ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΔΙΟΝΥΣΙΟΣ ΜΗΛΕΣΗΣ

«Η Αρχή του Gibbs και Εφαρμογές της»

ΣΤΗΝ ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ:

Θεωρία Πιθανοτήτων

ΕΠΙΒΛΕΠΩΝ:

Λουλάκης Μιχαήλ
Καθηγητής, Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

Λουλάκης Μιχαήλ
Καθηγητής, Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Παπανικολάου Βασίλειος
Ομότιμος Καθηγητής, Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Φουσκάκης Δημήτριος
Καθηγητής, Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Αθήνα, Ιούνιος 2023

Ευχαριστίες

Στο παρόν πόνημα, την κορωνίδα του πρώτου μου ακαδημαϊκού κύκλου, συνέβαλαν, ίσως και δίχως επίγνωση, άνθρωποι παλαιοί και πρόσφατοι, άνθρωποι προσωρινοί κι αιώνιοι, που στάθηκαν και στέκονται στο πλευρό μου. Θα ήθελα να ευχαριστήσω εγκάρδια αυτούς τους ανθρώπους, μα ειδικότερα τους καθηγητές μου, την οικογένειά μου και τους φίλους μου. Οι πρώτοι με στήριξαν, όταν δεν καταλάβαινα. Η δεύτερη με στήριξε, όταν δεν καταλάβαινε. Οι τρίτοι με στήριξαν και στις δύο περιπτώσεις.

Ιδιαίτερα θα ήθελα να ευχαριστήσω τον καθηγητή κι επιβλέποντά μου κύριο Λουλάκη Μιχαήλ, που με παρότρυνε να βουτήξω στον γοητευτικό και βαθυστόχαστο κόσμο της θεωρίας των μεγάλων αποκλίσεων. Η καθοδήγησή του, καθώς και η εμπιστοσύνη που μου επέδειξε, μου δίδαξαν πολλά. Τόσο για τα μαθηματικά, όσο και για τον εαυτό μου.

....

Διονύσιος Μήλεσης

Περίληψη

Στην παρούσα εργασία θα αποδείξουμε την αρχή του Gibbs για κατανομές με πεπερασμένες εκθετικές ροπές, καθώς και το ανάλογο της αρχής του Gibbs για κατανομές των οποίων οι εκθετικές ροπές δεν ορίζονται (κατανομές με βαριές ουρές). Τέλος, θα παρουσιάσουμε μια εφαρμογή της Αρχής του Gibbs για τη μελέτη μεγάλων απωλειών χαρτοφυλακίου.

Λέξεις Κλειδιά. Μεγάλη Απόκλιση, Σχετική Εντροπία, Εμπειρική Κατανομή, Αρχή του Gibbs, Απώλειες Χαρτοφυλακίου

Abstract

In the present thesis, we will prove the Gibbs Conditioning Principle for distributions with finite exponential moments, as well as the corresponding principle for distributions whose exponential moments are not defined (heavy-tailed distributions). Finally, we will present an application of the Gibbs Conditioning Principle for studying large portfolio losses.

Keywords. Large Deviation, Relative Entropy, Empirical Distribution, Gibbs Conditioning Principle, Portfolio Losses

Πρόλογος

Οι Dembo και Zeitouni σ' ένα περίφημο άρθρο τους [14] αποκάλεσαν την Αρχή του Gibbs ένα μέτα-θεώρημα της θεωρίας των μεγάλων αποκλίσεων. Η διαφορά μεταξύ θεωρήματος και μέτα-θεωρήματος είναι η εξής: ένα θεώρημα αποδεικνύεται *από* μια θεωρία, ενώ ένα μέτα-θεώρημα αποδεικνύεται *για* τη θεωρία. Μολονότι αυτός ο χαρακτηρισμός της Αρχής του Gibbs μπορεί να φαντάζει υπερβολικός εκ πρώτης, ωστόσο εμπεριέχει αρκετή ουσία, καθώς η αρχή αυτή δεν αποτελεί μονάχα απόρροια της θεωρίας των μεγάλων αποκλίσεων, αλλά συνιστά ένα γενικότερο αποτέλεσμα, που λέει με απλά λόγια ότι "υπό έναν περιορισμό στην ενέργεια ενός συστήματος, η φύση μεγιστοποιεί την εντροπία" - με απλοϊκό τρόπο, αυτό είναι σα να λέμε πως τα μικρά παιδιά μεγιστοποιούν την αταξία τους, μέχρι να εξαντληθούν. Αυτήν ακριβώς τη διαίσθηση παρέχουν για την αρχή του Gibbs οι Seppalainen και Rassoul-Agha στο διάσημο βιβλίο τους [20]. Δεν είναι παράλογο, συνεπώς, που η αρχή αυτή έχει βρει τόσες εφαρμογές στη στατιστική μηχανική, όπου πολλές φορές επιβάλλεται ένας ενεργειακός περιορισμός στο σύστημα και ζητείται ο τρόπος με τον οποίον αλλάζει το σύστημα υπό τον περιορισμό. Αυτό λοιπόν το "μέτα-θεώρημα" θα επιδιώξουμε εμείς να αποδείξουμε, μολονότι θα του φερθούμε τελικά σα "θεώρημα", κι αυτό διότι η θεωρία μεγάλων αποκλίσεων καταφέρνει επιτυχώς να γενικεύσει ένα τέτοιο αποτέλεσμα σε χώρους πιο αφηρημένους με τρόπο πολύ κομψό.

Όσον αφορά τη διάρθρωση της εργασίας, στο πρώτο κεφάλαιο θα δούμε κάποιους βασικούς ορισμούς της θεωρίας των μεγάλων αποκλίσεων, με τελικό σκοπό την απόδειξη του θεωρήματος του Cramér, που εκ πρώτης δε φαίνεται να έχει καμία σχέση με την αρχή του Gibbs, κι ας είναι στην πραγματικότητα άρρηκτα συνδεδεμένη με αυτήν. Στη συνέχεια, θα ανεβούμε ένα σκαλοπάτι αφαίρεσης, μιλώντας για μεγάλες αποκλίσεις σε γενικότερους τοπολογικούς χώρους. Εκεί θα συναντήσουμε το λήμμα του Varadhan, που θα μας βοηθήσει να αποδείξουμε το θεώρημα του Sanov, συνδυάζοντας αρχές της τοπολογίας, της συναρτησιακής ανάλυσης, της κυρτής ανάλυσης και της θεωρίας πιθανοτήτων. Στη διαδρομή αυτή θα δούμε δύο έννοιες υψίστης σημασίας στις μεγάλες αποκλίσεις: την εμπειρική κατανομή τυχαίων μεταβλητών και τη σχετική εντροπία δύο κατανομών. Κι αφού θα έχουμε εξετάσει όλα αυτά, στο τρίτο κεφάλαιο θα αποδείξουμε την αρχή της μέγιστης εντροπίας και, τελικά, την αρχή του Gibbs, όπου θα εξετάσουμε δύο περιπτώσεις: την αρχή του Gibbs για κατανομές με πεπερασμένες εκθετικές ροπές και το ανάλογο της αρχής αυτής για κατανομές των οποίων οι εκθετικές ροπές δεν ορίζονται. Τέλος, στο τέταρτο κεφάλαιο θα παρουσιάσουμε μια εφαρμογή της αρχής του Gibbs στη μελέτη των μεγάλων απωλειών χαρτοφυλακίου, επιδιώκοντας έτσι να δείξουμε πως η αρχή αυτή δεν είναι μονάχα κομψή, αλλά, πρωτίστως, και χρήσιμη.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Τι αναζητάει η θεωρία των μεγάλων αποκλίσεων;

Η θεωρία των μεγάλων αποκλίσεων έχει σημειώσει τα τελευταία χρόνια μεγάλη εξέλιξη στον κόσμο των μαθηματικών και της επιστήμης. Η εξέλιξη αυτή είναι διττή. Αφενός, έχει αποτελέσει το εφαλτήριο για την εξαγωγή πολύ ισχυρών θεωρητικών συμπερασμάτων, που έχουν συμβάλλει στην βαθύτερη κατανόηση της θεωρίας πιθανοτήτων και του στοχαστικού λογισμού. Παραδείγματος χάριν, η οπτική των μεγάλων αποκλίσεων δίνει πολύ ισχυρά αποτελέσματα, μεταξύ άλλων, στην εργοδική θεωρία των μαρκοβιανών διαδικασιών (βλέπε για παράδειγμα τα [19] και [15]), αλλά και στη μελέτη των στοχαστικών δυναμικών συστημάτων με τη θεωρία τυχαίων διαταραχών των Freidlin-Wentzel [3]. Αφετέρου, χάρη στη θεωρία μεγάλων αποκλίσεων έχει αναδυθεί ένα ευρύ φάσμα εφαρμογών, που εκτείνεται από τη στατιστική φυσική [20], τη μελέτη της δυναμικής των γεωφυσικών ρευστών και του κλίματος [11], έως και την μαθηματική χρηματοοικονομία [6]. Ας δούμε όμως με τι ασχολείται αυτή η τόσο αισιόδοξη θεωρία.

Η θεωρία των μεγάλων αποκλίσεων μελετάει την εκτίμηση της πιθανότητας σπάνιων ενδεχομένων. Ως σπάνια ενδεχόμενα εννοούμε ενδεχόμενα που αποκλίνουν από τυπικές συμπεριφορές στον χώρο πιθανότητας που έχουμε θεωρήσει. Μια τέτοια συμπεριφορά θα μπορούσε, παραδείγματος χάριν, να αποτελεί αυτή που υποδεικνύει ο νόμος των μεγάλων αριθμών. Πράγματι, ας θεωρήσουμε μια ακολουθία $\{X_n\}_{n\in\mathbb{N}}$ ανεξάρτητων, ισόνομων τυχαίων στον χώρο πιθανότητας που χαρακτηρίζεται από την τριπλέτα $(\Omega, \mathcal{F}, \mathbb{P})$. Έστω επιπλέον ότι έχουν πεπερασμένη μέση τιμή. Μάλιστα, μπορούμε χωρίς βλάβη της γενικότητας να υποθέσουμε ότι η μέση τους τιμή είναι 0. Τότε, ο ασθενής νόμος των μεγάλων αριθμών μας δίνει ότι:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \delta\right) = 0, \quad \forall \delta \geq 0 \tag{1.1}$$

όπου $S_n = \sum\limits_{k=1}^{n} X_k$.

Η παραπάνω σχέση μας δείχνει στην ουσία ότι το ενδεχόμενο ο μέσος όρος των τυχαίων μεταβλητών να αποκλίνει από τη μέση τους τιμή γίνεται σπάνιο, καθώς το n γίνεται μεγάλο. Πόσο γρήγορα όμως γίνεται η πραγματοποιείται σύγκλιση;

Το παραπάνω ερώτημα δεν είναι καθόλου τετριμμένο. Ευτυχώς, μπορούμε να το απαντήσουμε για την κανονική κατανομή δίχως να διαφύγουμε σε τεχνικές των μεγάλων αποκλίσεων. Έστω λοιπόν ότι οι τυχαίες μεταβλητές ακολουθούν την τυποποιημένη κανονική κατανομή $\mathcal{N}(0,1)$. Η κατανομή του αθροίσματος n τέτοιων τυχαίων μεταβλητών είναι κανονική με μέση τιμή $n \cdot 0 = 0$ και διασπορά $n \cdot 1 = n$, οπότε η κατανομή του μέσου όρου θα είναι κανονική με μέση τιμή 0 και διασπορά $\frac{n}{n^2} = \frac{1}{n}$. Έστω μ_n η κατανομή του μέσου όρου των $\{X_i\}_{1\le i\le n}$, δηλαδή:

$$\mu_n(A) = \mathbb{P}\left(\frac{S_n}{n} \in A\right)$$

για $A \in \mathcal{F}$.

