίνει ότι η LDP θα ικανοποιείται με συνάρτηση ταχύτητας* $I(x) = x$ *για* $x \geq 0$ *και* $I(x) = \infty$ *αλλιώς. Κάνοντας τον υπολογισμό βλέπουμε ότι* $I(\cdot) = \Lambda^*(\cdot)$*. Επομένως,* $\mathcal{F} = \{0\}$ *ενώ* $D_{\Lambda^*} = [0, \infty)$ *και συνεπώς το θεώρημα Gärtner-Ellis δίνει ένα τετριμμένο κάτω φράγμα για σύνολα που δεν περιέχουν το 0.*

(δ) Η υπόθεση που κάναμε συνεπάγεται ότι $\Lambda^*(x) \leq \liminf_{n\to\infty} \Lambda_n^*(x)$ *για* $\Lambda_n^* = \sup_{\theta\in\mathbb{R}^d}\{\langle\theta, x\rangle - \frac{1}{n}\Lambda_n(n\theta)\}$*. Ωστόσο η κατά σημείο σύγκλιση της* Λ_n^* *στη* Λ^* *δεν είναι εγγυημένη. Ένα παράδειγμα που δεν έχουμε κατά σημείο σύγκλιση είναι το εξής, όταν* $\mathbb{P}(Z_n = \frac{1}{n}) = 1$*, έχουμε* $\Lambda_n(\theta) = \frac{\theta}{n} \to 0 = \Lambda(\theta)$*, ενώ* $\Lambda_n^*(0) = \infty$ *και* $\Lambda^*(0) = 0$*.*

Πριν παρουσιάσουμε την απόδειξη του θεωρήματος, θα παρουσιάσουμε και θα αποδείξουμε δύο βοηθητικά λήμματα.

Λήμμα 4. *Έστω ότι ισχύει η υπόθεση 1 , τότε έχουμε*

(α) Η $\Lambda(\theta)$ *είναι κυρτή συνάρτηση και* $\Lambda(\theta) > -\infty$ *παντού, επιπλέον η* $\Lambda^*(x)$ *είναι καλή και κυρτή συνάρτηση ταχύτητας.*

(β) Έστω $y = \nabla\Lambda(\eta)$ *για κάποιο* $\eta \in D_\Lambda^o$*. Τότε,*

$$\Lambda^*(y) = \langle\eta, y\rangle - \Lambda(\eta). \tag{2.13}$$

Επιπλέον $y \in \mathcal{F}$*, με το* η *να είναι το εκτεθειμένο υπερεπίπεδο που αντιστοιχεί στο* y*.*

Απόδειξη. (α) Εφόσον οι Λ_n είναι κυρτές συναρτήσεις, το ίδιο θα ισχύει και για τις $\frac{\Lambda_n}{n}$ καθώς και για το οριό τους τη Λ. Επιπλέον, $\Lambda_n(0) = 0$ και συνεπώς $\Lambda(0) = 0$ που συνεπάγεται ότι η Λ^* είναι μη αρνητική.

Έστω $\Lambda(\theta) = -\infty$ για κάποιο $\theta \in \mathbb{R}^d$, τότε λόγω της κυρτότητας της Λ θα έχουμε $\Lambda(\alpha\theta) = -\infty$ για κάθε $\alpha \in (0, 1]$. Εφόσον $\Lambda(0) = 0$, έπεται από την κυρτότητα της Λ ότι $\Lambda(-\alpha\theta) = \infty$ για κάθε $\alpha \in (0, 1]$, που είναι άτοπο γιατί υποθέσαμε ότι $0 \in D_\Lambda^o$. Επομένως, $\Lambda > -\infty$ παντού.

Εφόσον $0 \in D_\Lambda^o$, έπεται ότι θα υπάρχει $\delta > 0$ τέτοιο ώστε $\bar{B}_{0,\delta} \subset D_\Lambda^o$ και $c = \sup_{\theta \in \bar{B}_{0,\delta}} \Lambda(\theta) < \infty$ επειδή η Λ είναι κυρτή και συνεχής πάνω στο D_Λ^o. Επομένως,

$$\Lambda^*(x) \geq \sup_{\theta \in \bar{B}_{0,\delta}} \{\langle \theta, x \rangle - \Lambda(\theta)\} \geq \sup_{\theta \in \bar{B}_{0,\delta}} \{\langle \theta, x \rangle\} - \sup_{\theta \in \bar{B}_{0,\delta}} \{\Lambda(\theta)\} = \delta|x| - c \tag{2.14}$$

Συνεπώς για κάθε $\alpha < \infty$ το σύνολο $\{x : \Lambda^*(x) \leq \alpha\}$ (level-set) είναι φραγμένο. Η συνάρτηση Λ^* είναι κυρτή και κάτω ημισυνεχής άρα είναι καλή συνάρτηση ταχύτητας.

(β) Η απόδειξη της σχέσης (2.13) είναι απλή επανάληψη του σκέλους (β) του Λήμματος 2. Ας υποθέσουμε τώρα ότι για κάποιο $x \in \mathbb{R}^d$,

$$\Lambda(\eta) = \langle \eta, y \rangle - \Lambda^*(y) \leq \langle \eta, x \rangle - \Lambda^*(x).$$

Τότε για κάθε $\theta \in \mathbb{R}^d$,

$$\langle \theta, x \rangle \leq \Lambda(\eta + \theta) - \Lambda(\eta).$$

Πιο συγκεκριμένα,

$$\langle \theta, x \rangle \leq \lim_{\epsilon \to 0} [\Lambda(\eta + \epsilon\theta) - \Lambda(\eta)] = \langle \theta, \nabla\Lambda(\eta) \rangle.$$

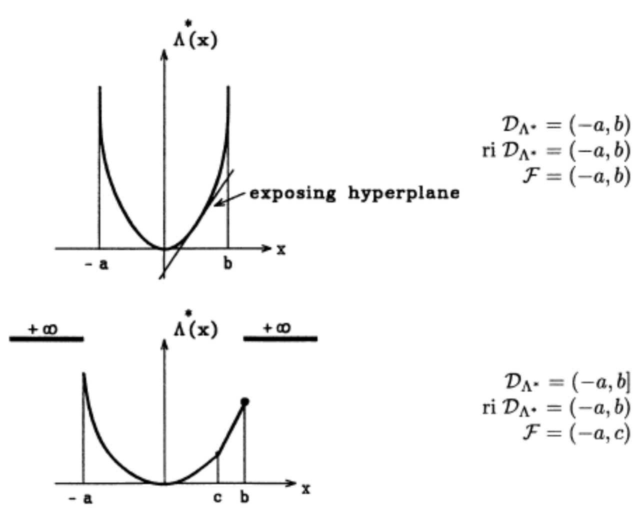
Εφόσον αυτή η ανισότητα ισχύει για κάθε $\theta \in \mathbb{R}^d$, απαραίτητα θα πρέπει να ισχύει $x = \nabla\Lambda(\eta) = y$. Επομένως το y είναι εκτεθειμένο σημείο της Λ^* με εκτεθειμένο υπερεπίπεδο το $\eta \in D_\Lambda^o$ και αυτό ολοκληρώνει την απόδειξη. $\square$

Για τη συνέχεια θα κάνουμε μια σύμβαση συμβολισμού, για κάθε μη κενό κυρτό σύνολο C ορίζουμε το σχετικό εσωτερικό (relative interior) του C,

Ορισμός 10.

$$riC := \{y \in C : x \in C \Rightarrow y - \epsilon(x - y) \in C, \quad \textit{για κάποιο} \quad \epsilon > 0\}.$$

Ένα γραφικό παράδειγμα :



Το επόμενο λήμμα μας είναι χρήσιμο για την απόδειξη του σκέλους (γ) του θεωρήματος.

Λήμμα 5 (Rockafellar). *Εάν* $\Lambda : \mathbb{R}^d \to (-\infty, \infty]$ *είναι μια ουσιωδώς λεία κάτω ημισυνεχής και κυρτή συνάρτηση, τότε* $riD_{\Lambda^*} \subset \mathcal{F}$.

Απόδειξη. Αρχικά ορίζουμε την έννοια του υποδιαφορικού μιας συνάρτησης γιατί θα μας χρειαστεί για την απόδειξη, έστω μια κυρτή συνάρτηση g τότε θα καλούμε υποδιαφορικό της g στο x το σύνολο $\{\theta : g(y) \geq g(x) + \langle \theta, y - x \rangle, \forall y \in \mathbb{R}^d\}$. Επισημαίνουμε τώρα ότι εάν το D_{Λ^*} είναι κενό δεν έχουμε κάτι να αποδείξουμε, το ζητούμενο ισχύει τετριμμένα. Συνεπώς από εδώ και στο εξής θα θεωρούμε ότι το D_{Λ^*} είναι μη κενό.

Τώρα σταθεροποιούμε ένα σημείο $x \in \mathrm{ri}D_{\Lambda^*}$ και ορίζουμε τη συνάρτηση,

$$f(\theta) \triangleq \Lambda(\theta) - \langle \theta, x \rangle + \Lambda^*(x).$$

Εάν $f(\theta) = 0$ τότε προφανώς το λ ανήκει στο υποδιαφορικό της Λ^* στο x. Η απόδειξη ότι $x \in \mathcal{F}$ θα στηριχτεί στο να δείξουμε ότι ένα τέτοιο θ υπάρχει και $\theta \in D_{\Lambda}^{o}$.

Παρατηρούμε τώρα ότι η $f : \mathbb{R}^d \to [0, +\infty]$ είναι μια κυρτή κάτω ημισυνεχής συνάρτηση και $\inf_{\theta \in \mathbb{R}^d} f(\theta) = 0$. Είναι εύκολο να δούμε ότι ο μετασχηματισμός Fenchel-Legendre της f είναι $f^*(\cdot) = \Lambda^*(\cdot + x) - \Lambda^*(x)$. Επομένως $x \in \mathrm{ri}D_{\Lambda^*}$ συνεπάγεται $0 \in \mathrm{ri}D_{f^*}$.

Για τη συνέχεια της απόδειξης θα χρειαστούμε δύο λήμματα κυρτής ανάλυσης, τα οποία παρουσιάζονται παρακάτω:

Λήμμα 6. *Εάν* $f : \mathbb{R}^d \to [0, +\infty]$ *είναι κυρτή, κάτω ημισυνεχής συνάρτηση με* $0 \in riD_{f^*}$, *τότε* $f(\eta) = 0$ *για κάποιο* $\eta \in \mathbb{R}^d$.

Απόδειξη. Η απόδειξη κάνει χρήση του θεωρήματος Hahn-Banach και παραλείπεται καθώς δεν προσθέτει στο αντικείμενο της εργασίας, ο ενδιαφερόμενος μπορεί να ανατρέξει στο [1]. □

Από το Λήμμα 6 θα υπάρχει $\eta \in D_\Lambda$ τέτοιο ώστε $f(\eta) = 0$. Έστω τώρα $\tilde{\Lambda}(\cdot) = \Lambda(\cdot + \eta) - \Lambda(\eta)$, από τις υποθέσεις που έχουμε κάνει η $\tilde{\Lambda}$ είναι μια ουσιωδώς λεία, κυρτή συνάρτηση με $\tilde{\Lambda}(0) = 0$ και επιπλέον είναι εύκολο να δούμε ότι $\tilde{\Lambda}^*(x) = f(\eta) = 0$.

Θα μας χρειαστεί το επόμενο λήμμα.

Λήμμα 7. *Έστω* f *ουσιωδώς λεία, κυρτή συνάρτηση. Εάν* $f(0) = 0$ *και* $f^*(x) = 0$ *για κάποιο* $x \in \mathbb{R}^d$, *τότε* $0 \in D_f^o$.

Απόδειξη. Η απόδειξη παραλείπεται καθώς δεν προσθέτει στο αντικείμενο της εργασίας, ο ενδιαφερόμενος μπορεί να ανατρέξει στο [1]. □

Από το Λήμμα 7 η συνάρτηση $\tilde{\Lambda}$ θα είναι φραγμένη σε μια περιοχή του 0. Επομένως, $\eta \in D_\Lambda^o$ και από τις υποθέσεις μας προκύπτει ότι η $f(\cdot)$ είναι διαφορίσιμη στο η. Επιπλέον, $f(\eta) = \inf_{\lambda \in \mathbb{R}^d} f(\lambda)$ το οποίο συνεπάγεται ότι $\nabla f(\eta) = 0$, δηλαδή $x = \nabla \Lambda(\eta)$. Τώρα, από το (β) του Λήμματος 4 έπεται ότι $x \in \mathcal{F}$ και εφόσον αυτό θα ισχύει για κάθε $x \in \mathrm{ri} D_{\Lambda^*}$ η απόδειξη ολοκληρώθηκε. □

Είμαστε τώρα σε θέση να παρουσιάσουμε την απόδειξη του θεωρήματος Gärtner-Ellis.