Τότε παίρνουμε ότι:

$$\begin{aligned}
\mu_n((-\delta,\delta)^c) &= \mathbb{P}\left(\left|\frac{S_n}{n}\right| \ge \delta\right) \\
&= 2\int_\delta^\infty \sqrt{\frac{n}{2\pi}} e^{-\frac{nx^2}{2}} dx \\
&= \frac{2}{\sqrt{2\pi}} \int_{\delta\sqrt{n}}^\infty e^{-\frac{x^2}{2}} dx
\end{aligned}$$

Είναι μια κλασική και χρήσιμη άσκηση στις Πιθανότητες η απόδειξη της ακόλουθης σχέσης:

$$\frac{x}{x^2+1} e^{-\frac{x^2}{2}} \le \int_x^\infty e^{-\frac{u^2}{2}} du \le \frac{1}{x} e^{-\frac{x^2}{2}} \tag{1.2}$$

η οποία ισχύει για κάθε $x > 0$. Μια απόδειξη αυτής της σχέσης είναι η εξής:

Για το άνω φράγμα θα έχουμε:

$$\int_x^\infty e^{-\frac{u^2}{2}} du = \int_x^\infty 1 \cdot e^{-\frac{u^2}{2}} du \le \int_x^\infty \frac{u}{x} e^{-\frac{u^2}{2}} du = \frac{1}{x} e^{-\frac{x^2}{2}}$$

Ενώ για το κάτω φράγμα ορίζουμε τη συνάρτηση:

$$g(x) = \int_x^\infty e^{-\frac{u^2}{2}} du - \frac{x}{x^2+1} e^{-\frac{x^2}{2}}$$

κατόπιν, υπολογίζουμε την παράγωγό της:

$$g'(t) = -\frac{2e^{-\frac{x^2}{2}}}{(x^2+1)^2} < 0$$

και άρα, η g είναι γνησίως αύξουσα. Αλλά $g(0) > 0$ και $\lim_{x\to+\infty} g(x) = 0$, άρα η g πρέπει να είναι πάντα θετική, κι έτσι αποδεικνύουμε και το κάτω φράγμα.

Συνεχίζοντας, λοιπόν, θέτουμε $x = \delta\sqrt{n}$ στην 1.2 και παίρνουμε:

$$\begin{aligned}
\frac{\delta\sqrt{n}}{n\delta^2+1} e^{-\frac{n\delta^2}{2}} &\le \int_{\delta\sqrt{n}}^{\infty} e^{-\frac{x^2}{2}} dx \le \frac{1}{\delta\sqrt{n}} e^{-\frac{n\delta^2}{2}} \\
\frac{\delta\sqrt{n}}{n\delta^2+1} e^{-\frac{n\delta^2}{2}} &\le \mu_n((-\delta,\delta)^c) \le \frac{1}{\delta\sqrt{n}} e^{-\frac{n\delta^2}{2}} \\
\log\left(\frac{\delta\sqrt{n}}{n\delta^2+1}\right) - \frac{n\delta^2}{2} &\le \log\mu_n((-\delta,\delta)^c) \le -\log\delta\sqrt{n} - \frac{n\delta^2}{2} \\
\frac{1}{n}\log\left(\frac{\delta\sqrt{n}}{n\delta^2+1}\right) - \frac{\delta^2}{2} &\le \frac{1}{n}\log\mu_n((-\delta,\delta)^c) \le -\frac{1}{n}\log\delta\sqrt{n} - \frac{\delta^2}{2}
\end{aligned}$$

Όμως για τα όρια του αριστερού και του δεξιού μέλους ισχύει ότι:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\log\left(\frac{\delta\sqrt{n}}{n\delta^2+1}\right) &= 0 \\
\lim_{n\to\infty} \frac{1}{n}\log\delta\sqrt{n} &= 0
\end{aligned}$$

Και άρα περνώντας στο όριο έχουμε ότι:

$$\lim_{n\to\infty} \frac{1}{n}\log\mu_n((-\delta,\delta)^c) = \lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left(\left|\frac{S_n}{n}\right| \ge \delta\right) = -\frac{\delta^2}{2}$$

Συνεπώς, η ασυμπτωτική συμπεριφορά της πιθανότητας στην (1.1) για μεγάλο n θα είναι:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \ge \delta\right) \approx e^{-\frac{n\delta^2}{2}} \tag{1.3}$$

Όπως θα δούμε παρακάτω, μια τέτοια συμπεριφορά δεν είναι χαρακτηριστική μόνο της κανονικής κατανομής. Μάλιστα, ένα μεγάλο κομμάτι της θεωρίας μεγάλων αποκλίσεων ασχολείται με τον υπολογισμό και τις συνέπειες του ρυθμού αυτής της σύγκλισης.

Άλλο ένα ερώτημα με το οποίο ασχολείται η θεωρία μεγάλων αποκλίσεων είναι το ακόλουθο: Έστω ότι ένα ενδεχόμενο A ως προς μια κατανομή μ είναι άτυπο, δηλαδή έχει πολύ μικρή πιθανότητα. Υπάρχει κάποια κατανομή $\tilde{\mu}$ στον ίδιο χώρο πιθανότητας, κάτω από την οποία το A γίνεται τυπικό; Το ερώτημα αυτό είναι ιδιαίτερα σημαντικό σε τομείς, όπως η χρηματοοικονομία και η στατιστική μηχανική. Την απάντησή του θα μας τη δώσει η αρχή του Gibbs που αποτελεί έναν από τους κύριους στόχους της παρούσας εργασίας.

1.2 Γνωριμία με τις βασικότερες έννοιες των μεγάλων αποκλίσεων

Εάν ο $\mathcal{X}$ είναι ένας τοπολογικός χώρος, τότε μια συνάρτηση $f : \mathcal{X} \to [0, +\infty]$ ονομάζεται *κάτω ημισυνεχής*, εάν για κάθε $a \in \mathbb{R}$ τα σύνολα:

$$\Psi_a = \{x \in \mathcal{X} : f(x) \leq a\}$$

είναι κλειστά. Αντίστοιχα, ονομάζεται *άνω ημισυνεχής*, εάν για κάθε $a \in \mathbb{R}$ τα σύνολα:

$$\Psi_a = \{x \in \mathcal{X} : f(x) \geq a\}$$

είναι κλειστά. Προφανώς, η f είναι συνεχής αν και μόνο αν είναι ταυτόχρονα άνω και κάτω ημισυνεχής. Τα Ψ_a λέγονται *level sets* της συνάρτησης f. Εάν ο $\mathcal{X}$ είναι μετρικός χώρος, τότε η κλειστότητα καθορίζεται από ακολουθίες, κατά τον γνωστό τρόπο της ανάλυσης (ένα σύνολο $A \subseteq \mathcal{X}$ είναι κλειστό εάν τα όρια των συγκλίνουσων ακολουθιών του A ανήκουν στο A). Εάν ο $\mathcal{X}$ είναι γενικός τοπολογικός χώρος, τότε η κλειστότητα καθορίζεται από δίκτυα, αρκεί να μπορούμε να βρούμε σε αυτόν ένα διατεταγμένο σύνολο. Περισσότερες πληροφορίες μπορεί κανείς να βρει στο [17].

Έστω για παράδειγμα ότι ο $\mathcal{X}$ είναι μετρικός χώρος. Υπάρχει ένας ισοδύναμος χαρακτηρισμός για τις κάτω ημισυνεχείς συναρτήσεις, που περιλαμβάνει τη σύγκλιση ακολουθιών. Συγκεκριμένα, η f είναι κάτω ημισυνεχής αν και μόνο αν για κάθε ακολουθία $\{x_n\}_{n\in\mathbb{N}} \subseteq \mathcal{X}$ με $x_n \to x \in \mathcal{X}$:

$$\liminf_{n\to+\infty} f(x_n) \geq f(x)$$

Ακόμα, η επόμενη Πρόταση είναι πολύ χρήσιμη και θα τη χρειαστούμε παρακάτω (βλ. [15]).

Πρόταση 0.1. *Έστω ένας τοπολογικός χώρος $\mathcal{X}$ και μια συνάρτηση $f : \mathcal{X} \to [-\infty, +\infty]$, η οποία είναι κάτω ημισυνεχής. Εάν τα Ψ_a είναι συμπαγή, τότε σε κλειστά σύνολα η συνάρτηση f λαμβάνει την ελάχιστη τιμή της*

Τη συνάρτηση $I : \mathcal{X} \to [0, +\infty]$ θα την λέμε *συνάρτηση ρυθμού*, εάν είναι μια κάτω ημισυνεχής συνάρτηση. Εάν ακόμα ισχύει ότι τα Ψ_a είναι συμπαγή, τότε θα ονομάζουμε την I *καλή συνάρτηση ρυθμού*. Από την παραπάνω πρόταση, μια καλή συνάρτηση ρυθμού πιάνει το infimum της σε κλειστά σύνολα. Θα συμβολίζουμε με $\mathcal{B}(\mathcal{X})$ την κλάση συνόλων Borel στον τοπολογικό χώρο $\mathcal{X}$. Η επόμενη έννοια αποτελεί τον θεμέλιο λίθο της θεωρίας των μεγάλων αποκλίσεων.

Ορισμός 0.1 (Αρχή των Μεγάλων Αποκλίσεων). *Τα μέτρα πιθανότητας $\{\mu_n\}$ θα λέμε ότι ικανοποιούν την αρχή των μεγάλων αποκλίσεων με συνάρτηση ρυθμού I, εάν για $A \in \mathcal{B}(\mathcal{X})$:*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mu_n(A) \leq - \inf_{x\in\bar{A}} I(x)$$
$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu_n(A) \geq - \inf_{x\in A^o} I(x)$$

Τα δεξιά μέλη ονομάζονται άνω και κάτω φράγμα αντίστοιχα.

Επίσης, δεν είναι δύσκολο να δει κανείς (βλ. [15]) ότι η παραπάνω συνθήκη είναι ισοδύναμη με την:

$$
\begin{aligned}
\limsup_{n\to+\infty} \frac{1}{n}\log\mu_n(F) &\leq -\inf_{x\in F} I(x), \quad \text{εάν } F\subseteq \mathcal{X} \text{ κλειστό} \\
\liminf_{n\to+\infty} \frac{1}{n}\log\mu_n(G) &\geq -\inf_{x\in G} I(x), \quad \text{εάν } G\subseteq \mathcal{X} \text{ ανοιχτό}
\end{aligned} \tag{1.4}
$$

Είναι φυσικό να διερωτηθεί κανείς, εάν υπάρχει πάντοτε μια τέτοια συνάρτηση I που να ικανοποιεί την Αρχή των Μεγάλων Αποκλίσεων. Για την απόδειξη, παραπέμπουμε στο θεώρημα 2.13. στο [20].

Στην εργασία θα χρησιμοποιούμε την συντομογραφία LDP (Large Deviations Principle) για να αναφερόμαστε στην αρχή των μεγάλων αποκλίσεων. Υπάρχει και μια ασθενέστερη εκδοχή της 1.4, η οποία ονομάζεται *ασθενής αρχή των μεγάλων αποκλίσεων* (weak LDP):

$$
\begin{aligned}
\limsup_{n\to+\infty} \frac{1}{n}\log\mu_n(K) &\leq -\inf_{x\in F} I(x), \quad \text{εάν } K\subseteq \mathcal{X} \text{ συμπαγές} \\
\liminf_{n\to+\infty} \frac{1}{n}\log\mu_n(G) &\geq -\inf_{x\in G} I(x), \quad \text{εάν } G\subseteq \mathcal{X} \text{ ανοιχτό}
\end{aligned} \tag{1.5}
$$

Δηλαδή το άνω φράγμα ισχύει για συμπαγή και όχι για κλειστά σύνολα. Φυσικά, θα ήταν πολύ ωραίο κανείς να μπορούσε να χρησιμοποιήσει την 1.5 για να συμπεράνει την 1.4, αφού με τη συμπάγεια δουλεύουμε ευκολότερα στις αποδείξεις, όπως θα δούμε παρακάτω. Αυτό όμως δεν γίνεται πάντοτε. Τη θέση του διαμεσολαβητή για την μετάβαση από την ασθενή LDP στην πλήρη LDP κατέχει η έννοια της *εκθετικής σφιχτότητας*:

Ορισμός 0.2. *Η ακολουθία των μέτρων πιθανότητας $\{\mu_n\}_{n\in\mathbb{N}}$ στον $\mathcal{X}$ ονομάζεται εκθετικά σφιχτή, εάν για κάθε $a<\infty$ υπάρχει ένα συμπαγές $K_a\subset\mathcal{X}$, ώστε:*

$$\limsup_{n\to+\infty} \frac{1}{n}\log\mu_n(K_a^c) < -a$$

Η σύνδεση που περιγράψαμε πριν διατυπώνεται και αποδεικνύεται στο παρακάτω Λήμμα:

Λήμμα 1. *Εάν η ακολουθία $\{\mu_n\}_{n\in\mathbb{N}}$ μέτρων πιθανότητας είναι εκθετικά σφιχτή, τότε η 1.5 συνεπάγεται την 1.4, δηλαδή η ασθενής LDP συνεπάγεται την πλήρη LDP.*

Απόδειξη. Χρειάζεται να αποδείξουμε μόνο ότι το άνω φράγμα της 1.5 επιτυγχάνεται εν τέλει για κλειστά σύνολα. Το κάτω φράγμα παραμένει το ίδιο.