Απόδειξη. (Gärtner-Ellis) (α) Το άνω φράγμα που παρουσιάζεται στη σχέση (2.11), για συμπαγή σύνολα προκύπτει με το ίδιο επιχείρημα όπως στην απόδειξη του θεωρήματος Cramer στο $\mathbb{R}^d$. Η επέκταση για όλα τα κλειστά σύνολα προκύπτει δείχνοντας ότι η ακολουθία των μέτρων πιθανότητας $\{\mu_n\}$ είναι εκθετικά σφιχτή. Αυτό θα προσπαθήσουμε να αποδείξουμε. Έστω e_j να είναι το j-οστό μοναδιαίο διάνυσμα βάσης του $\mathbb{R}^d$ για $j = 1, ..., d$. Εφόσον, $0 \in D_\Lambda^o$, υπάρχει $\theta_j > 0, \eta_j > 0$ τέτοια ώστε $\Lambda(\theta_j e_j) < \infty$ και $\Lambda(-\eta_j e_j) < \infty$ για $j = 1, ..., d$. Επομένως από την ανισότητα Chebyshev θα έχουμε,

$$\mu_n^j((-\infty, -\epsilon)) \le e^{-n\eta_j \epsilon + \Lambda_n(-n\eta_j e_j)},$$
$$\mu_n^j([\epsilon, \infty)) \le e^{-n\theta_j \epsilon + \Lambda_n(n\theta_j e_j)}$$

όπου μ_n^j είναι οι κατανομές των συντεταγμένων του τυχαίου διανύσματος Z_n.

Επομένως για $j = 1, ..., d$, έχουμε

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mu_n^j((-\infty, -\epsilon)) = -\infty,$$
$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mu_n^j([\epsilon, \infty)) = -\infty$$

Για τη συνέχεια θα χρειαστούμε το εξής λήμμα:

Λήμμα 8. *Έστω N ένας φυσικός αριθμός, τότε για κάθε $a_\epsilon^i \ge 0$,*

$$\limsup_{\epsilon \to 0} \epsilon \log \left(\sum_{i=1}^N a_\epsilon^i \right) = \max_{i=1,...,N} \limsup_{\epsilon \to 0} \epsilon \log a_\epsilon^i.$$

Απόδειξη. Παρατηρούμε ότι για κάθε ϵ,

$$0 \le \epsilon \log \left(\sum_{i=1}^N a_\epsilon^i \right) - \max_{i=1,...,N} \epsilon \log a_\epsilon^i \le \epsilon \log N.$$

Εφόσον το N είναι σταθερό $\epsilon \log N \to 0$ καθώς $\epsilon \to 0$ και

$$\max_{i=1,\dots,N} \limsup_{\epsilon\to 0} \epsilon \log a_\epsilon^i = \limsup_{\epsilon\to 0} \max_{i=1,\dots,N} \epsilon \log a_\epsilon^i$$

□

Άρα από την ανισότητα Boole και το Λήμμα 8 θα έχουμε,

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(([-\epsilon, \epsilon]^d)^c) = -\infty,$$

δηλαδή η ακολουθία $\{\mu_n\}$ είναι εκθετικά σφιχτή.

(β) Προκειμένου να αποδείξουμε το κάτω φράγμα (2.12) για κάθε κλειστό σύνολο αρκεί να αποδείξουμε ότι για κάθε $y \in \mathcal{F}$ ισχύει,

$$\lim_{\delta\to 0} \liminf_{n\to\ 0} \frac{1}{n} \log \mu_n(B_{y,\delta}) \geq -\Lambda^*(y). \tag{2.15}$$

Σταθεροποιούμε ένα $y \in \mathcal{F}$ και έστω ένα $\eta \in D_\Lambda{}^{\mathrm{o}}$ να είναι ένα εκτιθέμενο υπερεπίπεδο για το y. Τότε για κάθε n αρκετά μεγάλο, $\Lambda(n\eta) < \infty$ και τα αντίστοιχα μέτρα πιθανότητας είναι καλώς ορισμένα μέσω της σχέσης,

$$\frac{d\nu_n}{d\mu_n}(z) = e^{[n\langle \eta, z\rangle - \Lambda_n(n\eta)]}.$$

Κάνοντας μερικούς υπολογισμούς θα έχουμε,

$$\frac{1}{n} \log \mu_n(B_{y,\delta}) = \frac{1}{n}\Lambda_n(n\eta) - \langle \eta, y\rangle + \frac{1}{n} \log \int_{z\in B_{y,\delta}} e^{n\langle \eta, y-z\rangle} d\nu_n(z) \geq$$

$$\geq \frac{1}{n}\Lambda_n(n\eta) - \langle \eta, y\rangle - |\eta|\delta + \frac{1}{n} \log \nu_n(B_{y,\delta})$$

Επομένως,

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B_{y,\delta}) \geq \Lambda(\eta) - \langle \eta, y\rangle + \lim_{\delta\to 0} \liminf_{n\to 0} \frac{1}{n} \log \nu_n(B_{y,\delta})$$

$$\geq -\Lambda^*(y) + \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \nu_n(B_{y,\delta}), \tag{2.16}$$

όπου η τελευταία ανισότητα έχει προκύψει από τον ορισμό της Λ^*.

Αυτό που πρέπει να παρατηρήσουμε τώρα είναι ότι από τη στιγμή που έχουμε αφαιρέσει την υπόθεση της ανεξαρτησίας, ο ασθενής νόμος των μεγάλων αριθμών δεν ισχύει πια, συνεπώς θα πρέπει από εδώ και πέρα να ακολουθήσουμε διαφορετική στρατηγική.

Η στρατηγική μας θα είναι να χρησιμοποιήσουμε το άνω φράγμα που αποδείξαμε στο (α) μέρος . Προς αυτό το σκοπό θα δείξουμε πρώτα ότι η ν_n ικανοποιεί την υπόθεση που κάναμε στην αρχή της παραγράφου με λογαριθμική ροπογεννήτρια τη συνάρτηση, $\tilde{\Lambda} := \Lambda(\cdot+\eta)-\Lambda(\eta)$. Πράγματι, $\tilde{\Lambda}(0) = 0$ και εφόσον $\eta \in D_\Lambda{}^o$, έπεται ότι $\tilde{\Lambda}(\theta) < \infty$ για κάθε $|\theta|$ αρκετά μικρό. Έστω ότι με $\tilde{\Lambda}_n(\cdot)$ συμβολίζουμε τη λογαριθμική ροπογεννήτρια που αντιστοιχεί στην κατανομή ν_n. Τότε για κάθε $\theta \in \mathbb{R}^d$,

$$\frac{1}{n}\tilde{\Lambda}_n(n\theta) := \frac{1}{n}\log\left[\int_{\mathbb{R}^d} e^{n\langle\theta,z\rangle} d\nu_n(z)\right] = \frac{1}{n}\Lambda_n(n(\theta+\eta)) - \lambda_n(n\eta) \to \tilde{\Lambda}(\theta),$$

επειδή $\Lambda_n(n\eta) < \infty$ για κάθε n αρκετά μεγάλο. Ορίζουμε

$$\tilde{\Lambda}^*(x) := \sup_{\theta\in\mathbb{R}^d} = \Lambda^*(x) - \langle\eta, x\rangle + \Lambda(\eta). \tag{2.17}$$

Εφόσον η υπόθεση 1 ισχύει για την ν_n, εφαρμόζοντας το Λήμμα 2 για την $\tilde{\Lambda}$ έπεται ότι η $\tilde{\Lambda}^*$ είναι καλή συνάρτηση ταχύτητας. Επιπλέον από το (α) μέρος αυτής της απόδειξης έχουμε ότι ένα άνω φράγμα της μορφής (2.11) θα ισχύει για την ν_n με συνάρτηση ταχύτητας την $\tilde{\Lambda}^*$ η οποία είναι καλή. Πιο συγκεκριμένα, για το κλειστό σύνολο $B^c_{y,\delta}$ η σχέση (2.11) δίνει

$$\limsup_{n\to\infty}\frac{1}{n}\log\nu_n(B^c_{y,\delta}) \leq -\inf_{x\in B^c_{y,\delta}}\tilde{\Lambda}^*(x) = -\tilde{\Lambda}^*(x_0),$$

για κάποιο $x_0 \neq y$, όπου η ισότητα προκύπτει επειδή $\tilde{\Lambda}^*$ είναι καλή, Επιπλέον το y είναι εκτεθειμένο σημείο της Λ^* και το η είναι εκτεθειμένο υπερεπίπεδο. Επομένως, εφόσον $\tilde{\Lambda}^*(y) \geq [\langle\eta, y\rangle\Lambda(y)]$ και $x_0 \neq y$, θα έχουμε

$$\tilde{\Lambda}^*(x_0) \geq [\Lambda^*(x_0) - \langle\eta, x_0\rangle] - [\Lambda^*(y) - \langle\eta, y\rangle] > 0.$$

Συνεπώς για κάθε $\delta > 0$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\nu_n(B^c_{y,\delta}) < 0.$$

Από αυτήν την ανισότητα έπεται ότι για κάθε $\delta > 0$ έχουμε $\nu_n(B^c_{y,\delta}) \to 0$ επομένως $\nu_n(B_{y,\delta}) \to 1$. Πιο συγκεκριμένα, θα έχουμε

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\nu_n(B_{y,\delta}) = 0.$$

Επομένως το ζητούμενο ισχύει από τη σχέση (2.16).

(γ) Έχοντας δείξει τα μέρη (α) και (β) του θεωρήματος αρκεί τώρα να αποδείξουμε ότι για κάθε ανοιχτό σύνολο G,

$$\inf_{x \in G \bigcap \mathrm{ri} D_\Lambda^*} \Lambda^*(x) \leq \inf_{x \in G} \Lambda^*(x).$$

Εάν $G \bigcap D_\Lambda^* = \emptyset$ τότε δεν έχουμε κάτι να αποδείξουμε. Επομένως υποθέτουμε ότι $D_{\Lambda^*} \neq \emptyset$, το οποίο σημαίνει ότι $\exists z \in \mathrm{ri} D_\Lambda^*$. Συνεπώς εάν σταθεροποιήσουμε ένα $y \in G \bigcap D_\Lambda^*$ θα έχουμε ότι για κάθε αρκούντως μικρό $\alpha > 0$ ισχύει,

$$\alpha z + (1-\alpha) y \in G \bigcap \mathrm{ri} D_\Lambda^*,$$

άρα

$$\inf_{x \in G \bigcap \mathrm{ri} D_\Lambda^*} \Lambda^*(x) \leq \lim_{\alpha \searrow 0} \Lambda^*(\alpha z + (1-\alpha) y) \leq \Lambda^*(y),$$

εφόσον το y είναι αυθαίρετο η απόδειξη έχει ολοκληρωθεί. □

Σχόλιο 21. *Παρόλο που οι αποδείξεις που έχουμε παρουσιάσει είναι για τον $\mathbb{R}^d$, ουσιωδώς δηλαδή για χώρους πεπερασμένης διάστασης, με μικρές αλλαγές τα επιχειρήματα που χρησιμοποιήσαμε ισχύουν και στη γενική περίπτωση ενός απειροδιάστατου πολωνικού χώρου.*

2.4 Το Λήμμα Varadhan

Το λήμμα Varadhan είναι ένα πολύ σημαντικό αποτέλεσμα της θεωρίας των μεγάλων αποκλίσεων, το οποίο έχει πολύ γενικό χαρακτήρα και μπορεί να αποδειχθεί για έναν τυχαίο τοπολογικό χώρο, ωστόσο θα μείνουμε πιστοί στο χαρακτήρα της εργασίας έως τώρα και θα το παρουσιάσουμε για τον $\mathbb{R}^d$. Τα επιχειρήματα που θα χρησιμοποιήσουμε είναι τα ίδια που θα χρησιμοποιούσαμε για την απόδειξη σε έναν οποιονδήποτε πολωνικό χώρο.

Θεώρημα 4 (Λήμμα Varadhan)**.** *Έστω η ακολουθία μέτρων πιθανότητας $\{\mu_n\}$ ικανοποιεί την LDP στον $\mathbb{R}^d$ με συνάρτηση ταχύτητας I και $h : \mathbb{R}^d \to \mathbb{R}$ είναι συνεχής και φραγμένη τότε,*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) = - \inf_{x\in\mathbb{R}^d} [h(x) + I(x)].$$

Απόδειξη. Θα ξεκινήσουμε αποδεικνύοντας το κάτω φράγμα δηλαδή,

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \geq - \inf_{x\in\mathbb{R}^d} [h(x) + I(x)].$$

Σταθεροποιούμε ένα $x_0 \in \mathbb{R}^d$, τότε αρκεί να δείξουμε ότι

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \geq -h(x_0) - I(x_0).$$

Προφανώς, εάν $I(x_0) = \infty$ δεν έχουμε να αποδείξουμε κάτι, συνεπώς υποθέτουμε ότι $I(x_0) < \infty$. Επιπλέον σταθεροποιούμε ένα $\epsilon > 0$ και θα αποδείξουμε ότι

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \geq -h(x_0) - I(x_0) - \epsilon.$$

Εφόσον το $\epsilon > 0$ που πήραμε είναι αυθαίρετο αυτό το επιχείρημα είναι αρκετό για την ολοκλήρωση της απόδειξης. Έστω $G \subset \mathbb{R}^d$ ανοιχτό τέτοιο ώστε $x_0 \in G$ και

$$h\big|_G \leq h(x_0) + \epsilon,$$

όπου με $h\big|_G$ συμβολίζουμε τον περιορισμό της h πάνω στο σύνολο G. Επιπλέον πρέπει να σημειωθεί ότι ένα τέτοιο σύνολο υπάρχει γιατί η h είναι συνεχής συνάρτηση. Επομένως θα έχουμε,

$$\int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \geq \int_{\mathbb{R}^d} e^{-n(h(x_0)+\epsilon)} d\mu_n(x) + \int_{\mathbb{R}^d/G} e^{-nh(x_0)} d\mu_n(x) \geq$$

$$\geq e^{-n(h(x_0)+\epsilon)} \mu_n(G).$$

Επομένως

$$\lim_{n\to\infty} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \geq -h(x_0) - \epsilon + \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \underset{LDP}{\geq}$$

$$\geq -h(x_0) - \epsilon - I(G) \geq -h(x_0) - \epsilon + I(x_0),$$

εφόσον $I(x_0) \geq I(G)$ και αυτό ολοκληρώνει το ζητούμενο.

Περνάμε τώρα στην απόδειξη του άνω φράγματος, δηλαδή

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \leq \inf_{x\in\mathbb{R}^d} [h(x) + I(x)].$$

Εφόσον η h είναι φραγμένη θα υπάρχει ένα $M \in \mathbb{R}$ τέτοιο ώστε $-M \leq h(x) \leq M, \forall x \in \mathbb{R}^d$. Σε αυτήν την περίπτωση σταθεροποίουμε ένα $\epsilon > 0$ και κατασκεύαζουμε μια διαμέριση

$$\alpha_0 = -M < \alpha_1 < \alpha_2 < ... < \alpha_k = M, \quad \text{τέτοια ώστε} \quad \alpha_i - \alpha_{i-1} < \epsilon, \forall i \in \{1, 2, ..., k\}.$$

Επιπλέον ορίζουμε το σύνολο,

$$F_i = \{y \in \mathbb{R}^d : \alpha_{i-1} \leq h(y) \leq \alpha_i\}, \quad i \in \{1, 2, ..., k\}.$$

Εφόσον η h είναι συνεχής το F_i θα είναι κλειστό σύνολο για κάθε $i \in \{1, 2, ..., k\}$. Παρατηρούμε τώρα ότι θα έχουμε $\mathbb{R}^d = \bigcap_{i=1}^k F_i$, εφόσον τα F_i δεν είναι ξένα θα έχουμε

$$\int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \leq \sum_{i=1}^k \int_{F_i} e^{-nh(x)} d\mu_n(x) \underset{h(x)\geq\alpha_i, x\in F_i}{\leq} \sum_{i=1}^k \int_{F_i} e^{-n\alpha_{i-1}} d\mu_n(x) =$$

$$= \sum_{i=1}^k e^{-n\alpha_{i-1}} \mu_n(F_i).$$

Επομένως,

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \leq \limsup_{n\to\infty} \frac{1}{n} \log(\sum_{i=1}^k e^{-n\alpha_{i-1}} \mu_n(F_i)) =$$

$$\max_{1\leq i\leq k} [\limsup_{n\to\infty} \frac{1}{n} \log(e^{-n\alpha_{i-1}} \mu_n(F_i))] = \max_{1\leq i\leq k} [-\alpha_{i-1} + \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F_i)] \leq$$

$$\underset{LDP}{\leq} \max_{1\leq i\leq k} [-\alpha_{i-1} - I(F_i)] = \max_{1\leq i\leq k} [\sup_{x\in F_i} (-\alpha_{i-1} - I(x))], \tag{2.18}$$

εφόσον $I(F_i) = \inf_{x\in F_i} I(x)$.

Σταθεροποιούμε τώρα ένα $i \in \{1, ..., k\}$ και παρατηρούμε ότι

$$x \in F_i \Rightarrow \alpha_{i-1} \leq h(x) \leq \alpha_i.$$

Επιπλέον,

$$\alpha_i - \alpha_{i-1} < \epsilon \Rightarrow \forall x \in F_i, \quad h(x) < \alpha_i < \alpha_{i-1} + \epsilon \Rightarrow -\alpha_{i-1} < -h(x) + \epsilon$$

Επομένως,

$$\begin{aligned} -\alpha_{i-1} - I(x) < -h(x) + \epsilon - I(x), \quad \forall x \in F_i \Rightarrow \\ \Rightarrow \sup_{x \in F_i}(-\alpha_{i-1} - I(x)) \leq \sup_{x \in F_i}(-h(x) + \epsilon - I(x)) \Rightarrow \\ \Rightarrow \sup_{x \in F_i}(-\alpha_{i-1} - I(x)) \leq -\inf_{x \in F_i}(h(x) + I(x)) + \epsilon. \end{aligned}$$

Συνεπώς, συνδυάζοντας το φράγμα της σχέσης (2.18) με το φράγμα που μόλις υπολογίσαμε θα πάρουμε,

$$\begin{aligned} \limsup_{n \to \infty} \frac{1}{n} \log \int_{\mathbb{R}^d} e^{-nh(x)} d\mu_n(x) \leq \max_{1 \leq i \leq k} [\sup_{x \in F_i}(-\alpha_{i-1} - I(x))] \leq \\ \leq \max_{1 \leq i \leq k} [-\inf_{x \in F_i}(h(x) + I(x)) + \epsilon] = -\min_{1 \leq i \leq k} [\inf_{x \in F_i}(h(x) + I(x))] + \epsilon = \\ = -\inf_{x \in \mathbb{R}^d}[h(x) + I(x)], \end{aligned}$$

κι εφόσον το $\epsilon > 0$ είναι αυθαίρετο το ζητούμενο έχει αποδειχθεί. $\square$

Σχόλιο 22. *(α) Παρατηρούμε ότι για την απόδειξη του άνω φράγματος στο λήμμα Varadhan χρησιμοποιήσαμε το άνω φράγμα της LDP, ενώ για το κάτω φράγμα αξιοποιήσαμε το κάτω φράγμα της LDP.*

(β) Η υπόθεση του φραγμένου της h μας χρειάστηκε μόνο στην απόδειξη του άνω φράγματος.

2.5 Το θεώρημα Sanov & Μέθοδος Τύπων

Σε αυτήν την ενότητα θεωρούμε ότι όλες οι τυχαίες μεταβλητές ενδιαφέροντος παίρνουν τιμές σε ένα πεπερασμένο σύνολο $\Sigma = \{\alpha_1, ..., \alpha_N\}$, θα δούμε τότε ότι χρησιμοποιώντας συνδυαστικές τεχνικές μπορούμε να εξάγουμε πολύ χρήσιμα αποτελέσματα.

Ορισμός 11. *Έστω ακολουθία τυχαίων μεταβλητών $X_1, X_2, ..., X_n$ θα ονομάζουμε εμπειρικό μέτρο που παράγεται από τις $X_1, X_2, ..., X_n$ την ποσότητα*

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

όπου δ_x είναι το μέτρο Dirac.

Σχόλιο 23. *Έστω A μετρήσιμο σύνολο τότε,*

$$\delta_x(A) = \mathbb{1}_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}$$

Παρατηρούμε τώρα ότι το L_n είναι ένα μέτρο πιθανότητας πάνω στο Σ. Επεκτείνουμε αυτήν την παρατήρηση και ορίζουμε το σύνολο

$$\mathcal{M}(\Sigma) = \{\nu = (\nu_1, ..., \nu_N) \in [0,1]^N : \sum_{i=1}^{N} \nu_i = 1\},$$

το οποίο είναι ουσιαστικά το σύνολο όλων των μέτρων πιθανότητας που μπορούν να οριστούν πάνω στο Σ, επομένως $L_n \in \mathcal{M}(\Sigma)$. Δεν είναι δύσκολο να δούμε ότι το $\mathcal{M}(\Sigma)$ μπορεί με φυσιολογικό τρόπο να ταυτιστεί με το σύνολο των $|\Sigma|-$διάστατων πραγματικών διανυσμάτων των οποίων τα στοιχεία είναι μη αρνητικά και αθροίζονται στο 1, το σύνολο αυτό θα το συμβολίζουμε με $\mathbb{R}^{|\Sigma|}$.

Ορισμός 12. *Ο τύπος L_n^y του διανύσματος $y = (y_1, ..., y_n) \in |\Sigma|^n$ είναι το εμπειρικό μέτρο που παράγεται από αυτό το διάνυσμα, δηλαδή $L_n^y = (L_n^y(\alpha_1), ..., L_n^y(\alpha_N))$ είναι το στοιχείο του $\mathcal{M}(\Sigma)$ για το οποίο*

$$L_n^y(\alpha_i) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\alpha_i}(y_k) \quad \text{όπου} \quad i = 1, .., |\Sigma|.$$

Δηλαδή το $L_n^y(\alpha_i)$ είναι το κλάσμα των εμφανίσεων του α_i στο y. Θα συμβολίζουμε με $\mathcal{L}_n$ το σύνολο όλων των δυνατών τύπων διανυσμάτων μήκους n,τότε $\mathcal{L}_n = \{\nu : \nu = L_n^y, \text{για κάποιο } y\} \subset \mathbb{R}^{|\Sigma|}$.

Λήμμα 9. *(α) Ισχύει:* $|\mathcal{L}_n| \leq (n+1)^{|\Sigma|}$.

(β) Για κάθε μέτρο πιθανότητας $\nu \in \mathcal{M}(\Sigma)$ ισχύει:

$$d_V(\nu, \mathcal{L}_n) \triangleq \inf_{\nu' \in \mathcal{L}_n} d_V(\nu, \nu') \leq \frac{|\Sigma|}{2n}$$

όπου για τη μετρική ισχύει $d_V(\nu, \nu') = \sup_{A \subset \Sigma} (\nu(A) - \nu'(A))$.

Απόδειξη. (α) Κάθε στοιχείο του διανύσματος L_n^y ανήκει στο σύνολο $\{0, \frac{1}{n}, ... \frac{n-1}{n}, 1\}$, του οποίου η πληθικότητα είναι $n+1$. Επιπλέον το διάνυσμα L_n^y καθορίζεται από το πολύ $|\Sigma|$ ποσότητες άρα το ζητούμενο προκύπτει κατά προφανη τρόπο.

(β) Παρατηρούμε ότι το σύνολο $\mathcal{L}_n$ περιέχει όλα τα μέτρα πιθανότητας τα οποία αποτελούνται από $|\Sigma|$ συντεταγμένες οι οποίες προέρχονται από το σύνολο $\{0, \frac{1}{n}, ... \frac{n-1}{n}, 1\}$ επομένως για κάθε $\nu \in \mathcal{M}(\Sigma)$ υπάρχει $\nu' \in \mathcal{L}_n$ τέτοιο ώστε

$$|\nu(\alpha_i) - \nu'(\alpha_i)| < \frac{1}{n}, i = 1, ..., |\Sigma|,$$

αρα το ζητούμενο προκύπτει με απλή άθροιση:

$$d_V(\nu, \nu') = \frac{1}{2} \sum_{i=1}^{|\Sigma|} |\nu(\alpha_i) - \nu'(\alpha_i)| \leq \frac{|\Sigma|}{2n}$$

□

Σχόλιο 24. *Από το παραπάνω λήμμα συμπαιρένουμε δύο πράγματα:*

(α) Η πληθικότητα του συνόλου $\mathcal{L}_n$ μεγαλώνει πολυωνυμικά ως προς το n.

(β) Για αρκετά μεγάλο n τα $\mathcal{L}_n$ προσεγγίζουν όσο καλά θέλουμε οποιοδήποτε μέτρο στο $\mathcal{M}(\Sigma)$.

Θα συνεχίσουμε τη μελέτη μας δίνοντας τους εξής ορισμούς:

Ορισμός 13. *Θα ονομάζουμε κλάση τύπου $T_n(\nu)$ ενός διανύσματος πιθανότητας $\nu \in \mathcal{L}_n$ το σύνολο $T_n(\nu) = \{y \in \Sigma^n : L_n^y = \nu\}$.*

Ουσιαστικά η χρησιμότητα του $T_n(\nu)$ έγκειται στο γεγονός ότι αποτελείται από όλες τις δυνατές μεταθέσεις του διανύσματος y.

Ορισμός 14. *(α) Θα ονομάζουμε εντροπία ενός διανύσματος πιθανότητας ν την ποσότητα*

$$H(\nu) := -\sum_{i=1}^{|\Sigma|} \nu(\alpha_i) \log \nu(\alpha_i).$$

(β) Θα ονομάζουμε σχετική εντροπία ενός διανύσματος πιθανότητας ν ως προς ένα διάνυσμα πιθανότητας μ την ποσότητα

$$H(\nu|\mu) := -\sum_{i=1}^{|\Sigma|} \nu(\alpha_i) \log \frac{\nu(\alpha_i)}{\mu(\alpha_i)}.$$

Σχόλιο 25. *Η συνάρτηση $H(\cdot|\mu)$ είναι μια καλή συνάρτηση ταχύτητας. Εφαρμόζοντας την ανισότητα Jensen στην κυρτή συνάρτηση $x \log x$ έπεται ότι η $H(\cdot|\mu)$ είναι μη αρνητική. Επιπλέον η $H(\cdot|\mu)$ είναι πεπερασμένη και συνεχής πάνω στο συμπαγές σύνολο $\{\nu \in \pm : \Sigma_\nu \subset \Sigma_\mu\}$, επειδή η $x \log x$ είναι συνεχής για $0 \leq x \leq 1$. Επιπλέον $H(\cdot|\mu) = \infty$ έξω από αυτό το σύνολο, επομένως είναι καλή συνάρτηση ταχύτητας.*

Στα λήμματα που ακολουθούν υπολογίζουμε τις πιθανότητες των ενδεχομένων της μορφής $\{L_n^y = \nu\}, \nu \in \mathcal{L}_n$. Αρχικά δείχνουμε ότι τα στοιχεία της ίδιας κλάσης τύπου είναι ισοπίθανα και στη συνέχεια υπολογίζεται ο εκθετικός ρυθμός αύξησης του μεγέθους της κάθε κλάσης τύπου.