Έστω $a<\infty$ και ένα συμπαγές σύνολο $K\subset\mathcal{X}$, όπως στον ορισμό της εκθετικής σφιχτότητας. Εάν $F \subset \mathcal{X}$ είναι κλειστό, τότε το $F \bigcap K$ είναι συμπαγές (ως η τομή κλειστού και συμπαγούς συνόλου). Άρα:

$$\begin{aligned}
\limsup_{n\to+\infty} \frac{1}{n}\log \mu_n(F) &\leq \limsup_{n\to+\infty} \frac{1}{n}\log\left(\mu_n(F\cap K) + \mu_n(K^c)\right) \\
&\leq \limsup_{n\to+\infty} \left(\frac{1}{n}\log \mu_n(F\cap K) \vee \frac{1}{n}\log \mu_n(K^c)\right) \\
&\leq (-\inf_{x\in F\cap K} I(x)) \vee (-a) \\
&\leq (-\inf_{x\in F} I(x)) \vee (-a)
\end{aligned}$$

Η δεύτερη ανισότητα προέκυψε λόγω της ταυτότητας $a + b \leq 2(a \vee b)$, ενώ η τρίτη προέκυψε από την ασθενή LDP και την εκθετική σφιχτότητα. Στέλνοντας $a \to +\infty$ παίρνουμε:

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

δηλαδή το άνω φράγμα της πλήρους LDP. □

Η εκθετική σφιχτότητα μας δίνει κάτι ακόμα ισχυρό, που διατυπώνεται στο επόμενο Λήμμα.

Λήμμα 2. *Έστω ότι τα μέτρα πιθανότητας $\{\mu_n\}_{n\in\mathbb{N}}$ στον πολωνικό χώρο $\mathcal{X}$ ικανοποιούν την ασθενή LDP με συνάρτηση ρυθμού I. Εάν τα $\{\mu_n\}_{n\in\mathbb{N}}$ είναι εκθετικά σφιχτά, τότε η I είναι καλή συνάρτηση ρυθμού.*

Απόδειξη. Έστω $a < +\infty$ και $K_a \subset \mathcal{X}$, ώστε να ικανοποιείται η εκθετική σφχιτότητα. Καθώς ο $\mathcal{X}$ είναι μετρικός, το K_a είναι κλειστό (αφού είναι συμπαγές), άρα το K_a^c είναι ανοιχτό. Από το κάτω φράγμα της LDP θα έχουμε:

$$-\inf_{x\in K_a^c} I(x) \leq \liminf_{n\to+\infty} \frac{1}{n}\log \mu_n(K_a^c) \leq \limsup_{n\to+\infty} \frac{1}{n}\log \mu_n(K_a^c) < -a$$

λόγω της εκθετικής σφιχτότητας. Άρα:

$$\inf_{x\in K_a^c} I(x) > a$$

Άρα εάν $x \in \mathcal{X}$ είναι τέτοιο ώστε $I(x) \leq a$, τότε $x \in K_a$. Ειδάλλως, εάν $x \in K_a^c$ θα είχαμε ότι $I(x) \leq a < \inf_{x\in K_a^c} I(x)$, δηλαδή θα καταλήγαμε σε αντίφαση. Άρα:

$$\{x \in \mathcal{X} : I(x) \leq a\} \subset K_a$$

Εφόσον ικανοποιείται η ασθενής LDP με συνάρτηση ρυθμού I, τότε τα $\{x \in \mathcal{X} : I(x) \leq b\}$ είναι κλειστά για κάθε $b \in \mathbb{R}$. Επιπλέον, εάν διαλέξουμε $a > b$ και $a > 0$, τότε το $\{x \in \mathcal{X} : I(x) \leq b\}$ περιέχεται στο K_a, αφού ισχύει:

$$\{x \in \mathcal{X} : I(x) \leq b\} \subset \{x \in \mathcal{X} : I(x) \leq a\} \subset K_a$$

όμως το K_a είναι συμπαγές και κάθε κλειστό υποσύνολο συμπαγούς συνόλου είναι συμπαγές. Άρα, τα $\{x \in \mathcal{X} : I(x) \leq b\}$ είναι συμπαγή για κάθε $b \in \mathbb{R}$, δηλαδή η I είναι καλή συνάρτηση ρυθμού. □

Συνεπώς, εάν αποδείξουμε ότι ισχύει η ασθενής LDP με συνάρτηση ρυθμού I για τα εκθετικώς σφιχτά μέτρα πιθανότητας $\{\mu_n\}_{n\in\mathbb{N}}$, τότε θα ισχύει γι' αυτά και η πλήρης LDP με *καλή* συνάρτηση ρυθμού I.

1.3 Εκτίμηση της μεγάλης απόκλισης του εμπειρικού μέσου

Στο κεφάλαιο αυτό θα πάρουμε μια πρώτη γεύση της δύναμης των μεγάλων αποκλίσεων. Συγκεκριμένα, θα εκτιμήσουμε την μεγάλη απόκλιση του εμπειρικού μέσου ισόνομων, ανεξάρτητων τυχαίων μεταβλητών. Την εκτίμηση αυτή θα μας τη δώσει το θεώρημα του Cramér. Η συζήτηση αυτή είναι χρήσιμη, καθώς μας δίνει μια αίσθηση του τρόπου με τον οποίον δουλεύουν οι μεγάλες αποκλίσεις, καθώς και η αλλαγή μέτρου (εκθετικό πλάγιασμα) που μετατρέπει άτυπα ενδεχόμενα σε τυπικά. Πριν όμως αποδείξουμε το θεώρημα του Cramér θα αφιερώσουμε χρόνο στην παρουσίαση αποτελεσμάτων που προκύπτουν από τον συνδυασμό της θεωρίας πιθανοτήτων και της κυρτής ανάλυσης.

1.3.1 Το θεώρημα του Cramér στον $\mathbb{R}$

Έστω $\{X_n\}_{n\in\mathbb{N}}$ ανεξάρτητες, ισόνομες τυχαίες μεταβλητές ορισμένες σε έναν χώρο πιθανότητας $(\Omega, \mathcal{F}, \mathbb{P})$ με τιμές στον $\mathbb{R}$ και κοινή κατανομή μ. Υποθέτουμε ότι η ροπογεννήτριά τους είναι πεπερασμένη, δηλαδή:

$$M(\lambda) = \int_{\mathbb{R}} e^{\lambda x} d\mu(x) < \infty, \quad \forall \lambda \in \mathbb{R}$$

Η παραπάνω συνθήκη ονομάζεται *συνθήκη του Cramér*. Οι κατανομές που δεν ικανοποιούν τη συνθήκη αυτή ονομάζονται *κατανομές με βαριές ουρές* και αποτελούν το αντικείμενο μελέτης του Κεφαλαίου 4. Την ιδιότητα της πεπερασμένης ροπογεννήτριας την ικανοποιούν όλες σχεδόν οι γνωστές κατανομές, όπως η ομοιόμορφη, η κανονική, η Bernoulli, η εκθετική κλπ. Ενδιαφέρον είναι το γεγονός ότι η συνθήκη του Cramér μας επιτρέπει να παραγωγίσουμε την $M(\lambda)$, δηλαδή να περάσουμε την παράγωγο εντός του ολοκληρώματος (ή ισοδύναμα, εντός της μέσης τιμής).

Πράγματι, έστω λ τέτοιο ώστε $M(\lambda) < \infty$ και $\delta > 0$, ώστε να ισχύει $M(\tilde{\lambda}) < \infty$ για $\tilde{\lambda} \in [\lambda - \delta, \lambda + \delta]$. Τότε, για $-\delta \le \varepsilon \le \delta$ έχουμε:

$$\frac{M(\lambda+\varepsilon) - M(\lambda)}{\varepsilon} = \mathbb{E}\left(\frac{e^{(\lambda+\varepsilon)X_1} - e^{\lambda X_1}}{\varepsilon}\right) = \mathbb{E}\left(e^{\lambda X_1}\frac{e^{\varepsilon X_1} - 1}{\varepsilon}\right)$$

Θέλουμε στην ουσία να πάρουμε το όριο για $\varepsilon \to 0$ στην τελευταία σχέση, εντός της μέσης τιμής. Για την εναλλαγή μέσης τιμής και ορίου θα χρησιμοποιήσουμε το θεώρημα κυριαρχημένης σύγκλισης. Συγκεκριμένα, για ε πολύ μικρό αναπτύσσουμε σε δυναμοσειρά:

$$\frac{e^{\varepsilon X_1}-1}{\varepsilon} = X_1 + \frac{\varepsilon X_1^2}{2} + \ldots \leq |X_1| + \frac{\delta X_1^2}{2} + \ldots = \frac{e^{\delta |X_1|}-1}{\delta}$$

Οπότε:

$$e^{\lambda X_1}\frac{e^{\varepsilon X_1}-1}{\varepsilon} \leq e^{\lambda X_1}\frac{e^{\delta |X_1|}-1}{\delta} \leq \frac{e^{(\lambda-\delta)X_1}+e^{(\lambda+\delta)X_1}}{\delta}$$

Η τελευταία ανισότητα ισχύει εξαιτίας της ταυτότητας:

$$e^{\delta|X_1|}-1 \leq e^{-\delta X_1}+e^{\delta X_1}$$

Τώρα, οι ποσότητες στο δεξί μέλος της παραπάνω ανισότητας έχουν πεπερασμένη μέση τιμή, αφού έτσι διαλέξαμε τα $\lambda-\delta$ και $\lambda+\delta$. Επίσης, δεν εξαρτώνται από το ε, ενώ είναι Borel μετρήσιμες ως συνεχείς μετασχηματισμοί Borel μετρήσιμων συναρτήσεων. Συνεπώς, το θεώρημα κυριαρχημένης σύγκλισης εφαρμόζεται και μας δίνει:

$$M'(\lambda) = \lim_{\varepsilon\to 0}\frac{M(\lambda+\varepsilon)-M(\lambda)}{\varepsilon} = \mathbb{E}\left(e^{\lambda X_1}\lim_{\varepsilon\to 0}\frac{e^{\varepsilon X_1 - 1}}{\varepsilon}\right) = \mathbb{E}[X_1 e^{\lambda X_1}]$$

Αφού ορίσαμε λοιπόν την $M(\lambda)$ και είδαμε ότι είναι παραγωγίσιμη υπό τη συνθήκη του Cramér, ορίζουμε ως $\Lambda(\lambda) = \log M(\lambda)$ τη λογαριθμοποιημένη ροπογεννήτρια της X_1. Ορίζουμε επιπλέον τον μετασχηματισμό Fenchel-Legendre της $\Lambda(\lambda)$:

$$I(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\}$$

Η συνάρτηση I θα είναι και η συνάρτηση ρυθμού στο θεώρημα του Cramér. Πριν το δούμε αυτό όμως, ας διατυπώσουμε κάποιες ιδιότητες της I, που θα μας φανούν πολύ χρήσιμες στην πορεία. Θα συμβολίζουμε με $\bar{x} = \mathbb{E}[X_1]$ τη μέση τιμή των X_n. Παρακάτω υποθέτουμε ότι $\bar{x} < \infty$ για χάριν απλότητος. Όμως τα αποτελέσματα γενικεύονται για $\bar{x} = +\infty$ ή $-\infty$.

Ιδιότητες της I:

1. $I(x) \geq 0$.

   Αυτό δεν είναι δύσκολο να το δούμε. Πράγματι, αφού $\Lambda(0) = \log\mathbb{E}[1] = 0$, τότε:

   $$I(x) \geq 0\cdot x - \Lambda(0) = 0$$

2. Η I είναι κυρτή.

   Αυτό προκύπτει από τον ορισμό της, αφού για $a \in [0,1]$:

   $$\begin{aligned} aI(x_1)+(1-a)I(x_2) &= \sup_{\lambda\in\mathbb{R}}\{a\lambda x_1 - a\Lambda(\lambda)\} + \sup_{\lambda\in\mathbb{R}}\{(1-a)\lambda x_2 - (1-a)\Lambda(\lambda)\} \\ &\geq \sup_{\lambda\in\mathbb{R}}\{a\lambda x_1 - a\Lambda(\lambda) + (1-a)\lambda x_2 - (1-a)\Lambda(\lambda)\} \\ &= \sup_{\lambda\in\mathbb{R}}\{(ax_1+(1-a)x_2)\lambda - \Lambda(\lambda)\} \\ &= I(ax_1+(1-a)x_2) \end{aligned}$$

3. Η I είναι κάτω ημισυνεχής.

   Αφού είμαστε στον $\mathbb{R}$ θα χρησιμοποιήσουμε τον ισοδύναμο, ακολουθιακό ορισμό της κάτω ημισυνέχειας. Συγκεκριμένα, έστω $\{x_n\}_{n\in\mathbb{N}}$, ώστε $x_n \to x$. Τότε, για κάθε $\lambda \in \mathbb{R}$:

   $$\liminf_{n\to+\infty} I(x_n) \geq \liminf_{n\to+\infty}(\lambda x_n - \Lambda(\lambda)) = \lambda x - \Lambda(\lambda)$$

   Οπότε αφού αυτό ισχύει για κάθε $\lambda \in \mathbb{R}$:

   $$\liminf_{n\to+\infty} I(x_n) \geq \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\} = I(x)$$

   δηλαδή η I είναι κάτω ημισυνεχής.