Στις παρακάτω διατυπώσεις θα συμβολίζουμε με $\mathbb{P}_\mu$ την από κοινού κατανομή μιας ακολουθίας ανεξάρτητων και ισόνομων τυχαίων μεταβλητών Y_i οι οποίες κατανέμονται σύμφωνα με το μέτρο $\mu \in \mathcal{M}(\Sigma)$.

Λήμμα 10. *Έστω $y \in T_n(\nu), \nu \in \mathcal{L}_n$ τότε*

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = e^{-n[H(\nu)+H(\nu|\mu)]}.$$

Απόδειξη. Έστω $Y = (Y_1, ..., Y_n)$ τότε το εμπειρικό μέτρο L_n^Y το οποίο παράγεται από το τυχαίο διάνυσμα Y, συγκεντρώνεται σε τύπους $\nu \in \mathcal{L}_n$ για τους οποίους ισχύει $\Sigma_\nu \subset \Sigma_\mu$, δηλαδή $H(\nu|\mu) < \infty$. Οπότε χωρίς βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι $L_n^y = \nu$ και $\Sigma_\nu \subset \Sigma_\mu$ τότε,

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = \prod_{i=1}^{|\Sigma|} \mu(\alpha_i)^{n\nu(\alpha_i)} = e^{-n[H(\nu)+H(\nu|\mu)]}$$

Όπου η τελευταία ισότητα έχει προκύψει από την ταυτότητα:

$$H(\nu) + H(\nu|\mu) = -\sum_{i=1}^{|\Sigma|} \nu(\alpha_i) \log \nu(\alpha_i).$$

□

Πιο συγκεκριμένα επειδή $H(\mu|\mu) = 0$ για κάθε $\mu \in \mathcal{L}_n$ και κάθε $y \in T_n(\mu)$ έπεται ότι:

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = e^n - H(\mu)$$

Λήμμα 11. *Για κάθε $\nu \in \mathcal{L}_n$ ισχύει:*

$$(n+1)^{|\Sigma|} e^{nH(\nu)} \leq |T_n(\nu)| \leq e^{nH(\nu)}.$$

Απόδειξη. Κάτω από το μέτρο $\mathbb{P}_\nu$ κάθε κλάση έχει πιθανότητα το πολύ 1 και επιπλέον κάθε πραγματοποίηση είναι ισοπίθανη, οπότε για κάθε $\nu \in \mathcal{L}_n$ μπορούμε να γράψουμε:

$$1 \geq \mathbb{P}_\nu(L^Y_\nu) = \mathbb{P}_\nu((Y_1, ..., Y_n) \in T_n(\nu)) = e^{-nH(\nu)}|T_n(\nu)|$$

άρα έχουμε αποδείξει το άνω φράγμα. Στρέφουμε τώρα την προσοχή μας στο κάτω φράγμα.

Έστω $\nu' \in \mathcal{L}_n$ τέτοιο ώστε $\Sigma_{\nu'} \subset \Sigma_\nu$ τότε μπορούμε να γράψουμε:

$$\frac{\mathbb{P}_\nu(L^Y_n = \nu)}{\mathbb{P}_\nu(L^Y_n = \nu')} = \frac{|T_n(\nu)| \prod_{i=1}^{|\Sigma_\nu|} \nu(\alpha_i)^{n\nu(\alpha_i)}}{|T_n(\nu')| \prod_{i=1}^{|\Sigma_\nu|} \nu(\alpha_i)^{n\nu'(\alpha_i)}} = \prod_{i=1}^{|\Sigma_\nu|} \frac{(n\nu'(\alpha_i))!}{(n\nu(\alpha_i))!} \nu(\alpha_i)^{n\nu(\alpha_i) - n\nu'(\alpha_i)}$$

Η τελευταία έκφραση που έχουμε φτάσει περιέχει όρους της μορφής $\frac{m!}{k!}\frac{k}{n}^{k-m}$, δεν είναι δύσκολο τώρα να παρατηρήσουμε, (έαν κάνουμε τον έλεγχο για τις περιπτώσεις $m \geq k$ και $m < k$), ότι για κάθε $m, k \in \mathbb{N}$ ισχύει $\frac{m!}{k!} \geq k^{m-k}$, οπότε από την παραπάνω ισότητα έχουμε:

$$\frac{\mathbb{P}_\nu(L^Y_n = \nu)}{\mathbb{P}_\nu(L^Y_n = \nu')} \geq \prod_{i=1}^{|\Sigma_\nu|} n^{n\nu'(\alpha_i) - n\nu(\alpha_i)} = 1$$

τώρα αξίζει να παρατηρήσουμε ότι $\mathbb{P}_\nu(L^Y_n = \nu') > 0$ μόνο όταν $\Sigma_{\nu'} \subset \Sigma_\nu$ και $\nu' \in \mathcal{L}_n$ άρα από τα παραπάνω έχει νόημα να γράψουμε $\mathbb{P}_\nu(L^Y_n = \nu) \geq \mathbb{P}_\nu(L^Y_n = \nu')$ για κάθε $\nu, \nu' \in \mathcal{L}_n$, άρα έχουμε:

$$1 = \sum_{\nu' \in \mathcal{L}_n} \mathbb{P}_\nu(L^Y_n = \nu') \leq |\mathcal{L}_n| \mathbb{P}_\nu(L^Y_n = \nu') = |\mathcal{L}_n| e^{-H(\nu)} |T_n(\nu)|$$

και από το Λήμμα 9 προκύπτει το κάτω φράγμα που αναζητούσαμε. Συνεπώς έχουμε ολοκληρώσει το ζητούμενο. □

Λήμμα 12. *(Μεγάλες Αποκλίσεις) Για κάθε $\nu \in \mathcal{L}_n$ ισχύει:*

$$(n+1)^{-|\Sigma|}e^{-nH(\nu|\mu)} \le \mathbb{P}_\mu(L_n^Y = \nu) \le e^{-nH(\nu|\mu)}.$$

Απόδειξη. Από το Λήμμα 9 έχουμε:

$$\mathbb{P}_\mu(L_n^Y) = |T_n(\nu)|\mathbb{P}_\mu((Y_1, ..., Y_n) = y, L_n^y = \nu) = |T_n(\nu)|e^{-nH(\nu)+H(\nu|\mu)}$$

τώρα το ζητούμενο προκύπτει κατά προφανή τρόπο εάν εφαρμόσουμε το Λήμμα 10. □

Είμαστε τώρα σε θέση να παρουσιάσουμε το θεώρημα Sanov.

Θεώρημα 5 (Sanov)**.** *Για κάθε σύνολο Γ διανυσμάτων πιθανότητας τα οποία ανήκουν στο $\mathcal{M}(\Sigma)$ ισχύει:*

$$-\inf_{\nu\in\Gamma^o} H(\nu|\mu) \le \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\mu(L_n^Y \in \Gamma) \le \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\mu(L_n^Y \in \Gamma) \le -\inf_{\nu\in\Gamma} H(\nu|\mu).$$

Σχόλιο 26. *Βλέπουμε εδώ πως τα απολέσματα που έχουμε παρουσιάσει σε αυτό το κεφάλαιο δένουν με τα εισαγωγικά. Στην ουσία το θεώρημα Sanov μας λέει ότι η ακολουθία μέτρων $\mathbb{P}_\mu(L_n^Y \in \cdot)$ ακολουθεί την LDP με συνάρτηση ταχύτητας την $H(\cdot|\mu)$.*

Απόδειξη. Έχουμε

$$\mathbb{P}_\mu(L_n^Y \in \Gamma) = \sum_{\nu\in\Gamma\cap\mathcal{L}_n} \mathbb{P}_\mu(L_n^Y = \nu) \le \sum_{\nu\in\Gamma\cap\mathcal{L}_n} e^{-nH(\nu|\mu)}$$

και από το Λήμμα 11 :

$$\mathbb{P}_\mu(L_n^Y \in \Gamma) \le |\Gamma\cap\mathcal{L}_n|e^{-n\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)} \le (n+1)^{|\Sigma|}e^{-n\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)}.$$

Επιπλέον

$$\mathbb{P}_\mu(L_n^Y \in \Gamma) = \sum_{\nu\in\Gamma\cap\mathcal{L}_n} \mathbb{P}_\mu(L_n^Y = \nu) \ge \sum_{\nu\in\Gamma\cap\mathcal{L}_n} (n+1)^{-|\Sigma|}e^{-nH(\nu|\mu)} \ge$$
$$\ge (n+1)^{-|\Sigma|}e^{-n\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)}$$

Ακόμη έχουμε $\lim_{n\to\infty}\frac{1}{n}\log(n+1)^{|\Sigma|} = 0$, άρα το κανονικοποιημένο λογαριθμικό όριο των παραπάνω σχέσεων θα είναι:

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\mu(L_n^Y \in \Gamma) = -\liminf_{n\to\infty}\{\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\}$$

και

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\mu(L_n^Y \in \Gamma) = -\limsup_{n\to\infty}\{\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\}$$

Το ζητούμενο άνω όριο προκύπτει τώρα επειδή για κάθε n ισχύει $\Gamma \cap \mathcal{L}_n \subset \Gamma$. Για το κάτω όριο θα έχουμε, έστω ν ένα εσωτερικό σημείο του Γ τέτοιο ώστε $\Sigma_\nu \subset \Sigma_\mu$, τότε για κάποιο αρκετά μικρό $\delta > 0$ έχουμε ότι το σύνολο $\{\nu' : d_V(\nu, \nu') < \delta\}$ περιέχεται στο Γ, τότε όμως από το (β) του Λήμματος 9 θα υπάρχει μια ακολουθία $\{\nu_n\} \in \Gamma \cap \mathcal{L}_n$ τέτοια ώστε $\nu_n \to \nu$ καθώς το $n \to \infty$. Χωρίς βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι $\Sigma_{\nu_n} \subset \Sigma_\mu$ τότε,

$$-\limsup_{n\to\infty}\{\inf_{\nu' \in \Gamma\cap\mathcal{L}_n} H(\nu'|\mu)\} \geq -\lim_{n\to\infty} H(\nu_n|\mu) = -H(\nu|\mu).$$

Τώρα αξίζει να παρατηρήσουμε ότι $H(\nu|\mu)) = \infty$ όταν $\nu(\alpha_i) > 0$ και $\mu(\alpha_i) = 0$ οπότε η προηγούμενη ανισότητα θα μας δώσει :

$$-\limsup_{n\to\infty}\{\inf_{\nu \in \Gamma\cap\mathcal{L}_n} H(\nu|\mu)\} \geq -\inf_{\nu\in\Gamma^o} H(\nu|\mu)$$

και το ζητούμενο προκύπτει με απλή αντικατάσταση από τα παραπάνω. □

2.5.1 Σύνδεση Θ. Sanov και Θ. Cramér

Διαισθητικά θα πρέπει να μας φαίνεται ξεκάθαρο ότι η LDP για τον εμπειρικό μέσο $\bar{X}_n$ (Θεώρημα Cramér) μπορεί να εξαχθεί μέσω της LDP για το εμπειρικό μέτρο που παράγεται από τις $X_1, ..., X_n$ (Θεώρημα Sanov). Αυτή η σύνδεση είναι το αντικείμενο αυτής της ενότητας. Παρακάτω θα δούμε πως μια εκδοχή του θεωρήματος του Cramér προκύπτει ως εφαρμογή του θεωρήματος Sanov.

Πιο συγκεκριμένα μελετάμε την ακολουθία εμπειρικών μέσων $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$, όπου $X_j = f(Y_j)$ με $f : \Sigma \longrightarrow \mathbb{R}$ και Y_j είναι ισόνομες και ανεξάρτητες με κατανομή μ η οποία έχει τις ιδιότητες που περιγράψαμε στην προηγούμενη παράγραφο. Χωρίς βλάβη της γενικότητας υποθέτουμε επιπλέον ότι $\Sigma = \Sigma_\mu$ και $f(\alpha_1) < ... < f(\alpha_{|\Sigma|})$.

Σημειώνουμε ότι στην περίπτωση που μελετάμε εδώ οι $\bar{X}_n$ παίρνουν τιμές στο συμπαγές σύνολο $K := [f(\alpha_1), ..., f(\alpha_{|\Sigma|})]$. Επιπλέον $\bar{X}_n = \langle \mathbf{f}, L_n^Y \rangle$, όπου $\mathbf{f} = (f(\alpha_1), ..., f(\alpha_{|\Sigma|}))$, επομένως για κάθε σύνολο και κάθε $n \in \mathbb{N}$,

$$\bar{X}_n \in A \Leftrightarrow L_n^Y \in \{\nu : \langle \mathbf{f}, \nu \rangle \in A\} = \Gamma.$$

Θεώρημα 6. *Για κάθε σύνολο $A \subset \mathbb{R}$ ισχύει*

$$-\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in A) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in A) \leq -\inf_{x\in A} I(x) \quad (2.19)$$

όπου η συνάρτηση ταχύτητας $I(x) = \inf_{\{\nu:\langle \mathbf{f},\nu\rangle = x\}} H(\nu|\mu)$. Η $I(x)$ είναι συνεχής για $x \in K$ και ισχύει

$$I(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \Lambda(\theta)\}, \quad (2.20)$$

όπου

$$\Lambda(\theta) = \log \sum_{i=1}^{|\Sigma|} \mu(\alpha_i) e^{-\theta f(\alpha_i)}.$$

Σχόλιο 27. *Εφόσον η συνάρτηση ταχύτητας I είναι συνεχής στο K προκύπτει από το θεώρημα Cramer ότι $A \subset \bar{A}^o \subset K$ τότε,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\bar{X}_n \in A) = -\inf_{x\in A} I(x).$$

Απόδειξη. Από την ανισότητα Jensen για κάθε $\nu \in \mathcal{M}(\Sigma)$ και κάθε $\theta \in \mathbb{R}$ θα έχουμε:

$$\Lambda(\theta) = \log \sum_{i=1}^{|\Sigma|} \mu(\alpha_i) e^{\theta f(\alpha_i)} \geq \sum_{i=1}^{|\Sigma|} \nu(\alpha_i) \log \frac{\mu(\alpha_i) e^{\theta f(\alpha_i)}}{\nu(a_i)} = \theta \langle \mathbf{f}, \nu \rangle - H(\nu|\mu)$$

με την ισότητα να συμβαίνει για $\nu_\theta(\alpha_i) = \mu(\alpha_i) e^{\theta f(\alpha_i) - \Lambda(\theta)}$. Συνεπώς για κάθε θ και για κάθε x ισχύει:

$$\theta x - \Lambda(\theta) \leq \inf_{\nu:\langle \mathbf{f},\nu\rangle = x} H(\nu|\mu) = I(x)$$

μς την ισότητα να συμβαίνει για $x = \langle \mathbf{f}, \nu_\theta \rangle$.

Η συνάρτηση $\Lambda(\theta)$ είναι διαφορίσιμη με $\Lambda'(\theta) = \langle \mathbf{f}, \nu_\theta \rangle$. Συνεπώς η σχέση (2.20) αληθεύει για κάθε $x \in \{\Lambda'(\theta) : \theta \in \mathbb{R}\}$. Παρατηρούμε τώρα ότι η Λ' είναι γνησίως αύξουσα εφόσον η Λ είναι αυστηρά κυρτή και επιπλέον $f(\alpha_1) = \inf_\theta \Lambda'(\theta)$ και $f(\alpha_{|\Sigma|}) = \sup_\theta \Lambda'(\theta)$. Επομένως η (2.20) αληθεύει για κάθε $x \in K^o$. Ας θεωρήσουμε τώρα ως x το άκρο $x = f(\alpha_1)$ του K και έστω $\nu(\alpha_1) = 1$ ώστε $\langle \mathbf{f}, \nu \rangle = x$, τότε

$$-\log \mu(\alpha_1) = H(\nu|\mu) \geq I(x) \geq \sup_\theta \{\theta x - \Lambda(\theta)\} \geq \lim_{\theta\to-\infty} [\theta x - \Lambda(\theta)] = -\log \mu(\alpha_1).$$

Η απόδειξη για το άλλο άκρο είναι εντελώς όμοια, επίσης η συνέχεια της I είναι ευθεία συνέπεια της συνέχειας της συνάρτησης σχετικής εντροπίας. ☐

Σχόλιο 28. *Η σύνδεση της εντροπίας H με τη συνάρτηση ταχύτητας I μας επιτρέπει να αποκτήσουμε μια καλύτερη εποπτεία της φύσης των μεγάλων αποκλίσεων. Μια μεγάλη απόκλιση συμβαίνει με το λιγότερο πιθανό απ΄ όλους τους λιγότερο πιθανούς τρόπους.*

Κεφάλαιο 3

Μεγάλες Αποκλίσεις για Μαρκοβιανές Αλυσίδες

Στο κεφάλαιο αυτό ο σκοπός μας είναι να εξάγουμε εκτιμήσεις μεγάλων αποκλίσεων για μαρκοβιανές αλυσίδες σε πεπερασμένο χώρο καταστάσεων. Η πορεία που θα ακολουθήσουμε είναι εξής, αρχίκα θα δείξουμε πως μπορούμε με χρήση του θεωρήματος Cramér και του θεωρήματος Gärtner-Ellis να εξάγουμε μια LDP για αθροιστικά συναρτησιακά μαρκοβιανών αλυσίδων σε πεπερασμένο χώρο καταστάσεων, ενώ στη συνέχεια θα χρησιμοποιήσουμε επιχειρήματα συνδυαστικής και θα εφαρμόσουμε το θεώρημα Sanov.

Έστω μια Μαρκοβιανή αλυσίδα $X_1, X_2, ..., X_n, ...$ η οποία παίρνει τιμές πάνω σε ένα πεπερασμένο σύνολο Σ. Συνεχίζοντας την πρακτική που ακολουθήσαμε στο προηγούμενο κεφάλαιο θα θεωρήσουμε χωρίς βλάβη της γενικότητας ότι $\Sigma = \{1, .., r\}$ και $|\Sigma| = r$.

Έστω $\mathbf{\Pi}$=$\pi(i, j)$ όπου $i, j = 1, ..., |\Sigma|$, να είναι ένας στοχαστικός πίνακας, δηλαδή ένας πίνακας τα στοιχεία του οποίου είναι μη αρνητικά και το άθροισμα των στοιχείων της κάθε γραμμής ισούται με 1. Έστω Μαρκοβιανή αλυσίδα $X_1, X_2, ..., X_n, ...$ με πίνακα μετάβασης Π και αρχική κατάσταση $\sigma \in \Sigma$ θα συμβολίζουμε με P_σ^π το μέτρο πιθανότητας που προκύπτει από τον Π και τη σ, δηλαδή:

$$P_\sigma^\pi(X_1 = x_1, ..., X_n = x_n) = \pi(\sigma, x_1) \prod_{i=1}^{n-1} \pi(x_i, x_{i-1}).$$

Αντιστοίχως, τη μέση τιμή ως προς το μέτρο P_σ^π θα τη συμβολίζουμε με $\mathbb{E}_\sigma^\pi$.

Έστω $\mathbf{A}$ ένας πίνακας, θα συμβολίζουμε με $\mathbf{A}^m$ την m-οστή δύναμη του πίνακα. Επιπλέον θα λέμε ότι ο πίνακας $\mathbf{A}$ είναι *ανάγωγος*, εάν για κάθε ζεύγος δεικτών i, j υπάρχει φυσικός αριθμός $m = m(i, j)$ τέτοιος ώστε $\mathbf{A}^m(i, j) > 0$. Ουσιαστικά η αναγωγιμότητα του πίνακα $\mathbf{A}$ είναι ισοδύναμη με το να πούμε ότι για κάθε ζεύγος δεικτών i, j είναι δυνατόν να βρούμε μια ακολουθία δεικτών $i_1, ..., i_m$ έτσι ώστε $i_1 = i$ και $i_m = j$ με $\mathbf{A}(i_k, i_{k+1}) > 0$ για κάθε $k = 1, ..., m-1$, σε όρους Μαρκοβιανών αλυσίδων αυτό σημαίνει ότι υπάρχει ακολουθία μεταβάσεων που οδηγεί από την κατάσταση i στην κατάσταση j με θετική πιθανότητα. Το παρακάτω θεώρημα μας δίνει κάποιες βασικές ιδιότητες των ανάγωγων πινάκων.

Θεώρημα 7 (Perron-Frobenius)**.** *Έστω $\mathbf{A}$=$\alpha(i, j)$ όπου $i, j = 1, ..., |\Sigma|$ είναι ένας ανάγωγος πίνακας. Τότε ο $\mathbf{A}$ έχει μια ιδιοτιμή ρ (ιδιοτιμή Perron-Frobenius) τέτοια ώστε:*

(α) Η ρ είναι πραγματική και $\rho > 0$.

(β) Για οποιαδήποτε άλλη ιδιοτιμή λ του $\mathbf{A}$, ισχύει $|\lambda| \leq \rho$.

(γ) Υπάρχουν δεξιά και αριστερά ιδιοδιανύσματα τα οποία αντιστοιχούν στη ρ και έχουν αυστηρά θετικές συντεταγμένες.

(δ) Το δεξί και αριστερό ιδιοδιάνυσμα μ,θ αντίστοιχα είναι μοναδικά έως μια σταθερά.

(ε) Για κάθε $i \in \Sigma$ και κάθε $u = (u_1, ..., u_{|\Sigma|})$ τέτοιο ώστε $u_j > 0$ για κάθε j, θα ισχύει

$$\lim_{n\to\infty} \frac{1}{n} \log[\sum_{i=1}^{|\Sigma|} \boldsymbol{A}^n(i, j)u_j] = \lim_{n\to\infty} \frac{1}{n} \log[\sum_{i=1}^{|\Sigma|} u_j \boldsymbol{A}^n(j, i)] = \log \rho.$$

Απόδειξη. Θα αποδείξουμε την ιδιότητα (ε) η οποία είναι αυτή που θα χρησιμοποιήσουμε άμεσα στη συνέχεια. Έστω $\alpha := \sup_i \theta_i$, $\beta := \inf_i \theta_i$, $\gamma := \sup_j \phi_j$, $\delta := \inf_j \phi_j$, όπου θ είναι το δεξί ιδιοδιάνυσμα το οποίο αντιστοιχεί στη ρ. Τότε για κάθε $i, j \in \Sigma$,

$$\frac{\gamma}{\beta}\mathbf{A}^n(i, j)\theta_j \geq \mathbf{A}^n(i, j)\phi_j \geq \frac{\delta}{\alpha}\mathbf{A}^n(i, j)\theta_j.$$

Επομένως,

$$\lim_{n\to\infty} \frac{1}{n} \log[\sum_{i=1}^{|\Sigma|} \mathbf{A}^n(i, j)\phi_j] = \lim_{n\to\infty} \frac{1}{n} \log[\sum_{i=1}^{|\Sigma|} \mathbf{A}^n(i, j)\theta_j] =$$
$$= \lim_{n\to\infty} \frac{1}{n} \log(\rho^n\theta_i) = \log \rho.$$

Με όμοιο τρόπο παίρνουμε και

$$\lim_{n\to\infty} \frac{1}{n} \log[\sum_{i=1}^{|\Sigma|} \phi_j \mathbf{A}^n(j, i)] = \log \rho.$$

□

Σχόλιο 29. *Για το υπόλοιπο του κεφαλαίου θα χρησιμοποιούμε το συμβολισμό $\phi >> 0$ για να συμβολίζουμε αυστηρώς θετικά διανύσματα, δηλαδή $\phi_j > 0, \quad \forall j \in \Sigma$.*

3.1 LDP για Αθροιστικά Συναρτησοειδή Μαρκοβιανών Αλυσίδων

Το αντικείμενο αυτής της ενότητας είναι η μελέτη συναρτησοειδών της μορφής

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k,$$

(ουσιαστικά είναι ο εμπειρικός μέσος), με $X_k = f(Y_k)$, όπου Y_k είναι τυχαίες μεταβλητές και $f : \Sigma \to \mathbb{R}^d$ είναι μια δοσμένη ντετερμινιστική συνάρτηση. Τώρα εάν οι Y_k είναι ανεξάρτητες τότε και οι X_k θα είναι ανεξάρτητες και κατ΄ επέκταση έπεται από το θεώρημα Cramér ότι ο μετασχηματισμός Fenchel-Legendre της λογαριθμικής ροπογεννήτριας της X_1 θα είναι η συνάρτηση ταχύτητας για την οποία η $\{Z_n\}$ ικανοποιεί την LDP. Από την άλλη το θεώρημα Gärtner-Ellis μας δίνει μια ένδειξη ότι η συνάρτηση ταχύτητας μπορεί να εκφραστεί ως προς κάποιον μετασχηματισμό Fenchel-Legendre ακόμα και όταν οι Y_k δεν είναι ανεξάρτητες αλλά παρουσιάζουν μαρκοβιανή εξάρτηση.

Για να βρούμε μια εναλλακτική αναπαράσταση για τη λογαριθμική ροπογεννήτρια $\Lambda(\theta)$ θα αντιστοιχίσουμε σε κάθε $\theta \in \mathbb{R}^d$ έναν μη αρνητικό πίνακα $\mathbf{\Pi}_\theta$ τα στοιχεία του οποίου ικανοποιούν τη σχέση

$$\pi_\theta(i,j) = \pi(i,j)e^{\langle \theta, f(j)\rangle} \quad i,j \in \Sigma.$$

Επειδή οι ποσότητες $e^{\langle \theta, f(j)\rangle}$ είναι πάντοτε θετικές, εάν ο $\pi(i,j)$ είναι ανάγωγος τότε και ο $\pi_\theta(i,j)$ θα είναι ανάγωγος. Επιπλέον για κάθε $\theta \in \mathbb{R}^d$ θα συμβολίζουμε με $\rho(\mathbf{\Pi}_\theta)$ την ιδιοτιμή Perron-Frobenius του πίνακα $\mathbf{\Pi}_\theta$. Η συγκεκριμένη ιδιοτιμή είναι άκρως σημαντική διότι, όπως θα αποδείξουμε στη συνέχεια θα παίξει το ρόλο της $\Lambda(\theta)$.