4. $I(\bar{x}) = 0$ και $\inf\limits_{x\in\mathbb{R}} I(x) = 0$.

   Η ανισότητα του Jensen μας δίνει:

   $$\Lambda(\lambda) = \log \mathbb{E}[e^{\lambda X_1}] \geq \mathbb{E}[\log e^{\lambda X_1}] = \lambda\bar{x}$$

   Δηλαδή $\lambda\bar{x} - \Lambda(\lambda) \leq 0$ για κάθε $\lambda \in \mathbb{R}$. Άρα:

   $$I(\bar{x}) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\} \leq 0$$

   Όμως από παραπάνω ιδιότητα πρέπει $I(\bar{x}) \geq 0$. Συνεπώς, $I(\bar{x}) = 0$.

   Τέλος, το γεγονός ότι βρήκαμε κάποιο $x \in \mathbb{R}$ ώστε $I(x) = 0$ μας υποδεικνύει ότι $\inf_{x\in\mathbb{R}} I(x) = 0$, αφού $I(x) \geq 0$ για όλα τα $x \in \mathbb{R}$.

5. Η I είναι αύξουσα για $x > \bar{x}$ και φθίνουσα για $x < \bar{x}$.

   Έστω ότι $x \geq \bar{x}$. Τότε, για $\lambda < 0$ έχουμε:

   $$\lambda x - \Lambda(\lambda) \leq \lambda\bar{x} - \Lambda(\lambda) \leq I(\bar{x}) - \Lambda(\lambda) \leq I(\bar{x}) = 0$$

   Και άρα: $I(x) = \sup\limits_{\lambda\geq 0}\{\lambda x - \Lambda(\lambda)\}$. Επειδή η $\lambda x - \Lambda(\lambda)$ είναι αύξουσα ως συνάρτηση του x για $\lambda \geq 0$, τότε η I είναι επίσης αύξουσα για $x \geq \bar{x}$, αφού το sup διατηρεί την διάταξη, δηλαδή για $\lambda \geq 0$ εάν $\bar{x} \leq x_1 \leq x_2$:

   $$I(x_1) = \sup_{\lambda\geq 0}\{\lambda x_1 - \Lambda(\lambda)\} \leq \sup_{\lambda\geq 0}\{\lambda x_2 - \Lambda(\lambda)\} = I(x_2)$$

   Έστω τώρα ότι $x \leq \bar{x}$. Εάν $\lambda > 0$ θα έχουμε:

   $$\lambda x - \Lambda(\lambda) \leq \lambda\bar{x} - \Lambda(\lambda) \leq I(\bar{x}) = 0$$

   Και άρα: $I(x) = \sup\limits_{\lambda\leq 0}\{\lambda x - \Lambda(\lambda)\}$. Η συνέχεια της απόδειξης είναι ίδια με πριν.

Είμαστε έτοιμοι να διατυπώσουμε το θεώρημα του Cramér.

Θεώρημα 3 (Θεώρημα του Cramér). *Η οικογένεια κατανομών* $\{\mu_n\}_{n\in\mathbb{N}}$ *των εμπειρικών μέσων* $\frac{1}{n}\sum_{i=1}^{n} X_i$ *ικανοποιεί την LDP με συνάρτηση ρυθμού:*

$$I(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\}$$

Απόδειξη. Για το άνω φράγμα της LDP θεωρούμε ένα κλειστό, μη-κενό σύνολο $F \subset \mathbb{R}$. Εάν $\inf_{x\in F} I(x) = 0$, τότε το άνω φράγμα ικανοποιείται, αφού το αριστερό μέλος είναι πάντοτε μη-θετικό. Άρα, έστω ότι $\inf_{x\in F} I(x) > 0$. Για κάθε $x \in \mathbb{R}$ και $\lambda \geq 0$ παίρνουμε λόγω της ανισότητας Markov:

$$\begin{aligned}
\mu_n([x,\infty)) &= \mathbb{P}\left(\frac{S_n}{n} \geq x\right) \\
&= \mathbb{P}\left(e^{\lambda S_n} \geq e^{\lambda n x}\right) \\
&\leq e^{-\lambda n x}\mathbb{E}[e^{\lambda S_n}] \\
&= e^{-\lambda n x}\mathbb{E}[e^{\lambda X_1}]^n \\
&= e^{-\lambda n x}M^n(\lambda) \\
&= e^{-n(\lambda x - \Lambda(\lambda))}
\end{aligned}$$

Άρα για κάθε $x \geq \bar{x}$, βελτιστοποιώντας το δεξί μέλος και εφαρμόζοντας την ιδιότητα από την απόδειξη της ιδιότητας 5.:

$$\mu_n([x,\infty)) \leq e^{-nI(x)} \tag{1.6}$$

Όμοια για $\lambda < 0$, $x < \bar{x}$ και εφαρμόζοντας την ανισότητα Markov:

$$\begin{aligned}
\mu_n((-\infty,x]) &= \mathbb{P}\left(\frac{S_n}{n} \leq x\right) \\
&= \mathbb{P}\left(e^{\lambda S_n} \geq e^{\lambda n x}\right) \\
&\leq e^{-\lambda n x}\mathbb{E}[e^{\lambda S_n}] \\
&= e^{-\lambda n x}\mathbb{E}[e^{\lambda X_1}]^n \\
&= e^{-\lambda n x}M^n(\lambda) \\
&= e^{-n(\lambda x - \Lambda(\lambda))}
\end{aligned}$$

Και άρα βελτιστοποιώντας ως προς λ στο δεξί μέλος και χρησιμοποιώντας την αντίστοιχη σχέση από την απόδειξη της ιδιότητας 5. για την I:

$$\mu_n((-\infty,x]) \leq e^{-nI(x)} \tag{1.7}$$

Τώρα, αφού $I(\bar{x}) = 0$ και καθώς έχουμε υποθέσει ότι $\inf_{x\in F} I(x) > 0$, τότε το $\bar{x}$ δε μπορεί να ανήκει στο F, άρα θα περιέχεται στο F^c, που είναι ανοιχτό, αφού το F είναι κλειστό.

Έστω (x_1, x_2) η ένωση όλων των ανοιχτών διαστημάτων $(\alpha, \beta) \subset F^c$ που περιέχουν το $\bar{x}$, δηλαδή το μεγαλύτερο ανοιχτό υποδιάστημα του F^c που περιέχει το $\bar{x}$. Ας παρατηρήσουμε ότι είτε το x_1 είτε το x_2 πρέπει να είναι πεπερασμένα, ειδάλλως $F^c = \mathbb{R} \Rightarrow F = \emptyset$, αλλά έχουμε υποθέσει το F μη κενό.

Εάν το x_1 είναι πεπερασμένο, τότε θα περιέχεται στο F, αφού δεν περιέχεται στο F^c (καθώς το διάστημα (x_1, x_2) είναι ανοιχτό). Συνεπώς, $I(x_1) \geq \inf\limits_{x \in F} I(x)$. Ομοίως, εάν το x_2 είναι πεπερασμένο, τότε θα περιέχεται στο F και άρα $I(x_2) \geq \inf\limits_{x \in F} I(x)$.

Παρατηρούμε επίσης ότι μπορούμε να εφαρμόσουμε την (1.6) για $x = x_2$, αφού $x_2 \geq \bar{x}$ και την (1.7) για $x = x_1$, αφού $x_1 \leq \bar{x}$. Τέλος, έχουμε:

$$(x_1, x_2) \subseteq F^c \Leftrightarrow F \subseteq (x_1, x_2)^c = (-\infty, x_1] \cup [x_2, +\infty)$$

Βάζοντας όλα αυτά μαζί και χρησιμοποιώντας την υποπροσθετικότητα του μέτρου:

$$\begin{aligned}
\mu_n(F) &\leq \mu_n((-\infty, x_1] \cup [x_2, +\infty)) \\
&\leq \mu_n((-\infty, x_1]) + \mu_n([x_2, +\infty)) \\
&\leq 2e^{-nI(x)}
\end{aligned}$$

Και άρα:

$$\mu_n(F) \leq 2e^{-n \inf\limits_{x \in F} I(x)}$$

Οπότε:

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

Δηλαδή πήραμε το άνω φράγμα.

Η περίπτωση του κάτω φράγματος είναι πιο ενδιαφέρουσα και αισθητικά κομψή. Θα δείξουμε πρώτα ότι αν $x \geq \bar{x}$ και $\delta > 0$, τότε:

$$\liminf_{n \to +\infty} \log \mu_n([x, x+\delta)) = \liminf_{n \to +\infty} \frac{1}{n} \log \mathbb{P}\left(\{x \leq \frac{S_n}{n} < x + \delta\right) \geq -I(x)$$

Έστω ότι υπάρχει πεπερασμένο λ^*, ώστε:

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\} = \lambda^* x - \Lambda(\lambda *)$$

Η συνθήκη του Cramér μας επιτρέπει να παραγωγίσουμε την $M(\lambda)$, περνώντας την παράγωγο εντός της μέσης τιμής, όπως είδαμε σε προηγούμενη συζήτηση. Συνεπώς, αφού το sup λαμβάνεται, θα πρέπει:

$$x - \Lambda'(\lambda^*) = 0 \Leftrightarrow x = \frac{M'(\lambda^*)}{M(\lambda^*)} = \frac{\mathbb{E}[X_1 e^{\lambda^* X_1}]}{M'(\lambda^*)} = \int y \frac{e^{\lambda^* y}}{M(\lambda^*)} d\mu$$

Από τη σχέση αυτή παρατηρούμε ότι εάν ορίσουμε ένα νέο μέτρο πιθανότητας μ^*:

$$\frac{d\mu^*}{d\mu}(y) = \frac{e^{\lambda^* y}}{M(\lambda^*)}$$

Τότε κάποια τυχαία μεταβλητή με κατανομή την μ^* θα έχει μέση τιμή x. Έστω ότι μ_n^* είναι η κατανομή του μέσου όρου $\frac{S_n}{n}$, όταν οι X_i είναι ανεξάρτητες, ισόνομες τυχαίες μεταβλητές με κατανομή μ^*. Αυτό σημαίνει ότι ο μέσος όρος των X_i συγκεντρώνεται γύρω από το x, κάτω από την νέα, αλλαγμένη κατανομή. Η αλλαγή μέτρου λοιπόν έκανε τυπική την προηγουμένως άτυπη συμπεριφορά των X_i. Έστω $\varepsilon > 0$ με $\varepsilon < \delta$. Παίρνουμε ότι:

$$\begin{aligned}
\mu_n([x, x+\delta)) &= \int_{nx \le \sum_{i=1}^n x_i \le n(x+\delta)} d\mu(x_1)\dots d\mu(x_n) \\
&\ge \int_{nx \le \sum_{i=1}^n x_i \le n(x+\varepsilon)} d\mu(x_1)\dots d\mu(x_n) \\
&\ge \int_{nx \le \sum_{i=1}^n \le n(x+\varepsilon)} e^{\lambda^* \sum_{i=1}^n x_i - n\lambda^* x - n|\lambda^*|\varepsilon} d\mu(x_1)\dots d\mu(x_n) \\
&= e^{-n\lambda^* x - n|\lambda^*|\varepsilon} M^n(\lambda^*) \mu_n^* \left([x, x+\varepsilon]\right)
\end{aligned}$$

Εάν $\mathbb{P}^*$ είναι το νέο μέτρο πιθανότητας που υιοθετείται εξαιτίας του μ^*, τότε από το κεντρικό οριακό θεώρημα θα έχουμε:

$$\lim_{n\to+\infty} \mu_n^*([x, x+\varepsilon]) = \lim_{n\to+\infty} \mathbb{P}^* \left(0 \le \frac{S_n}{n} - x \le \varepsilon\right) = \lim_{n\to+\infty} \mathbb{P}^* \left(0 \le \frac{\frac{S_n}{n} - x}{\sigma/\sqrt{n}} \le \frac{\varepsilon\sqrt{n}}{\sigma}\right) = \frac{1}{2}$$

Συνεπώς,

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu_n([x, x+\delta)) \ge -(\lambda^* x - \Lambda(\lambda^*)) - |\lambda^*|\varepsilon = -I(x) - |\lambda^*|\varepsilon$$

Οπότε στέλνοντας το $\varepsilon \to 0$ παίρνουμε την 1.3.1.

Έστω τώρα ότι δεν υπάρχει τέτοιο λ^*. Αυτό μπορεί να συμβαίνει μόνο εάν $\mu((x, +\infty)) = 0$. Στην περίπτωση αυτή θα έχουμε ότι:

$$\lambda x - \Lambda(\lambda) = \log e^{\lambda x} - \log \int_{-\infty}^{x} e^{\lambda y} d\mu(y) = -\log \int_{-\infty}^{x} e^{\lambda(y-x)} d\mu(y)$$

Οπότε για τη συνάρτηση I έχουμε:

$$I(x) = \sup_{\lambda\in\mathbb{R}} \{\lambda x - \Lambda(\lambda)\} = -\lim_{\lambda\to\infty} \{\log \int_{-\infty}^{x} e^{\lambda(y-x)} d\mu(y)\} = -\log \mu(\{x\})$$

Όμως ισχύει ότι:

$$\mu_n([x, x+\delta]) \ge \mathbb{P}\left(X_1 = \dots = X_n = x\right) = \mu^n(\{x\})$$

Συνεπώς:

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu_n([x, x+\delta]) \geq \log \mu(\{x\}) = -I(x)$$

Οπότε πήραμε ξανά την (1.3.1).

Εάν όμως G είναι ένα ανοιχτό υποσύνολο του $\mathbb{R}$ και $x \in G$, τότε θα υπάρχει $\delta > 0$, ώστε $(x-\delta, x+\delta) \subset G$. Ισχύει όμως ότι:

$$\mu_n(G) \geq \mu_n((x-\delta, x+\delta)) \geq \mu_n([x, x+\delta))$$

Συνεπώς, από τη σχέση 1.3.1 παίρνουμε:

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu_n(G) \geq -I(x)$$

Βελτιστοποιώντας ως προς x παίρνουμε το επιθυμητό κάτω φράγμα:

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} I(x)$$

Και η απόδειξη του θεωρήματος του Cramér έχει ολοκληρωθεί. $\square$

Το σπουδαιότερο σημείο στην ανωτέρω απόδειξη είναι αυτό της αλλαγής μέτρου. Πράγματι, κανείς θα μπορούσε να κάνει απλώς έναν μετασχηματισμό στις τυχαίες μεταβλητές X_i, ώστε να κάνει το άτυπο ενδεχόμενο τυπικό. Όμως, έτσι δε θα πετύχαινε το κάτω φράγμα. Η αλλαγή μέτρου αποτελεί γενικότερα μια τέχνη. Η λογική με την οποία την κάναμε εδώ ήταν ώστε να ελαχιστοποιείται η σχετική εντροπία ως προς την αρχική κατανομή των X_i. Φυσικά δεν έχουμε πει τι σημαίνει αυτό, αφού αποτελεί αντικείμενο της αρχής του Gibbs, στην οποία είναι αφιερωμένη η παρούσα εργασία. Το θεώρημα του Cramér μας δίνει μια πρώτη ματιά στο πως μπορούμε να κάνουμε ενδεχόμενα με μικρή πιθανότητα να έχουν μεγάλη πιθανότητα. Αυτή η αλλαγή μέτρου ονομάζεται *εκθετικό πλάγιασμα*. Θα το ξαναδούμε παρακάτω, όταν θα αποδείξουμε το θεώρημα του Sanov.

Αξίζει να σημειώσουμε ότι το θεώρημα του Cramér ισχύει ακόμα κι αν οι X_i δεν είναι ανεξάρτητες. Στην περίπτωση αυτή, το αντίστοιχο θεώρημα ονομάζεται Gartner-Ellis, όμως η απόδειξή του ξεφεύγει από τους σκοπούς της παρούσας εργασίας. Μια εκτενή παρουσίαση γι'αυτό το θεώρημα μπορεί κανείς να βρει στο [15].

Ας παρατηρήσουμε και κάτι ακόμα. Εάν οι X_i ακολουθούν την τυποποιημένη κανονική κατανομή, τότε η συνάρτηση ρυθμού στο θεώρημα του Cramér θα είναι (έπειτα από απλές πράξεις):

$$I(x) = \frac{x^2}{2}, \quad \forall x \in \mathbb{R}$$

Για κάθε $\delta > 0$ θεωρούμε το σύνολο:

$$A_\delta = (-\delta, \delta)^c = (-\infty, -\delta] \cup [\delta, \infty)$$

Παρατηρούμε ότι ισχύει: $\inf_{x \in A_\delta^o} I(x) = \inf_{x \in \bar{A}_\delta} I(x) = \frac{\delta^2}{2}$, αφού η I είναι συνεχής συνάρτηση. Έτσι, από την LDP και σύμφωνα με το θεώρημα του Cramér θα πάρουμε:

$$\lim_{n \to +\infty} \frac{1}{n} \log \mu_n((-\delta, \delta)^c) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\left| \frac{S_n}{n} \right| \geq \delta \right) = -\frac{\delta^2}{2}$$

Που αποτελεί ακριβώς την εκτίμηση που κάναμε στην εισαγωγή. Επομένως, το θεώρημα του Cramér μπορεί να μας δώσει τέτοιου τύπου εκτιμήσεις και για άλλες κατανομές, δυσκολότερες στη μελέτη.

Ακόμα, είναι σημαντικό να σημειώσουμε ότι το θεώρημα του Cramér γενικεύεται και στην περίπτωση του $\mathbb{R}^d$, δηλαδή για τυχαία διανύσματα και τον μέσο όρο τους. Η απόδειξη όμως αυτού είναι απαιτητική και ξεφεύγει από τους σκοπούς της εργασίας.

Κεφάλαιο 2

Μεγάλες αποκλίσεις σε γενικότερους τοπολογικούς χώρους

Στο κεφάλαιο αυτό θα περάσουμε σε ένα μεγαλύτερο στάδιο αφαίρεσης. Συγκεκριμένα, θα μελετήσουμε μεγάλες αποκλίσεις σε πιο αφηρημένους χώρους. Θα αποδείξουμε το λήμμα του Varadhan, το οποίο θα μας οδηγήσει στην απόδειξη του σημαντικού θεωρήματος του Sanov. Το θεώρημα αυτό θα αποτελέσει το τελευταίο μεγάλο εργαλείο που θα χρειαστούμε πριν περάσουμε στην απόδειξη της αρχής του Gibbs.

2.1 Το λήμμα του Varadhan

Τους τοπολογικούς χώρους τους χαρακτηρίζουν τα ανοιχτά τους σύνολα. Υπάρχουν διάφορες κλάσεις τοπολογικών χώρων, ανάλογα με τον τρόπο συνύπραξης ανοιχτών συνόλων και στοιχείων στους χώρους αυτούς. Εμάς μας ενδιαφέρουν δύο τέτοιες κλάσεις τοπολογικών χώρων: οι Hausdorff και οι κανονικοί Hausdorff.

Ένας τοπολογικός χώρος $\mathcal{X}$ ονομάζεται Hausdorff, εάν κάθε δύο στοιχεία του μπορούμε να τα ξεχωρίσουμε από ξένα ανοιχτά σύνολα. Συμβολικά, λέμε ότι ο $\mathcal{X}$ είναι Hausdorff εάν για κάθε $x, y \in \mathcal{X}$ με $x \neq y$ υπάρχουν U γειτονιά του x και V γειτονιά του y, ώστε $U \cap V = \emptyset$.

Σε έναν τοπολογικό χώρο κλειστά ονομάζονται τα σύνολα που είναι συμπληρώματα ανοιχτών. Ένας τοπολογικός χώρος $\mathcal{X}$ ονομάζεται κανονικός χώρος Hausdorff, εάν κλειστά σύνολα και σημεία που δεν ανήκουν σε αυτά μπορούν να διαχωριστούν από ανοιχτά σύνολα. Συμβολικά, ο $\mathcal{X}$ λέγεται κανονικός χώρος Hausdorff, εάν για κάθε $F \subset \mathcal{X}$ κλειστό και $x \notin F$ υπάρχουν G_1, G_2 ανοιχτά, ώστε $F \subset G_1, x \in G_2$ και $G_1 \cap G_2 = \emptyset$. Είναι σημαντικό να θυμόμαστε την εξής παρατήρηση.

Παρατήρηση. Ένας μετρικός χώρος είναι πάντοτε κανονικός.

Απόδειξη. Έστω $(\mathcal{X}, d)$ ένας μετρικός χώρος. Έστω επίσης ένα κλειστό σύνολο $F \subset \mathcal{X}$ και ένα σημείο $x \notin F$. Αρκεί να δείξουμε ότι υπάρχουν $U, W \subset X$ ανοιχτά με $x \in U$ και $F \subset W$, ώστε $U \bigcap V = \emptyset$. Για τον σκοπό αυτό, θέτουμε,

$$\rho = d(x, F) = \inf_{y \in F} d(x, y)$$

και ορίζουμε τα σύνολα,

$$U = \{y \in \mathcal{X} : d(x, y) < \frac{\rho}{2}\}$$
$$W = \{z \in \mathcal{X} : d(z, F) < \frac{\rho}{2}\}$$

Τότε, και τα δύο σύνολα είναι ανοιχτά. Πράγματι, το U είναι η ανοιχτή μπάλα κέντρου x κι ακτίνας $\frac{\rho}{2}$. Για το W, εάν $z \in W$, τότε $\inf_{y \in F} d(z, y) = d(z, F) < \frac{\rho}{2}$. Οπότε, το $V = \{y \in F : d(z, y) < \frac{\rho}{2}\}$ είναι μια ανοιχτή περιοχή του z, αφού $\forall y \in V : d(z, y) < \frac{\rho}{2}$, άρα και $d(z, F) = \inf_{y \in F} d(z, y) < \frac{\rho}{2}$. Ακόμα, εάν $y \in F$, τότε $d(y, F) = 0 < \frac{\rho}{2}$, δηλαδή $y \in W$, άρα $F \subseteq W$.

Αρκεί να αποδείξουμε τώρα ότι $U \cap W = \emptyset$. Έστω $y \in U \cap W$. Τότε,

$$y \in U \Leftrightarrow d(x, y) < \frac{\rho}{2}$$
$$y \in W \Leftrightarrow d(y, F) = \inf_{z \in F} d(y, z) < \frac{\rho}{2}$$

Άρα, ενθυμούμενοι τον ορισμό του ρ,

$$\rho = d(x, F) = \inf_{z \in F} d(x, z)$$

και υπολογίζοντας για τυχαίο $z \in F$ ότι,

$$\begin{aligned} &d(x, z) \leq d(x, y) + d(y, z) \\ &\Rightarrow \rho = d(x, F) = \inf_{z \in F} d(x, z) \leq d(x, y) + \inf_{z \in F} d(y, z) < \frac{\rho}{2} + \frac{\rho}{2} = \rho \end{aligned}$$

καταλήγουμε σε άτοπο. Άρα, $U \cap W = \emptyset$, οπότε τελικά ο $\mathcal{X}$ είναι κανονικός. □

Σχόλιο. Εάν x, y στοιχεία του μετρικού χώρου $(\mathcal{X}, d)$ και $Z \subseteq \mathcal{X}$, μπορούμε να ορίσουμε,

$$B = \{d(x, z) : z \in Z\}$$
$$C = \{d(x, y) + d(y, z) : z \in Z\}$$

και τότε, εάν $c \in C$, θα έχουμε ότι $c = d(x, y) + d(y, z)$ για κάποιο $z \in Z$. Επιπλέον, εάν $b \in B$, τότε $b = d(x, z)$ για κάποιο $z \in Z$, άρα η τριγωνική ανισότητα θα μας δώσει,

$$b = d(x, z) \leq d(x, y) + d(y, z) = c \Rightarrow \inf B \leq c$$

Αυτό όμως ισχύει για κάθε $c \in C$, οπότε το $\inf B$ είναι κάτω φράγμα του συνόλου C και επιπλέον $\inf B \leq \inf C$. Επομένως για κάθε $x, y \in \mathcal{X}$:

$$\begin{aligned} &d(x,z) \leq d(x,y) + d(y,z) \\ &\Rightarrow \inf_{z \in Z} d(x,z) = \inf B \leq \inf C = d(x,y) + \inf_{z \in Z} d(y,z) \end{aligned}$$

Γεγονός που αιτιολογεί την συνεπαγωγή στο τελευταίο βήμα της παραπάνω απόδειξης.