Θεώρημα 8. *Έστω μαρκοβιανή αλυσίδα $\{Y_k\}$ με ανάγωγο πίνακα μετάβασης $\mathbf{\Pi}$. Για κάθε $z \in \mathbb{R}^d$, ορίζουμε*

$$I(z) := \sup_{\theta \in \mathbb{R}^d} (\langle \theta, z\rangle - \log \rho(\mathbf{\Pi}_\theta)).$$

Τότε ο εμπειρικός μέσος Z_n ικανοποιεί την LDP με συνάρτηση ταχύτητας την κυρτή I, η οποία είναι επιπλέον καλή. Πιο συγκεκριμένα, για κάθε σύνολο $\Gamma \subset \mathbb{R}^d$ και κάθε αρχική κατάσταση $\sigma \in \Sigma$,

$$\begin{aligned} -\inf_{z \in \Gamma^o} I(z) &\le \liminf_{n\to\infty} \frac{1}{n}\log P_\sigma^\pi(Z_n \in \Gamma) \le \\ &\le \limsup_{n\to\infty} \frac{1}{n}\log P_\sigma^\pi(Z_n \in \Gamma) \le -\inf_{z\in\bar{\Gamma}} I(z). \end{aligned} \tag{3.1}$$

Απόδειξη. Ορίζουμε
$$\Lambda_n(\theta) := \log \mathbb{E}_\sigma^\pi(e^{\langle \theta, Z_n\rangle}).$$
Με βάση το θεώρημα Gärtner-Ellis αρκεί να αποδείξουμε ότι το όριο

$$\Lambda(\theta) := \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\theta) = \lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}_\sigma^\pi(e^{n\langle \theta, Z_n\rangle})$$

υπάρχει για κάθε $\theta \in \mathbb{R}^d$, επιπλέον ότι η Λ είναι πεπερασμένη και παντού διαφορίσιμη στο $\mathbb{R}^d$ και τέλος ότι $\Lambda(\theta) = \log \rho(\mathbf{\Pi}_\theta)$. Αρχικά θα έχουμε

$$\Lambda_n(n\theta) = \log \mathbb{E}_\pi^\sigma(e^{\langle \theta \sum_{k=1}^n X_k}) = \log \sum_{y_1,\ldots,y_n} P_\sigma^\pi(Y_1 = y_1, \ldots, Y_n = y_n)\prod_{k=1}^n e^{\langle \theta, f(y_k)\rangle}) =$$

$$= \log \sum_{y_1,\ldots,y_n} \pi(\sigma, y_1)e^{\langle \theta, f(y_1)\rangle}\ldots\pi(y_{n-1}, y_n)e^{\langle \theta, f(y_n)\rangle} = \log \sum_{y_n=1}^{|\Sigma|}(\mathbf{\Pi}_\theta)^n(\sigma, y_n).$$

Εφόσον ο $\mathbf{\Pi}_\theta$ είναι ανάγωγος από το μέρος (ε) του θεωρήματος Perron-Frobenius με $\phi = (1, \ldots, 1)$ θα πάρουμε
$$\Lambda(\theta) = \lim_{n\to\infty}\frac{1}{n}\Lambda_n(n\theta) = \log \rho(\mathbf{\Pi}_\theta).$$
Επιπλέον, επειδή το $|\Sigma|$ είναι πεπερασμένο η $\rho(\mathbf{\Pi}_\theta)$ όντας απομονωμένη ρίζα της χαρακτηριστικής εξίσωσης του $\mathbf{\Pi}_\theta$ είναι θετική, πεπερασμένη και διαφορίσιμη ως προς θ, επομένως η απόδειξη έχει ολοκληρωθεί.

(Το γεγονός αυτό προκύπτει από γνωστό θεώρημα της θεωρίας πινάκων η απόδειξη μπορεί να βρεθεί στο [8]). □

Σχόλιο 30. *(α) Η προηγούμενη απόδειξη στηρίχθηκε σε δύο ιδιότητες της μαρκοβιανής αλυσίδας, συγκεκριμένα το μέρος (ε) του θεωρήματος Perron-Frobenius και τη διαφορισιμότητα της $\rho(\mathbf{\Pi}_\theta)$ ως προς θ. Επομένως το θεώρημα ισχύει εφόσον η μαρκοβιανή αλυσίδα έχει αυτές τις δύο ιδιότητες.*

(β) Η καλή συνάρτηση ταχύτητας I είναι κυρτή και δεν εξαρτάται από την αρχική κατάσταση $\sigma \in \Sigma$. Και οι δύο αυτές ιδιότητες μπορεί να χαθούν όταν ο πίνακας $\mathbf{\Pi}$ δεν είναι ανάγωγος. Ένα παράδειγμα που εμφανίζεται εξάρτηση από την αρχική κατάσταση είναι το εξής, έστω $\Sigma = \{1, 2\}$ και έχουμε τον πίνακα μετάβασης ,

$$\mathbf{\Pi} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

και $f(j) = \mathbb{1}_2(j)$. Σε αυτήν την περίπτωση

$$\rho(\mathbf{\Pi}_\theta) = e^{\max(\theta - \log 2, 0)} \quad \& \quad Z_n = L_n(2) \to 0 \quad \text{σχεδόν βεβαίως.}$$

Οπότε τώρα, εάν $\sigma = 1$ προφανώς $Y_k = 1$ για όλα τα k και επομένως $Z_n = 0$ ικανοποιεί την LDP με την καλή συνάρτηση ταχύτητας

$$I(z) = \begin{cases} 0, & z = 0 \\ \infty, & z \neq 0 \end{cases}.$$

Από την άλλη για $\sigma = 2$ έχουμε,

$$I(z) = \begin{cases} z \log 2, & z \in [0,1] \\ \infty, & z \notin [0,1] \end{cases}.$$

3.2 Το Θεώρημα Sanov για το Εμπειρικό Μέτρο Μαρκοβιανών Αλυσίδων

Μια εξαιρετικά σημαντική εφαρμογή του θεωρήματος 8 είναι ότι μας δίνει την LDP την οποία ικανοποιεί το εμπειρικό μέτρο $L_n = (L_n(1), ..., L_n(|\Sigma|))$ μιας μαρκοβιανής αλυσίδας. Εδώ το L_n είναι ουσιαστικά το διάνυσμα των συχνοτήτων με τις οποίες η μαρκοβιανή αλυσίδα επισκέφτεται την κάθε κατάσταση, δηλαδή

$$L_n(i) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_i(Y_k), \quad i = 1, ..., |\Sigma|.$$

Έστω ότι ο $\mathbf{\Pi}$ είναι ανάγωγος πίνακας και μ είναι η στάσιμη κατανομή της μαρκοβιανής αλυσίδας, δηλαδή το μοναδικό μη αρνητικό αριστερό ιδιοδιάνυσμα του $\mathbf{\Pi}$ του οποίου οι συνιστώσες αθροίζονται στο 1. Τότε από το εργοδικό θεώρημα έπεται ότι $L_n \to \mu$ κατά πιθανότητα καθώς $n \to \infty$ (για μια αναλυτική παρουσίαση του εργοδικού θεωρήματος ο ενδιαφερόμενος μπορεί να ανατρέξει στο [6]). Επομένως η ακολουθία $\{L_n\}$ φαίνεται να είναι καλή υποψήφιος για να ικανοποιεί την LDP στο $\mathcal{M}(\Sigma)$ με κάποια συνάρτηση ταχύτητας.

Τώρα που έχουμε κάνει αυτήν τη σύντομη προεργασία είναι προφανές ότι $\{L_n\}$ πέφτει κάτω από την ομπρέλα του θεωρήματος 8 εάν πάρουμε $f(y) = (\mathbb{1}_1(y), ..., \mathbb{1}_{|\Sigma|}(y))$, επομένως από το θεώρημα 8 η $\{L_n\}$ ικανοποιεί την LDP με συνάρτηση ταχύτητας την

$$I(q) = \sup_{\theta \in \mathbb{R}^d} (\langle \theta, q \rangle - \log \rho(\mathbf{\Pi}_\theta)),$$

όπου $\pi_\theta(i,j) = \pi(i,j)e^{\theta j}$.

Το παρακάτω θεώρημα μας παρέχει έναν εναλλακτικό χαρακτηρισμό για την $I(q)$.

Θεώρημα 9.

$$I(q) = J(q) := \begin{cases} \sup\limits_{u>>0} \sum_{j=1}^{|\Sigma|} (q_j \log \left[\frac{u_j}{(u\mathbf{\Pi})_j}\right]), & q \in \mathcal{M}(\Sigma) \\ \infty, & q \notin \mathcal{M}(\Sigma). \end{cases}$$

Σχόλιο 31. *(a) Εάν οι τυχαίες μεταβλητές είναι ανεξάρτητες και ισόνομες, τότε οι γραμμές του* $\mathbf{\Pi}$ *ταυτίζονται μεταξύ τους, σε αυτήν την περίπτωση η* $J(q)$ *είναι η σχετική εντροπία* $H(q|\pi(1,\cdot))$.

(β) Η προηγούμενη ταυτότητα ισχύει ακόμα και στην περίπτωση που ο $\mathbf{\Pi}$ *δεν είναι στοχαστικός.*

Απόδειξη. Παρατηρούμε ότι για κάθε n και κάθε διαφορετική πραγματοποίηση της μαρκοβιανής αλυσίδας έχουμε $L_n \in \mathcal{M}(\Sigma)$ το οποίο είναι κλειστό υποσύνολο του $\mathbb{R}^{|\Sigma|}$, επομένως από το κάτω όριο της σχέσης (3.1) για το ανοιχτό σύνολο $(\mathcal{M}(\Sigma))^c$ θα λάβουμε

$$-\infty = \liminf_{n\to\infty} \frac{1}{n} \log P_\sigma^\pi(\{L_n \in (\mathcal{M}(\Sigma))^c\}) \geq - \inf_{q\notin\mathcal{M}(\Sigma)} I(q),$$

συνεπώς $I(q) = \infty$ για κάθε $q \notin \mathcal{M}(\Sigma)$.

Τώρα σταθεροποιούμε ένα διάνυσμα πιθανότητας $q \in \mathcal{M}(\Sigma)$, ένα διάνυσμα $u >> 0$ και για $j = 1, ..., |\Sigma|$ θέτουμε $\theta_j = \log\left[\frac{u_j}{(u\mathbf{\Pi})_j}\right]$. Εφόσον $u >> 0$ και $\mathbf{\Pi}$ είναι ανάγωγος έπεται ότι και $(u\mathbf{\Pi})_j >> 0$. Παρατηρούμε ότι $u\mathbf{\Pi}_\theta = u$, επομένως $u\mathbf{\Pi}_\theta^n = u$ και άρα από το (ε) του θεωρήματος Perron-Frobenius με ($\phi_i = u_i > 0$) έπεται ότι $\rho(\mathbf{\Pi}_\theta) = 1$. Επομένως, εξ' ορισμού θα έχουμε,

$$I(q) \geq \sum_{j=1}^{|\Sigma|} q_j \log\left[\frac{u_j}{(u\mathbf{\Pi})_j}\right].$$

Όμως το $u >> 0$ είναι αυθαίρετο άρα η παραπάνω ανισότητα συνεπάγεται ότι $I(q) \geq J(q)$.

Για την αντίστροφη φορά σταθεροποιούμε ένα αυθαίρετο διάνυσμα $\theta \in \mathbb{R}^{|\Sigma|}$ και $u^* >> 0$ να είναι το αριστερό ιδιοδιάνυσμα που αντιστοιχεί στην ιδιοτιμή $\rho(\mathbf{\Pi}_\theta)$ του ανάγωγου πίνακα $\mathbf{\Pi}_\theta$, τότε $u^*\mathbf{\Pi}_\theta = \rho(\mathbf{\Pi}_\theta)u^*$ και εξ' ορισμού του $\mathbf{\Pi}_\theta$ θα έχουμε

$$\langle\theta, q\rangle + \sum_{j=1}^{|\Sigma|} q_j \log\left[\frac{(u^*\mathbf{\Pi})_j}{u_j^*}\right] = \sum_{j=1}^{|\Sigma|} q_j \log\left[\frac{(u^*\mathbf{\Pi}_\theta)_j}{u_j^*}\right] =$$

$$= \sum_{j=1}^{|\Sigma|} q_j \log \rho(\mathbf{\Pi}_\theta) = \log \rho(\mathbf{\Pi}_\theta).$$

Επομένως,

$$\langle\theta, q\rangle - \log\rho(\mathbf{\Pi}_\theta) \leq \sup_{u>>0}\sum_{j=1}^{|\Sigma|} q_j \log\left[\frac{u_j}{(u\mathbf{\Pi})_j}\right] = J(q).$$

Εφόσον το θ είναι αυθαίρετο έπεται ότι $I(q) \leq J(q)$ το οποίο ολοκληρώνει την απόδειξη. $\square$

3.3 Το Θεώρημα Sanov για το Εμπειρικό Μέτρο Ζεύγους Μαρκοβιανών Αλυσίδων

Η συνάρτηση ταχύτητας η οποία ΄*κυβερνά*΄ την LDP για το εμπειρικό μέτρο μιας μαρκοβιανής αλυσίδας βρίσκεται και αυτή ακολουθώντας την ίδια συλλογιστική που έχουμε ακολουθήσει καθ΄ όλη την πορεία αυτής της εργασίας, δηλαδή μέσω ενός προβλήματος βελτιστοποίησης. Το κόστός που πληρώνουμε όμως στην περίπτωση της μαρκοβιανής εξάρτησης είναι ότι έχουμε χάσει την κομψή ερμηνεία της συνάρτησης ταχύτητας ως σχετικής εντροπίας όπως είχαμε στην περίπτωση των ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Έχει ενδιαφέρον να δούμε ότι, εάν χρησιμοποιήσουμε μια ελαφρώς διαφορετική τυχαία μεταβλητή για να εξάγουμε την LDP για την $\{L_n\}$ μπορούμε να πάρουμε μια συνάρτηση ταχύτητας η οποία να έχει τη μορφή σχετικής εντροπίας. Αυτό θα είναι το αντικείμενο αυτής της ενότητας.