Συνεχίζοντας, γνωρίζουμε ότι οι καλές κάτω ημισυνεχείς συναρτήσεις πιάνουν το ελάχιστό τους σε κλειστά σύνολα. Παρόμοια εκτίμηση δε μπορούμε να έχουμε εν γένει για κάτω ημισυνεχείς συναρτήσεις. Όμως σε κανονικούς χώρους η κατάσταση βελτιώνεται. Συγκεκριμένα, έχουμε το ακόλουθο Λήμμα:

Λήμμα 4. *Εάν $\mathcal{X}$ κανονικός χώρος Hausdorff και $\varphi : \mathcal{X} \to \mathbb{R}$ μια κάτω ημισυνεχής συνάρτηση, τότε για κάθε $x \in \mathcal{X}$ και $\delta > 0$ υπάρχει ανοιχτή περιοχή G του x, ώστε:*

$$\inf_{y \in \bar{G}} \varphi(y) \geq (\varphi(x) - \delta) \wedge \frac{1}{\delta}$$

Απόδειξη. Αφού η φ είναι κάτω ημισυνεχής, τότε για κάθε $c \in \mathbb{R}$ το σύνολο $F = \{y \in \mathcal{X} : \varphi(y) \leq c\}$ είναι εξ ορισμού κλειστό. Διαλέγουμε ως c το δεξί μέλος της σχέσης που θέλουμε να δείξουμε, δηλαδή $c = (\varphi(x) - \delta) \wedge \frac{1}{\delta}$. Τότε, $c \leq \varphi(x) - \delta < \varphi(x)$. Άρα $x \notin F$, εξ ορισμού του F. Σκοπός είναι να δείξουμε ότι η τελευταία σχέση ισχύει όχι μόνο για το x, αλλά και για τιμές γύρω απ' το x, δηλαδή σε μια ανοιχτή περιοχή του.

Η ύπαρξη της ανοιχτής περιοχής του x θα προκύψει από την κανονικότητα του χώρου. Πράγματι, αφού F κλειστό και $\mathcal{X}$ κανονικός, τότε υπάρχει ανοιχτή περιοχή G του x και ανοιχτή περιοχή W του F, ώστε $F \subset W$ και $G \cap W = \emptyset$. Η τελευταία σχέση μας δίνει ότι $G \subset W^c$. Όμως το W είναι ανοιχτό, δηλαδή το W^c είναι κλειστό και περιέχει το G. Επιπλέον, το γεγονός ότι $F \subset W$ μας δίνει ότι $W^c \subset F^c$. Οπότε:

$$\bar{G} \subset W^c \subset F^c$$

Άρα $y \in \bar{G} \Rightarrow y \in F^c$. Οπότε, $\varphi(y) > c$ για κάθε $y \in \bar{G}$.

Βελτιστοποιώντας ως προς $y \in \bar{G}$ παίρνουμε ότι $\inf_{y \in \bar{G}} \varphi(y) \geq c = (\varphi(x) - \delta) \wedge \frac{1}{\delta}$, δηλαδή το ζητούμενο. □

Παρατήρηση. Εάν G είναι η ανοιχτή περιοχή του x στο προηγούμενο Λήμμα, τότε προφανώς $G \subset \bar{G}$, αφού το G είναι ανοιχτό σύνολο. Οπότε, θα ισχύει ότι $\inf_{y \in G} \varphi(y) \geq \inf_{y \in \bar{G}} \varphi(y)$ και τότε η σχέση του παραπάνω Λήμματος ισχύει ακόμα κι αν πάρουμε το inf πάνω στο G και όχι στο $\bar{G}$.

Αφού έχουμε ορίσει τους απαραίτητους χώρους και έχουμε αποδείξει το παραπάνω Λήμμα, έχουμε στα χέρια μας όλα τα εργαλεία που χρειαζόμαστε για να αποδείξουμε το λήμμα του Varadhan. Όμως πρώτα ας κάνουμε μια μικρή συζήτηση, για να εκτιμήσουμε περισσότερο τη δύναμη αυτού του λήμματος.

Στη μελέτη της ασυμπτωτικής εκτίμησης ολοκληρωμάτων υπάρχει ένα διάσημο θεώρημα, αυτό του Laplace, το οποίο μας δίνει ότι αν φ είναι μια συνεχής συνάρτηση στο διάστημα $[0,1]$, τότε:

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^1 e^{n\varphi(x)}dx = \sup_{x\in[0,1]}\varphi(x)$$

Κανείς μπορεί να βρει μια καλή συζήτηση για τέτοιες εκτιμήσεις στο [13].

Μια τέτοια εκτίμηση μας λέει ότι, ενώ πολύ μεγάλες τιμές μπορεί λαμβάνονται σε ένα μικρό μέρος του χώρου μας, κι όμως συχνά αυτές οι τιμές παίζουν καθοριστικό ρόλο στη συμπεριφορά του ολοκληρώματος. Το Λήμμα του Varadhan αποτελεί τη γενίκευση της παραπάνω εκτίμησης για μια ακολουθία κατανομών μ_n που ικανοποιεί την LDP. Αποτελεί στην ουσία μια ακόμα χρήσιμη εφαρμογή της θεωρίας των μεγάλων αποκλίσεων για τον υπολογισμό πιο αφηρημένων ολοκληρωμάτων.

Λήμμα 5 (Varadhan). *Έστω $\{Z_n\}_{n\in\mathbb{N}}$ μια οικογένεια τυχαίων μεταβλητών με αντίστοιχες κατανομές $\{\mu_n\}_{n\in\mathbb{N}}$ και οι οποίες παίρνουν τιμές σε κάποιον κανονικό τοπολογικό χώρο Hausdorff $\mathcal{X}$.*

Έστω ότι οι $\{\mu_n\}_{n\in\mathbb{N}}$ ικανοποιούν την LDP με καλή συνάρτηση ρυθμού $I:\mathcal{X}\to[0,+\infty]$. Έστω επίσης $\varphi:\mathcal{X}\to\mathbb{R}$ μια οποιαδήποτε συνεχής συνάρτηση.

Εάν ισχύει κάποια από τις παρακάτω σχέσεις:

$$\lim_{M\to+\infty}\limsup_{n\to+\infty}\frac{1}{n}\mathbb{E}[e^{n\varphi(Z_n)}\mathbb{1}_{\varphi(Z_n)\geq M}] = -\infty \tag{2.1}$$

$$\limsup_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{\gamma n\varphi(Z_n)}] < +\infty, \text{για κάποιο} \quad \gamma>1 \tag{2.2}$$

Τότε,

$$\lim_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] = \lim_{n\to+\infty}\frac{1}{n}\log\int_X e^{n\varphi(x)}d\mu_n(x) = \sup_{x\in\mathcal{X}}\{\varphi(x)-I(x)\}$$

Απόδειξη. Ενθυμούμενοι ότι μια συνάρτηση είναι συνεχής αν και μόνο αν είναι ταυτόχρονα κάτω και άνω ημισυνεχής, θα αποδείξουμε το παραπάνω λήμμα σε 3 βήματα. Πρώτα θα αποδείξουμε με τη χρήση της (2.1) την επιθυμητή σχέση ως κάτω φράγμα για φ κάτω ημισυνεχή, έπειτα ως άνω φράγμα για φ άνω ημισυνεχή και τέλος θα δείξουμε ότι η (2.2) συνεπάγεται τη (2.1).

- **Κάτω φράγμα:**

Έστω $\varphi : \mathcal{X} \to \mathbb{R}$ κάτω ημισυνεχής συνάρτηση. Έστω επίσης ότι το κάτω φράγμα της LDP ισχύει.

Έστω $x \in \mathcal{X}$ και $\delta > 0$. Αφού η φ είναι κάτω ημισυνεχής και ο $\mathcal{X}$ κανονικός, τότε από το παραπάνω Λήμμα υπάρχει μια περιοχή G του x, ώστε $\inf_{y\in G}\varphi(y) \geq (\varphi(x)-\delta)\wedge \frac{1}{\delta}$. Άρα:

$$\mathbb{E}[e^{n\varphi(Z_n)}] \geq \mathbb{E}[e^{n\varphi(Z_n)}\mathbb{1}_{\{Z_n\in G\}}] \geq \mathbb{E}[e^{n\inf_{y\in G}\varphi(y)}\mathbb{1}_{\{Z_n\in G\}}] = e^{n\inf_{y\in G}\varphi(y)}\mu_n(G)$$

Εφαρμόζουμε τον λογάριθμο στην παραπάνω σχέση, διαιρούμε με n:

$$\frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] \geq \inf_{y\in G}\varphi(y) + \frac{1}{n}\log\mu_n(G) \geq (\varphi(x)-\delta)\wedge\frac{1}{\delta} + \frac{1}{n}\log\mu_n(G)$$

Πηγαίνοντας στο όριο και εφαρμόζοντας την το κάτω φράγμα της LDP για για τις μ_n παίρνουμε ότι:

$$\begin{aligned}\liminf_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] &\geq (\varphi(x)-\delta)\wedge\frac{1}{\delta} + \liminf_{n\to+\infty}\frac{1}{n}\log\mu_n(G)\\ &\geq (\varphi(x)-\delta)\wedge\frac{1}{\delta} - \inf_{y\in G}I(y)\\ &\geq (\varphi(x)-\delta)\wedge\frac{1}{\delta} - I(x)\end{aligned}$$

αφού $I(x) \geq \inf_{y\in G} I(y)$. Το $\delta > 0$ όμως ήταν αυθαίρετο. Οπότε, για $\delta \to 0^+$ παίρνουμε ότι:

$$\liminf_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] \geq \varphi(x) - I(x)$$

Βελτιστοποιώντας την παραπάνω ως προς x παίρνουμε ότι:

$$\liminf_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] \geq \sup_{x\in\mathcal{X}}\{\varphi(x) - I(x)\} \tag{2.3}$$

Πριν προχωρήσουμε στην απόδειξη του άνω φράγματος ας παρατηρήσουμε ότι για την απόδειξη του κάτω φράγματος δε χρειαστήκαμε κάποια από τις (2.1), (2.2). Όντως, είναι στο άνω φράγμα συγκεκριμένα που η ανάγκη για την (2.1) εμφανίζεται.

- **Άνω φράγμα:**

  Έστω $\varphi : \mathcal{X} \to \mathbb{R}$ άνω ημισυνεχής συνάρτηση, για την οποία ισχύει η (2.1). Έστω επίσης ότι το άνω φράγμα της LDP ισχύει για την $\{\mu_n\}_{n\in\mathbb{N}}$ με καλή συνάρτηση ρυθμού $I : \mathcal{X} \to [0, +\infty]$. Διακρίνουμε δύο περιπτώσεις για την φ:

  1. Έστω ότι η φ είναι άνω φραγμένη, δηλαδή υπάρχει σταθερά $M > 0$, ώστε $\varphi(x) \leq M < +\infty$ για κάθε $x \in \mathcal{X}$. Αυτό συνεπάγεται ότι

$\sup_{x\in\mathcal{X}} \varphi(x) \leq M < +\infty$. Τότε, η (2.1) ισχύει τετριμμένα αφού μετά από κάποιο μεγάλο $M > 0$ θα ισχύει ότι $\varphi(x) < M$ για κάθε $x \in \mathcal{X}$, δηλαδή $\mathbb{1}_{\{\varphi(Z_n)\geq M\}} = 0$ για άπειρα $n \in \mathbb{N}$.

Έστω τώρα $a < +\infty, \delta > 0$ και τα σύνολα $\Psi_a = \{x \in \mathcal{X} : I(x) \leq a\}$, που είναι συμπαγή, αφού η I είναι καλή συνάρτηση ρυθμού. Έστω λοιπόν ένα $x \in \Psi_a$. Τότε, αφενός υπάρχει γειτονιά A_x του x, ώστε $\inf_{y\in\bar{A}_x} I(y) \geq (I(x) - \delta) \wedge \dfrac{1}{\delta}$, αφού η I είναι κάτω ημισυνεχής και άρα ισχύει το προηγούμενο Λήμμα. Αφετέρου, ισχύουν οι συνεπαγωγές:

$$\varphi \text{ άνω ημισυνεχής} \Rightarrow -\varphi \text{ κάτω ημισυνεχής} \Rightarrow \inf_{y\in\bar{A}_x}(-\varphi(y)) \geq (-\varphi(x) - \delta) \wedge \frac{1}{\delta}$$

$$\Rightarrow \sup_{y\in\bar{A}_x}(\varphi(y)) \leq (\varphi(x) + \delta) \wedge \left(-\frac{1}{\delta}\right)$$

Το $\bigcup\limits_{x\in\Psi(a)} A_x$ είναι ένα κάλυμμα του συμπαγούς συνόλου $\Psi(a)$. Άρα, υπάρχει πεπερασμένο υποκάλυμμα, δηλαδή κάποιος φυσικός αριθμός $N \in \mathbb{N}$, ώστε $\Psi_a \subset \bigcup\limits_{i=1}^{N} A_{x_i}$. Τότε, κάνουμε την εξής εκτίμηση:

$$\begin{aligned}
\mathbb{E}[e^{n\varphi(Z_n)}] &= \mathbb{E}[e^{n\varphi(Z_n)}\mathbb{1}_{\{Z_n\in\bigcup\limits_{i=1}^{N}A_{x_i}\}}] + E[e^{n\varphi(Z_n)}\mathbb{1}_{\{Z_n\in(\bigcup\limits_{i=1}^{N}A_{x_i})^c\}}] \\
&\leq \mathbb{E}[e^{n\varphi(Z_n)}\sum_{i=1}^{N}\mathbb{1}_{\{Z_n\in A_{x_i}\}}] + E[e^{nM}\mathbb{1}_{\{Z_n\in(\bigcup\limits_{i=1}^{N}A_{x_i})^c\}}] \\
&= \sum_{i=1}^{N}\mathbb{E}[e^{n\varphi(Z_n)}\mathbb{1}_{\{Z_n\in A_{x_i}\}}] + e^{nM}\mu_n((\bigcup_{i=1}^{N}A_{x_i})^c) \\
&\leq \sum_{i=1}^{N} e^{n(\varphi(x_i+\delta)\wedge\left(-\frac{1}{\delta}\right)}\mu_n(\bar{A}_{x_i}) + e^{nM}\mu_n((\bigcup_{i=1}^{N}A_{x_i})^c) \\
&\leq 2\max\{\sum_{i=1}^{N} e^{n(\varphi(x_i+\delta)\wedge\left(-\frac{1}{\delta}\right)}\mu_n(\bar{A}_{x_i}), e^{nM}\mu_n((\bigcup_{i=1}^{N}A_{x_i})^c)\}
\end{aligned}$$

Και άρα περνώντας στο όριο και ενθυμούμενοι ότι η αυθαίρετη ένωση ανοιχτών είναι ανοιχτό (άρα $(\bigcup\limits_{i=1}^{N} A_{x_i})^c$ κλειστό) παίρνουμε τα παρακάτω:

$$\begin{aligned}
&\limsup_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] \leq \\
&\leq \max\{\limsup_{n\to+\infty}\frac{1}{n}\log\left(\sum_{i=1}^{N} e^{n(\varphi(x_i+\delta)\wedge\left(-\frac{1}{\delta}\right)}\mu_n(\bar{A}_{x_i})\right), M + \limsup_{n\to+\infty}\frac{1}{n}\log\mu_n((\bigcup_{i=1}^{N}A_{x_i})^c)\} \\
&\leq \max\{\limsup_{n\to+\infty}\frac{1}{n}\log\left(\max_{i=1,\ldots,N}\left(e^{n(\varphi(x_i+\delta)\wedge\left(-\frac{1}{\delta}\right)}\mu_n(\bar{A}_{x_i})\right)\right), M - \inf_{y\in(\bigcup\limits_{i=1}^{N}A_{x_i})^c} I(y)\}
\end{aligned}$$

$$= \max\{\limsup_{n\to+\infty} \frac{1}{n} \max_{i=1,\dots,N} \left(\log\left(e^{n(\varphi(x_i+\delta)\wedge\left(-\frac{1}{\delta}\right)}\mu_n(\bar{A}_{x_i})\right)\right), M - \inf_{y\in(\bigcup_{i=1}^{N} A_{x_i})^c} I(y)\}$$

$$= \max\{\max_{i=1,\dots,N} \left((\varphi(x_i+\delta) \wedge \left(-\frac{1}{\delta}\right) + \limsup_{n\to+\infty} \frac{1}{n}\log\mu_n(\bar{A}_{x_i})\right), M - \inf_{y\in(\bigcup_{i=1}^{N} A_{x_i})^c} I(y)\}$$

$$\leq \max\{\max_{i=1,\dots,N} \left((\varphi(x_i+\delta) \wedge \left(-\frac{1}{\delta}\right) - \inf_{y\in\bar{A}_{x_i}} I(y)\right), M - a\}$$

$$\leq \max\{\max_{i=1,\dots,N} \left((\varphi(x_i+\delta) \wedge \left(-\frac{1}{\delta}\right) - I(x_i) + \delta\right), M - a\}$$

Όπου στην προτελευταία ανισότητα χρησιμοποιήσαμε το γεγονός ότι:

$$(\bigcup_{i=1}^{N} A_{x_i})^c \in \Psi(a)^c \Rightarrow \inf_{y\in(\bigcup_{i=1}^{N} A_{x_i})^c} I(y) \geq a$$

Οπότε για $\delta \to 0^+$:

$$\begin{aligned}\limsup_{n\to+\infty} \frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] &\leq \max\{\max_{i=1,\dots,N}\{\varphi(x_i) - I(x_i)\}, M - a\} \\ &\leq \max\{\sup_{x\in\mathcal{X}}\{\varphi(x) - I(x)\}, M - a\}\end{aligned}$$

Και τέλος για $a \to +\infty$ και αφού $M > 0$ παίρνουμε ότι:

$$\limsup_{n\to+\infty} \frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] \leq \sup_{x\in\mathcal{X}}\{\varphi(x) - I(x))\}$$

2. Έστω ότι η φ είναι μια αυθαίρετη, συνεχής συνάρτηση. Έστω επίσης $M > 0$. Τότε, προφανώς ισχύει ότι:

$$\begin{cases} e^{n\varphi(Z_n)}, & \varphi(Z_n) < M \\ 0, & \varphi(Z_n) \geq M \end{cases} \leq \begin{cases} e^{n\varphi(Z_n)}, & \varphi(Z_n) < M \\ e^{nM}, & \varphi(Z_n) \geq M \end{cases}$$

και άρα για κάθε $n \in \mathbb{N}$:

$$e^{n\varphi(Z_n)}\mathbb{1}_{\{\varphi(Z_n)<M\}} \leq e^{n(\varphi(Z_n)\wedge M)}$$

οπότε αφού η $\varphi(Z_n) \wedge M$ είναι άνω φραγμένη, θα πάρουμε ότι:

$$\begin{aligned}&\limsup_{n\to+\infty} \frac{1}{n}\log\mathbb{E}[e^{n\varphi(Z_n)}] = \\ &= \limsup_{n\to+\infty} \frac{1}{n}\log\left(\mathbb{E}[e^{n\varphi(Z_n)}\mathbb{1}_{\{\varphi(Z_n)<M\}}] + \mathbb{E}[e^{n\varphi(Z_n)}\mathbb{1}_{\{\varphi(Z_n)\geq M\}}]\right)\end{aligned}$$

$$
\begin{aligned}
&\leq \limsup_{n\to+\infty} \frac{1}{n} \log \Big(\mathbb{E}[e^{n\varphi(Z_n)\wedge M}] + \mathbb{E}[e^{n\varphi(Z_n)} \mathbb{1}_{\{\varphi(Z_n)\geq M\}}] \Big) \\
&\leq \limsup_{n\to+\infty} \frac{1}{n} \log \Big(\max\{\mathbb{E}[e^{n\varphi(Z_n)\wedge M}], \mathbb{E}[e^{n\varphi(Z_n)} \mathbb{1}_{\{\varphi(Z_n)\geq M\}}]\} \Big) \\
&\leq \sup_{x\in\mathcal{X}} \{\varphi(x) \wedge M - I(x)\} \vee \limsup_{n\to+\infty} \frac{1}{n} \log \mathbb{E}[e^{n\varphi(Z_n)} \mathbb{1}_{\{\varphi(Z_n)\geq M\}}]
\end{aligned}
$$

Για $M \to +\infty$ και με τη χρήση της (2.1) παίρνουμε ότι:

$$
\limsup_{n\to+\infty} \frac{1}{n} \log \mathbb{E}[e^{n\varphi(Z_n)}] \leq \sup_{x\in\mathcal{X}} \{\varphi(x) \wedge M - I(x)\}
$$

Και άρα αποδείξαμε το άνω φράγμα για τυχαία άνω ημισυνεχή συνάρτηση φ. Οπότε μας απομένει να αποδείξουμε ότι η (2.2) συνεπάγεται την (2.1):

- **(2.2) $\Rightarrow$ (2.1)**

  Ας ορίσουμε για κάθε $n \in \mathbb{N}$ τη νέα τυχαία μεταβλητή X_n ως: $X_n \triangleq e^{n(\varphi(Z_n)-M)}$ Έστω επίσης κάποιο $\gamma > 1$, ώστε να ισχύει η (2.2). Ας παρατηρήσουμε πρώτα ότι:

  - Εάν $X_n \geq 1$, $\mathbb{P} - \sigma.\beta.$, τότε αφού $\gamma > 1$ θα έχουμε: $X_n \mathbb{1}_{\{X_n \geq 1\}} = X_n \leq (X_n)^\gamma$, $\mathbb{P} - \sigma.\beta.$
  - Εάν $0 \leq X_n < 1$, $\mathbb{P} - \sigma.\beta.$, τότε αφού $\gamma > 1$ θα έχουμε: $X_n \mathbb{1}_{\{X_n \geq 1\}} = 0 < (X_n)^\gamma$, $\mathbb{P} - \sigma.\beta.$

  Άρα πάντοτε θα ισχύει ότι: $X_n \mathbb{1}_{\{X_n \geq 1\}} \leq (X_n)^\gamma$. Τελικά, έχουμε ότι:

$$
\begin{aligned}
e^{-nM}\mathbb{E}[e^{n\varphi(Z_n)} \mathbb{1}_{\{\varphi(Z_n)\geq M\}}] &= \mathbb{E}[e^{n\varphi(Z_n - M)} \mathbb{1}_{\{\varphi(Z_n)\geq M\}}] \\
&= \mathbb{E}[X_n \mathbb{1}_{\{X_n\geq 1\}}] \\
&\leq \mathbb{E}[(X_n)^\gamma] \\
&= \mathbb{E}[e^{\gamma n(\varphi(Z_n)-M)}] \\
&= e^{-\gamma n M}\mathbb{E}[e^{\gamma n\varphi(Z_n)}]
\end{aligned}
$$

  Παίρνοντας όριο ως προς n:

$$
\limsup_{n\to+\infty} \frac{1}{n} \log \mathbb{E}[e^n \varphi(Z_n) \mathbb{1}_{\{\varphi(Z_n)\geq M\}}] \leq -(\gamma - 1)M + \limsup_{n\to+\infty} \frac{1}{n} \log \mathbb{E}[e^{\gamma n\varphi(Z_n)}]
$$

  Κι έπειτα ως προς $M > 0$ (θυμόμαστε ότι $\gamma > 1$):

$$
\lim_{M\to+\infty} \limsup_{n\to+\infty} \frac{1}{n} \mathbb{E}[e^{n\varphi(Z_n)} \mathbb{1}_{\varphi(Z_n)\geq M}] = -\infty
$$

  καταλήγουμε στην (2.1)

Συνεπώς αποδείξαμε το Λήμμα του Varadhan είτε ισχύει η (2.1) είτε η (2.2). □

Από την παραπάνω απόδειξη βλέπουμε ότι αρκεί η φ να είναι μια συνεχής και άνω φραγμένη συνάρτηση για να ισχύει η ζητούμενη σχέση του Λήμματος του Varadhan. Το γεγονός αυτό είναι αρκετά ισχυρό από μόνο του και αξίζει να το διατυπώσουμε ως Πόρισμα:

Πόρισμα 5.1. *Έστω ότι η οικογένεια κατανομών $\{\mu_n\}_{n\in\mathbb{N}}$, ορισμένη σε κάποιον κανονικό τοπολογικό χώρο Hausdorff $\mathcal{X}$, ικανοποιεί την LDP με καλή συνάρτηση ρυθμού $I : \mathcal{X} \to [0, +\infty]$. Έστω επίσης $\varphi : \mathcal{X} \to \mathbb{R}$ μια συνεχής και άνω φραγμένη συνάρτηση. Τότε:*

$$\lim_{n\to+\infty} \frac{1}{n} \log \int_X e^{n\varphi(x)} d\mu_n(x) = \sup_{x\in\mathcal{X}} \{\varphi(x) - I(x)\}$$

Πριν προχωρήσουμε στην επόμενη παράγραφο, αξίζει να κάνουμε μία παρατήρηση που δείχνει τον τρόπο με τον οποίον το λήμμα του Varadhan γενικεύει το αποτέλεσμα του Laplace. Συγκεκριμένα, έστω ότι $\mathcal{X} = [0, 1]$ και για κάθε $n \in \mathbb{N}$: $\mu_n = \lambda$, όπου λ το μέτρο Lebesgue στο $[0, 1]$. Τότε, η οικογένεια κατανομών $\{\mu_n\}_{n\in\mathbb{N}}$ ικανοποιεί την LDP με καλή συνάρτηση ρυθμού την $I(x) \equiv 0$. Πράγματι, αφενός έχουμε τις παρακάτω δύο εκτιμήσεις:

1. Για $F \in \mathcal{B}([0, 1])$ κλειστό:

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mu_n(F) = \limsup_{n\to+\infty} \frac{1}{n} \log \lambda(F) = \begin{cases} -\infty, & \lambda(F) = 0 \\ 0, & \lambda(F) \in (0, 1] \end{cases} \leq 0 = -\inf_{x\in F} I(x)$$

2. Για $G \in \mathcal{B}([0, 1])$ ανοιχτό:

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu_n(G) = \liminf_{n\to+\infty} \frac{1}{n} \log \lambda(G) = 0 \geq 0 = -\inf_{x\in G} I(x)$$

Στη δεύτερη περίπτωση χρησιμοποιήσαμε το γεγονός ότι κάθε μη-κενό, ανοιχτό σύνολο Borel στο $[0, 1]$ έχει θετικό μέτρο Lebesgue. Αυτό είναι συνέπεια του ότι κάθε ανοιχτό σύνολο Borel θα περιλαμβάνει τουλάχιστον ένα ανοιχτό διάστημα και τα ανοιχτά διαστήματα έχουν θετικό μέτρο.