Έστω το σύνολο $\Sigma^2 = \Sigma \times \Sigma$ το οποίο αποτελείται από διαδοχικές δυάδες τιμών της ακολουθίας $Y_1, ..., Y_n$ δηλαδή, $Y_0Y_1, ...Y_iY_{i+1}, ..., Y_{n-1}Y_n$, όπου $Y_0 = \sigma$. Ακολουθώντας αυτή την προσέγγιση βλέπουμε ότι μπορούμε να πάρουμε μια μαρκοβιανή αλυσίδα με χώρο καταστάσεων Σ^2 και πίνακα μετάβασης $\mathbf{\Pi}^2$ (αντί για Σ, $\mathbf{\Pi}$ που είχαμε στην προηγούμενη διατύπωση), όπου ο πίνακας $\mathbf{\Pi}^2$ ορίζεται από τη σχέση

$$\pi^2(k \times l, i \times j) = \mathbb{1}_l(i)\pi(i, j).$$

Χάριν ευκολίας υποθέτουμε ότι ο $\mathbf{\Pi}$ είναι αυστηρά θετικός, δηλαδή $\forall i, j \quad \pi(i, j) > 0$. Τότε ο $\mathbf{\Pi}^2$ είναι ανάγωγος και επομένως μπορούμε να εφαρμόσουμε τα αποτελέσματα της προηγούμενης παραγράφου για να εξάγουμε μια συνάρτηση ταχύτητας $I_2(q)$ για το εμπειρικό μέτρο ζεύγους

$$L_{n,2} \coloneqq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_y(Y_{i-1}Y_i), \quad y \in \Sigma^2$$

Όπως και πριν θα έχουμε $L_{n,2} \in \mathcal{M}(\Sigma^2)$, επιπλέον για κάθε μέτρο $q \in \mathcal{M}(\Sigma^2)$, έστω οι περιθώριες κατανομές του

$$q_1(i) \coloneqq \sum_{j=1}^{|\Sigma|} q(i, j) \quad \& \quad q_2(i) \coloneqq \sum_{j=1}^{|\Sigma|} q(j, i).$$

Επιπλέον για $q_1(i) > 0$, ορίζουμε

$$q_f(j|i) := \frac{q(i,j)}{q_1(i)}.$$

Θα λέμε ότι ένα μέτρο πιθανότητας $q \in \mathcal{M}(\Sigma^2)$ είναι *αναλλοίωτο κάτω από μετατόπιση* (shift invariant) εάν

$$q_1 = q_2,$$

δηλαδή οι περιθώριες ταυτίζονται.

Θεώρημα 10. *Έστω $\mathbf{\Pi}$ αυστηρά θετικός πίνακας, τότε για κάθε μέτρο πιθανότητας $q \in \mathcal{M}(\Sigma^2)$,*

$$I_2(q) = \begin{cases} \sum_i q_1(i) H(q_f(\cdot|i)|\pi(i,\cdot)), & \text{εάν } q \text{ αναλλοίωτο κάτω από μετατοπίσεις} \\ \infty, & \text{αλλιώς} \end{cases},$$

όπου $H(\cdot|\cdot)$ είναι η συνάρτηση σχετικής εντροπίας που ορίσαμε στην ενότητα (2.5) δηλαδή,

$$H(q_f(\cdot|i)|\pi(i,\cdot)) = \sum_{j=1}^{|\Sigma|} q_f(j|i) \log \frac{q_f(j|i)}{\pi(i,j)}.$$

Σχόλιο 32. *Εάν ο $\mathbf{\Pi}$ δεν είναι αυστηρά θετικός αλλά είναι αναλλοίωτος το θεωρήμα και πάλι ισχύει εάν αντικαταστήσουμε το Σ^2 με το σύνολο $\{(i,j) : \pi(i,j) > 0\}$.*

Απόδειξη. (Θεωρήματος 10) Από το θεώρημα 9 θα έχουμε

$$I_2(q) = \sup_{u>>0} \sum_{j=1}^{|\Sigma|} \sum_{i=1}^{|\Sigma|} q(i,j) \log \frac{u(i,j)}{(u\mathbf{\Pi}^2)(i,j)} =$$

$$= \sup_{u>>0} \sum_{j=1}^{|\Sigma|} \sum_{i=1}^{|\Sigma|} q(i,j) \log \frac{u(i,j)}{[\sum_k u(k,j)]\pi(i,j)}, \tag{3.2}$$

όπου η τελευταία ισότητα έχει προκύψει από τον ορισμό του $\mathbf{\Pi^2}$.

Έστω τώρα ότι το q δεν είναι αναλλοίωτο κάτω από τις μετατοπίσεις, τότε χωρίς βλάβη της γενικότητας θα υπάρχει j_0 τέτοιο ώστε $q_1(j_0) < q_2(j_0)$. Οπότε για $u(\cdot, j) = 1, \quad j \neq j_0$ και $u(\cdot, j_0) = e^\alpha$ θα έχουμε,

$$\sum_{j=1}^{|\Sigma|}\sum_{i=1}^{|\Sigma|} q(i,j)\log\frac{u(i,j)}{[\sum_k u(k,j)]\pi(i,j)} = \sum_{j=1}^{|\Sigma|}\sum_{i=1}^{|\Sigma|} q(i,j)\log\frac{u(i,j)}{[|\Sigma|u(1,j)]\pi(i,j)} =$$

$$= -\sum_{j=1}^{|\Sigma|}\sum_{i=1}^{|\Sigma|} q(i,j)\log[|\Sigma|\pi(i,j)] + \alpha[q_2(j_0) - q_1(j_0)].$$

Επομένως παίρνοντας $\alpha \to \infty$ στην παραπάνω σχέση έχουμε $I_2(q) = \infty$, που ήταν και το ζητούμενο.

Εξετάζουμε τώρα τον άλλο κλάδο της συνάρτησης ταχύτητας, έστω δηλαδή q αναλλοίωτο κάτω από τις μετατοπίσεις, τότε για κάθε $u >> 0$,

$$\sum_{j=1}^{|\Sigma|}\sum_{i=1}^{|\Sigma|} q(i,j)\log\frac{\sum_k u(k,i)q_2(j)}{\sum_k u(k,j)q_1(i)} = 0. \tag{3.3}$$

Έστω $u(i|j) = \frac{u(i,j)}{\sum_k u(k,j)}$ και $q_b(i|j) = \frac{q(i,j)}{q_2(j)} = \frac{q(i,j)}{q_1(j)}$, τότε από τις σχέσεις (3.2),(3.3) έπεται ότι

$$I_2(q) - \sum_{i=1}^{|\Sigma|} q_1(i)H(q_f(\cdot|i)|\pi(i,\cdot)) = \sup_{u>>0}\sum_{j=1}^{|\Sigma|}\sum_{i=1}^{|\Sigma|} q(i,j)\log\frac{u(i,j)}{[\sum_k u(k,j)]q(i,j)} =$$

$$= \sup_{u>>0}\sum_{j=1}^{|\Sigma|}\sum_{i=1}^{|\Sigma|} q(i,j)\log\frac{u(i|j)}{q_b(i|j)} = \sup_{u>>0}(-\sum_{j=1}^{|\Sigma|} q_2(j)H(q_b(\cdot|j)|u(\cdot|j))).$$

Παρατηρούμε τώρα ότι $I_2(q) \leq \sum_i q_1(i)H(q_f(\cdot|j)|\pi(\cdot|j)))$ διότι η H είναι μη αρνητική, ενώ εάν $q >> 0$ η επιλογή $u = q$ μας δίνει τη ζητούμενη ισότητα. Για να ολοκληρώσουμε την απόδειξη τώρα, έστω q το οποίο δεν ειναι αυστηρά θετικό θεωρούμε μια ακολουθία $u_n >> 0$ τέτοια ώστε $u_n \to q$, τότε για κάθε j,

$$q_2(j)H(q_b(\cdot|j)|u_n(\cdot|j))) \to 0$$

και έχουμε τελειώσει. □

3.4 Σπάνια Τμήματα σε Τυχαίους Περιπάτους

Η ενότητα αυτή είναι αφιερωμένη σε μια εφαρμογή των αποτελεσμάτων που έχουμε ήδη αποδείξει, σε τυχαίους περιπάτους. Έστω ο τυχαίος περίπατος με $S_0 = 0$ και $S_k = \sum_{i=1}^{k} X_i, \quad k = 1, 2, ..$ όπου οι X_i είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές οι οποίες παίρνουν τιμές στο $\mathbb{R}^d$.

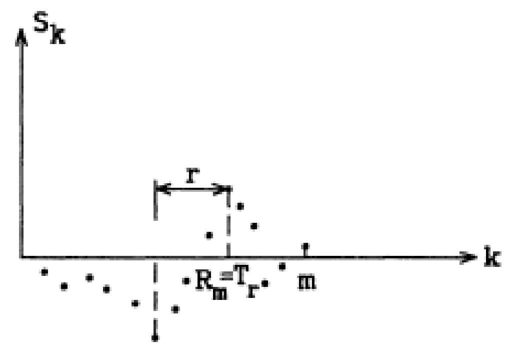
Έστω R_m να είναι το τμήμα του τυχαίου περιπάτου με το μέγιστο μήκος μέχρι τη στιγμή m, για το οποίο το εμπειρικό του μέτρο ανήκει σε ένα μετρήσιμο σύνολο $A \subset \mathbb{R}^d$, δηλαδή

$$R_m := \max\{l - k : 0 \leq k < l \leq m, \frac{S_l - S_k}{l - k} \in A\}.$$

Συσχετισμένοι με το R_m είναι οι χρόνοι διακοπής

$$T_r := \inf\{l : \frac{S_l - S_k}{l - k} \in A \quad \text{για κάποιο} 0 \leq k \leq l - r\},$$

ώστε $\{R_m \geq r\}$ αν και μόνο αν $\{T_r \leq m\}$.



Οι τυχαίες μεταβλητές R_m και T_r εμφανίζονται συχνά στη συγκριτική ανάλυση ακολουθιών γενετικού υλικού καθώς και στην ανάλυση του χρόνου εκτέλεσης αλγορίθμων. Το επόμενο θεώρημα μας παρέχει μια ασυμπτωτική εκτίμηση για τις R_m και T_r στην περίπτωση που $m \to \infty$ και $r \to \infty$ αντίστοιχα.

Θεώρημα 11. *Έστω σύνολο A τέτοιο ώστε το όριο*

$$I_A := -\lim_{n\to\infty}\frac{1}{n}\log\mu_n(A), \tag{3.4}$$

υπάρχει, όπου μ_n είναι η κατανομή του εμπειρικού μέσου $\bar{S}_n = \frac{1}{n}S_n$, τότε, σχεδόν βεβαίως θα έχουμε

$$\lim_{m\to\infty}\frac{R_m}{\log m} = \lim_{r\to\infty}\frac{r}{\log T_r} = \frac{1}{I_A}.$$

Απόδειξη. Έστω

$$C_{k,l} := \left\{\frac{S_l - S_k}{l-k} \in A\right\}.$$

Τότε για την αριστερή ουρά της κατανομής της T_r θα έχουμε

$$\{T_r \leq m\} \subset \bigcup_{k=0}^{m-r}\bigcup_{l=k+r}^{m} C_{k,l} \subset \bigcup_{k=0}^{m-1}\bigcup_{l=k+r}^{\infty} C_{k,l}.$$

Παρατηρούμε τώρα ότι υπάρχουν m δυνατές επιλογές για το k, ενώ $\mathbb{P}(C_{k,l}) = \mu_{l-k}(A), \quad l-k \geq r$, επομένως από την ανισότητα Boole θα έχουμε,

$$\mathbb{P}(T_r \leq m) \leq m\sum_{n=r}^{\infty}\mu_n(A).$$

Υποθέτουμε τώρα ότι $0 < I_A < \infty$, τότε θέτοντας $m = \left\lfloor e^{r(I_A-\epsilon)}\right\rfloor$ και χρησιμοποιώντας τη σχέση (3.4) έχουμε,

$$\sum_{r=1}^{\infty}\mathbb{P}(T_r \leq e^{r(I_A-\epsilon)}) \leq \sum_{r=1}^{\infty}e^{r(I_A-\epsilon)}\sum_{n=r}^{\infty}ce^{-n(I_A-\frac{\epsilon}{2})} \leq c'\sum_{r=1}^{\infty}e^{\frac{-r\epsilon^2}{2}} < \infty,$$

για κάθε $\epsilon > 0$ και για κάποιες θετικές σταθερές c, c', οι οποίες μπορεί να εξαρτώνται από το ϵ. Εάν τώρα $I_A = \infty$, χρησιμοποιώντας το ίδιο επιχείρημα για $m = \left\lfloor e^{\frac{r}{\epsilon}}\right\rfloor$ θα λάβουμε,

$$\sum_{r=1}^{\infty}\mathbb{P}(T_r \leq e^{r(I_A-\epsilon)}) < \infty, \quad \forall\epsilon > 0,$$

Από το λήμμα Borel-Cantelli έχουμε,

$$\liminf_{r\to\infty}\frac{1}{r}\log T_r \geq I_A, \quad \text{σχεδόν βεβαίως.}$$

Χρησιμοποιώντας την ισότητα $\{R_m \geq r\} = \{T_r \leq m\}$, από την παραπάνω σχέση έπεται ότι,

$$\limsup_{m\to\infty} \frac{R_m}{\log m} \leq \frac{1}{I_A}, \quad \text{σχεδόν βεβαίως.}$$

Τώρα εάν $I_A = \infty$ δεν έχουμε κάτι να αποδείξουμε. Οπότε εξετάζουμε την αντίθετη φορά για την περίπτωση που $I_A < \infty$. Σε αυτή την περίπτωση για να πετύχουμε το ζητούμενο θα πρέπει η δεξιά ουρά της κατανομής της T_r να είναι φραγμένη.

Έστω

$$B_l := \{\frac{1}{r}(S_{lr} - S_{(l-1)r}) \in A\}.$$

Τα B_l είναι ανεξάρτητα ενδεχόμενα με $\mathbb{P}(B_l) = \mu_r(A)$. Συνεπώς,

$$\bigcup_{l=1}^{\lfloor \frac{m}{r} \rfloor} B_l \subset \{T_r \leq m\} \Rightarrow \mathbb{P}(T_r > m) \leq 1 - \mathbb{P}(\bigcup_{l=1}^{\lfloor \frac{m}{r} \rfloor} B_l) = (1 - \mathbb{P}(B_1))^{\lfloor \frac{m}{r} \rfloor} \leq$$

$$\leq e^{-\lfloor \frac{m}{r} \rfloor \mathbb{P}(B_1)} = e^{-\lfloor \frac{m}{r} \rfloor \mu_r(A)}.$$

Συνδυάζοντας τώρα αυτήν την ανισότητα με τη σχέση (3.4) για $m = \lfloor e^{r(I_A+\epsilon)} \rfloor$ έπεται ότι για κάθε $\epsilon > 0$,

$$\sum_{r=1}^{\infty} \mathbb{P}(T_r > e^{r(I_A+\epsilon)}) \leq \sum_{r=1}^{\infty} exp(-\frac{c}{r} e^{r(I_A+\epsilon)} e^{-r(I_A+\frac{\epsilon}{2})}) \leq \sum_{r=1}^{\infty} exp(-c_2 e^{c_3 r}) < \infty,$$

όπου c_1, c_2, c_3 είναι θετικές σταθερές οι οποίες μπορεί να εξαρτώνται από το ϵ. Από το λήμμα Borel-Cantelli και την ισότητα $\{R_m \geq r\} = \{T_r\} \leq m$,

$$\liminf_{m\to\infty} \frac{R_m}{\log m} = \liminf_{r\to\infty} \frac{r}{\log T_r} \leq \frac{1}{I_A}, \quad \text{σχεδόν βεβαίως,}$$

το οποίο ολοκληρώνει την απόδειξη. $\square$

Παράρτημα Α΄

Βασικές Ανισότητες

Α΄.1 Ανισότητα Markov

Θεώρημα. *Έστω* $(\Omega, \mathcal{F}, \mathbb{P})$ *χώρος πιθανότητας και* X *μη αρνητική τυχαία μεταβλητή και* $\alpha > 0$, *τότε*

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha}.$$

Απόδειξη. Έχουμε

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx \overset{X\geq 0}{=} \int_{0}^{+\infty} xf(x)dx = \int_{0}^{\alpha} xf(x)dx + \int_{\alpha}^{+\infty} xf(x)dx \geq$$
$$\geq \int_{\alpha}^{+\infty} xf(x)dx \geq \int_{\alpha}^{+\infty} \alpha f(x)dx = \alpha \int_{\alpha}^{+\infty} f(x)dx = \alpha\mathbb{P}(X \geq \alpha) \Rightarrow$$
$$\Rightarrow \mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha}.$$

□

Α΄.2 Ανισότητα Chebyshev

Θεώρημα. *Έστω* $(\Omega, \mathcal{F}, \mathbb{P})$ *χώρος πιθανότητας και η τυχαία μεταβλητή* X *με πεπερασμένη μέση τιμή* μ *και μη μηδενική διασπορά* σ^2, *τότε για κάθε* $k > 0$

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{1}{k^2}.$$

Απόδειξη. Η απόδειξη γίνεται με εφαρμογή της ανισότητας Markov για την τυχαία μεταβλητή $Y = (X - \mu)^2 \geq 0$ και $\alpha = (k\sigma)^2 > 0$. □

A′.3 Ανισότητα Jensen

Θεώρημα. *Έστω $(\Omega, \mathcal{F}, \mathbb{P})$ χώρος πιθανότητας, X τυχαία μεταβλητή και ϕ κυρτή πραγματική συνάρτηση, τότε*

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)).$$

Απόδειξη. Έστω

$$x_0 \coloneqq \mathbb{E}(X).$$

Επιπλέον εφόσον η ϕ είναι κυρτή μπορούμε να επιλέξουμε α, β τέτοια ώστε,

$$\alpha x + \beta \leq \phi(x), \quad \forall x \in \mathbb{R} \quad \& \quad \alpha x_0 + \beta = \phi(x_0).$$

□

Οπότε τώρα θα έχουμε,

$$\phi(X) \geq \alpha X + \beta \Rightarrow \mathbb{E}(\phi(X)) \geq \mathbb{E}(\alpha X + \beta) = \alpha x_0 + \beta = \phi(x_0) = \phi(\mathbb{E}(X)).$$

A′.4 Ανισότητα Boole

Θεώρημα. *Έστω $(\Omega, \mathcal{F}, \mathbb{P})$ χώρος πιθανότητας και $\{A_n\} \in \mathcal{F}$, τότε*

$$\mathbb{P}(\bigcup_{n=1}^{+\infty} A_n) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Απόδειξη. Κατασκευάζουμε την εξής ακολουθία συνόλων

$$B_n \coloneqq A_n - \bigcup_{i=1}^{n-1} A_n.$$

Παρατηρούμε τώρα ότι για κάθε $n \in \mathbb{N}$, $B_n \in \mathcal{F}$ και $B_n \subset A_n$, άρα

$$\mathbb{P}(B_n) \leq \mathbb{P}(A_n).$$

Επιπλέον τα B_n είναι μεταξύ τους ξένα και δεν είναι δύσκολο να δούμε ότι

$$\bigcup_{n=1}^{+\infty} A_n = \bigcup_{n=1}^{+\infty} B_n.$$

Συνεπώς συνδυάζοντας όλα τα παραπάνω θα έχουμε,

$$\mathbb{P}(\bigcup_{n=1}^{+\infty} A_n) = \mathbb{P}(\bigcup_{n=1}^{+\infty} B_n) = \sum_{n=1}^{+\infty} \mathbb{P}(B_n) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

□

A′.5 Ανισότητα Hölder

Θεώρημα. *Έστω* $(S, \mathcal{F}, \mu)$ *χώρος πιθανότητας και* $p, q \in [1, +\infty]$ *με* $\frac{1}{p} + \frac{1}{q} = 1$*, τότε για κάθε μετρήσιμες συναρτήσεις* f, g *στο* S *έχουμε*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \,,$$

όπου τα p, q *καλούνται συζυγείς εκθέτες.*

Απόδειξη. Εφόσον έχουμε ότι $p \geq 1$ η συνάρτηση $\phi(x) = x^p$ θα είναι κυρτή, επομένως από την ανισότητα Jensen για κάθε μέτρο πιθανότητας ν και κάθε $\nu-$μετρήσιμη συνάρτηση h θα ισχύει

$$\int h d\nu \leq \left(\int h^p d\nu \right)^{\frac{1}{p}}.$$

Συνεπώς, παίρνουμε το ν να είναι τέτοιο ώστε:

$$d\nu = \frac{g^p}{\int g^p d\mu} d\mu \quad \& \quad h = fg^{1-q}.$$

Τότε

$$\|fg\|_1 = \int fg d\mu = \int g^q d\mu \frac{1}{\int g^q d\mu} \int fg^q g^{1-q} d\mu = \int g^q d\mu \int fg^{1-q} \frac{g^q}{\int g^q d\mu} d\mu \leq$$

$$\leq \int g^q d\mu \left(\int f^p g^{p(1-q)} \frac{g^q}{\int g^q d\mu} d\mu \right)^{\frac{1}{p}} = \int g^q d\mu \left(\frac{f^p}{\int g^q d\mu} \right)^{\frac{1}{p}} = \|f\|_p \|g\|_q \,.$$

□

Παράρτημα Β΄

Βασικά Θεωρήματα

Β΄.1 Ο Ασθενής Νόμος των Μεγάλων Αριθμών

Θεώρημα. *Έστω $(\Omega, \mathcal{F}, \mathbb{P})$ χώρος πιθανότητας και $X_1, ..., X_n, ...$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με μέση τιμή μ, τότε*

$$\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = 0, \quad \forall \epsilon > 0.$$

Απόδειξη. Δεν είναι απαραίτητο για την ισχύ του θεωρήματος αλλά χάριν ευκολίας της απόδειξης θα θεωρήσουμε ότι $\sigma^2 < \infty$, τότε

$$Var(\bar{X}_n) = \frac{\sigma^2}{n} \quad \& \quad \mathbb{E}(\bar{X}_n) = \mu.$$

Εφαρμόζοντας την ανισότητα Chebyshev για το μέσο $\bar{X}_n$ έχουμε,

$$\forall \epsilon > 0 \quad \mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \Rightarrow \mathbb{P}(|\bar{X}_n - \mu| < \epsilon) = 1 - \mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq 1 - \frac{\sigma^2}{n\epsilon^2} \to 1, \quad n \to \infty,$$

άρα

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mu, \quad n \to \infty,$$

που είναι ισοδύναμο με το ζητούμενο.

□

B′.2 Το Κεντρικό Οριακό Θεώρημα

Θεώρημα. *Έστω $(\Omega, \mathcal{F}, \mathbb{P})$ χώρος πιθανότητας και $X_1, ..., X_n$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με μέση τιμή μ και πεπερασμένη διασπορά σ^2, τότε*

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2), \quad n \to \infty.$$

Απόδειξη. Εάν η $X_i, i = 1, ..., n$ έχει μέση τιμή μ και διασπορά σ^2 τότε η τυχαία μεταβλητή $S_n = \sum_{i=1}^n X_n$ θα έχει μέση τιμή $n\mu$ και διασπορά $n\sigma^2$. Επιπλέον θεωρούμε την τυχαία μεταβλητή

$$Z_n \coloneqq \frac{S_n}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i,$$

όπου χάριν ευκολίας έχουμε πάρει $Y_i = \frac{X_i - \mu}{\sigma}$. Η χαρακτηριστική συνάρτηση της Z_n θα είναι

$$\phi_{Z_n}(t) = \phi_{Y_1}(\frac{t}{\sqrt{n}})\phi_{Y_2}(\frac{t}{\sqrt{n}})...\phi_{Y_n}(\frac{t}{\sqrt{n}}) = [\phi_{Y_1}(\frac{t}{\sqrt{n}})]^n.$$

Παίρνοντας τώρα την επέκταση κατά Taylor θα έχουμε

$$\phi_{Y_1}(\frac{t}{\sqrt{n}}) = 1 - \frac{t^2}{2n} + o(\frac{t^2}{2n}), \quad \frac{t}{\sqrt{n}} \to 0,$$

άρα

$$\phi_{Z_n}(\frac{t}{\sqrt{n}}) = \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{2n})\right)^n \to e^{-\frac{1}{2}t^2}, \quad n \to \infty,$$

άρα $Z_n \to N(0, 1)$ (θεώρημα συνέχειας Lévy) και το ζητούμενο προκύπτει άμεσα. □

B′.3 Το Θεώρημα Κυριαρχημένης Σύγκλισης

Θεώρημα. *Έστω $(\Omega, \mathcal{F}, \mu)$ χώρος μέτρου και έστω $\{f_n\}$ μια ακολουθία μετρήσιμων συναρτήσεων η οποία συγκλίνει κατά σημείο σε μία συνάρτηση f και η ακολουθία κυριαρχείται από μία μετρήσιμη συνάρτηση g, δηλαδή $|f_n(x)| \leq g(x)$ για κάθε $n \in \mathbb{N}$ και κάθε $x \in \Omega$, τότε η f είναι ολοκληρώσιμη και*

$$\lim_{n \to \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu$$

Απόδειξη. Η απόδειξη κάνει χρήση του Λήμματος Fatou και μπορεί να βρεθεί στο [6]. □

Βιβλιογραφία

[1] Amir Dembo & Ofer Zeitouni (1998), *Large Deviations Techniques and Applications*, Springer.

[2] Frank den Hollander (2000), *Large Deviations*, American Mathematical Society

[3] Parthanil Roy (2020), *Notes and Lectures on the Theory of Large Deviations*, Indian Statistical Institute

[4] Μιχάλης Λουλάκης (2005), *Θεωρία Μεγάλων Αποκλίσεων*, Θερινό Σχολείο Μαθηματικών Ηράκλειο

[5] Δημήτρης Χελιώτης (2015), *Ένα Δεύτερο Μάθημα στις Πιθανότητες*, Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών.

[6] Rick Durrett (2017), *Probability: Theory and Examples*, Cambridge University Press

[7] Hugo Touchette (2009), *The large deviation approach to statistical mechanics*, arXiv:0804.0327v2

[8] P. Lancaster (1969), *Theory of Matrices, Academic Press, New York*