Άρα, η οικογένεια κατανομών $\mu_n = \lambda$ ικανοποιεί την LDP. Αφετέρου, έχουμε ότι η $I(x) = 0$ για $x \in [0, 1]$ είναι καλή συνάρτηση ρυθμού. Πράγματι, αν $a \in \mathbb{R}$, τότε:

$$\Psi_a = \{x \in [0, 1] : I(x) \leq a\} = \{x \in [0, 1] : a \geq 0\} = \begin{cases} [0, 1], & a \geq 0 \\ \emptyset, & a < 0 \end{cases}$$

Το κενό είναι συμπαγές, αφού κάθε πεπερασμένο σύνολο είναι συμπαγές. Ακόμα, το $[0, 1]$ είναι συμπαγές ως κλειστό και φραγμένο διάστημα του $\mathbb{R}$. Άρα τα Ψ_a είναι συμπαγή για κάθε $a \in \mathbb{R}$ και συνεπώς η $I(x)$ είναι καλή συνάρτηση ρυθμού.

Άρα, εάν θεωρήσουμε μια πραγματική συνάρτηση $\varphi \in C([0, 1])$, τότε αυτή θα είναι φραγμένη, αφού είναι συνεχής σε συμπαγές διάστημα. Και τότε για την φ εφαρμόζεται το Λήμμα του Varadhan και παίρνουμε ότι:

$$\lim_{n\to+\infty} \frac{1}{n} \log \int_X e^{n\varphi(x)} d\mu_n(x) = \sup_{x\in\mathcal{X}} \{\varphi(x) - I(x)\}$$

Οπότε αντικαθιστώντας ως μ_n τα μέτρα Lebesgue, $\mathcal{X} = [0,1]$ και $I(x) = 0$ παίρνουμε:

$$\lim_{n\to+\infty} \frac{1}{n} \log \int_0^1 e^{n\varphi(x)} dx = \sup_{x\in[0,1]} \varphi(x)$$

δηλαδή το θεώρημα του Laplace, όπως το είδαμε παραπάνω. Άρα πράγματι το λήμμα του Varadhan αποτελεί μία φυσική γενίκευση, για την περίπτωση που οι κατανομές ικανοποιούν την LDP.

Αρκετά ενδιαφέρον είναι, τέλος, και το γεγονός ότι ισχύει το αντίστροφο του Λήμματος του Varadhan. Όμως, για να ισχύει, απαιτείται να κάνουμε τον χώρο μας λίγο πιο πλούσιο. Συγκεκριμένα, αρκεί να του δώσουμε την ιδιότητα όλα του τα σημεία να μπορούν να διαχωριστούν από συνεχείς συναρτήσεις. Ένας τέτοιος τοπολογικός χώρος λέγεται πλήρως κανονικός. Το λήμμα του Bryc, που είναι το αντίστροφο του Varadhan, έχει την εξής διατύπωση:

Λήμμα 6 (**Bryc**). *Έστω ότι η οικογένεια $\{\mu_n\}_{n\in\mathbb{N}}$ μέτρων πιθανότητας είναι εκθετικά σφιχτή σε κάποιον πλήρως κανονικό τοπολογικό χώρο Hausdorff $\mathcal{X}$. Υποθέτουμε ότι το όριο:*

$$\Lambda(f) \triangleq \lim_{n\to+\infty} \frac{1}{n} \log \int_{\mathcal{X}} e^{nf(x)} d\mu_n(x)$$

υπάρχει για όλες τις $f \in C_b(\mathcal{X})$. Τότε, η $\{\mu_n\}_{n\in\mathbb{N}}$ ικανοποιεί την LDP με καλή συνάρτηση ρυθμού:

$$I(x) = \sup_{f\in C_b(\mathcal{X})} \{f(x) - \Lambda(f)\}$$

Επιπλέον, για κάθε $f \in C_b(\mathcal{X})$:

$$\Lambda(f) = \sup_{x\in\mathcal{X}} \{f(x) - I(x)\}$$

Συνεπώς, το Λήμμα του Varadhan μας δίνει την ασυμπτωτική συμπεριφορά ολοκληρωμάτων, ως συνέπεια της LDP. Το Λήμμα του Bryc μας λέει ότι η γνώση της ασυμπτωτικής συμπεριφοράς επαρκώς πολλών ολοκληρωμάτων είναι ισοδύναμη με την LDP.

2.2 Τοπολογικοί διανυσματικοί χώροι

Ας περάσουμε τώρα στην μελέτη της LDP σε τοπολογικούς διανυσματικούς χώρους. Συγκεκριμένα, θα χτίσουμε σταδιακά τη θεωρία, ώστε με τη χρήση του Λήμματος του Varadhan να αποδείξουμε ένα γενικό άνω φράγμα για την LDP. Αυτό το αποτέλεσμα είναι βασικό, καθώς θα μας βοηθήσει να αποδείξουμε το θεώρημα του Sanov και συνεπώς την αρχή του Gibbs.

2.2.1 Η απόδειξη ενός γενικού άνω φράγματος για την LDP

Ας ξεκινήσουμε τη μελέτη θεωρώντας δύο πραγματικούς, διανυσματικούς χώρους $\mathcal{X}, \mathcal{Y}$. Έστω ότι υπάρχει μια διγραμμική μορφή:

$$\langle \cdot, \cdot \rangle : \quad \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$$

Παραθέτουμε τον εξής σημαντικό ορισμό:

Ορισμός 6.1. *Ο διανυσματικός χώρος $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ λέγεται τοπολογικός διανυσματικός χώρος, εάν είναι τοπολογικός χώρος με την ιδιότητα ότι οι πράξεις της πρόσθεσης και του βαθμωτού πολλαπλασιασμού (υπό το εσωτερικό γινόμενο $\langle \cdot, \cdot \rangle$) είναι συνεχείς.*

Με τέτοιους χώρους θέλουμε να δουλέψουμε στη συνέχεια. Μάλιστα, ισχύει ότι κάθε κανονικός τοπολογικός διανυσματικός χώρος είναι χώρος Hausdorff. Ας προχωρήσουμε τώρα στον ορισμό της ασθενούς τοπολογίας $\sigma(\mathcal{X}, \mathcal{Y})$, ξεκινώντας από ένα σχόλιο.

Σχόλιο. Μια συνάρτηση $f : \mathcal{X} \to \mathcal{Y}$ λέγεται συνεχής αν και μόνο αν για κάθε $C \subset \mathcal{Y}$ ανοιχτό: $f^{-1}(C) \subset \mathcal{X}$ είναι ανοιχτό.

Συνεπώς, εάν μειώσουμε τα ανοιχτά του $\mathcal{Y}$, η συνέχεια της f δε θα αλλάξει. Εάν όμως μειώσουμε τα ανοιχτά του $\mathcal{X}$, η συνέχεια της f μπορεί να καταρριφθεί, καθώς αν $C \subset \mathcal{Y}$ ανοιχτό, το $f^{-1}(C) \subset \mathcal{X}$ δεν θα είναι κατ'ανάγκη ανοιχτό.

Επομένως, έχει νόημα να ορίσουμε την μικρότερη τοπολογία υπό την οποία οι συναρτήσεις:

$$x \mapsto \langle x, y \rangle, \quad \forall y \in \mathcal{Y}$$

είναι συνεχείς. Την τοπολογία αυτή θα την λέμε ασθενή τοπολογία και θα τη συμβολίζουμε με $\sigma(\mathcal{X}, \mathcal{Y})$.

Παρακάτω, θα θεωρούμε σε ισχύ την εξής υπόθεση:

Υπόθεση. Για κάθε $x \in \mathcal{X} \setminus \{0\}$ υπάρχει $y \in \mathcal{Y}$, ώστε $\langle x, y \rangle \neq 0$.

Υπό την υπόθεση αυτή, η τοπολογία $\sigma(\mathcal{X}, \mathcal{Y})$ είναι Hausdorff, αφού αν $x_1, x_2 \in \mathcal{X}, x_1 \neq x_2$, τότε υπάρχει $y \in \mathcal{Y}$, ώστε:

$$\langle x_1 - x_2, y \rangle \neq 0 \Leftrightarrow \langle x_1, y \rangle \neq \langle x_2, y \rangle \Leftrightarrow f(x_1) \neq f(x_2)$$

όπου $f(x) = \langle x, y \rangle$. Εάν $r = \dfrac{f(x_1) + f(x_2)}{2}$, τότε θεωρώντας χωρίς βλάβη της γενικότητας ότι $f(x_1) < f(x_2)$, μπορούμε να βρούμε περιοχές U, V των $f(x_1), f(x_2)$

αντίστοιχα, ξένες μεταξύ τους. Για παράδειγμα:

$$U = \{h(x) : h(x) < r\}$$
$$V = \{g(x) : g(x) > r\}$$

Άρα αφού $f^{-1}(U) \cap f^{-1}(V) = f^{-1}(U \cap V) = f^{-1}(\emptyset) = \emptyset$ και f συνεχής στην τοπολογία $\sigma(\mathcal{X}, \mathcal{Y})$, τότε τα $f^{-1}(U), f^{-1}(V)$ είναι ξένες, ανοιχτές περιοχές των x_1, x_2 αντίστοιχα, οπότε η $\sigma(\mathcal{X}, \mathcal{Y})$ είναι Hausdorff.

Αφού έχουμε μια τοπολογία, θα έχουμε και μια βάση περιοχών για αυτή. Μια βάση για την $\sigma(\mathcal{X}, \mathcal{Y})$ είναι η συλλογή των συνόλων της μορφής:

$$U_C(x_0) = \bigcap_{y \in C} \{x \in \mathcal{X} : |\langle x, y\rangle - \langle x_0, y\rangle| < \varepsilon\}, \quad \varepsilon > 0, x_0 \in \mathcal{X}, \dim(C) < \infty$$

Ο $\mathcal{X}$ με την ασθενή τοπολογία είναι τοπικά κυρτός, αφού αν $x_0 \in \mathcal{X}$ και $V(x_0)$ είναι μια περιοχή του, τότε για $x_1, x_2 \in X$ και $\varepsilon > 0$, θα ισχύει:

$$|\langle x_1, y\rangle - \langle x_0, y\rangle| < \varepsilon$$
$$|\langle x_2, y\rangle - \langle x_0, y\rangle| < \varepsilon$$

και άρα για να αποδείξουμε την κυρτότητα, θα πάρουμε ένα τυχαίο $\lambda \in [0, 1]$ και θα υπολογίσουμε:

$$\begin{aligned}
|\langle \lambda x_1 + (1-\lambda)x_2, y\rangle - \langle x_0, y\rangle| &= |\lambda\langle x_1, y\rangle - \lambda\langle x_0, y\rangle + (1-\lambda)\langle x_2, y\rangle - (1-\lambda)\langle x_0, y\rangle| \\
&\leq \lambda|\langle x_1, y\rangle - \langle x_0, y\rangle| + (1-\lambda)|\langle x_2, y\rangle - \langle x_0, y\rangle| \\
&< \lambda\varepsilon + (1-\lambda)\varepsilon = \varepsilon
\end{aligned}$$

Άρα, τα $U_C(x_0)$ είναι κυρτά σύνολα, οπότε κάθε περιοχή του x_0 περιέχει μια κυρτή περιοχή του x_0.

Ακόμα, οι συναρτήσεις:

$$(a, x) \mapsto ax, \quad (\mathbb{R} \times \mathcal{X} \to \mathcal{X})$$
$$(x, y) \mapsto x + y, \quad (\mathcal{X} \times \mathcal{X} \to \mathcal{X})$$

είναι προφανώς συνεχείς ως προς την τοπολογία γινόμενο και άρα ο $\mathcal{X}$ είναι τοπολογικός διανυσματικός χώρος.

Τα παραπάνω ευρήματα είναι σημαντικά και αξίζει να τα διατυπώσουμε σε μία πρόταση:

Πρόταση 6.1. *Υπό την ανωτέρω υπόθεση, ο τοπολογικός διανυσματικός χώρος $\mathcal{X}$ με την τοπολογία $\sigma(\mathcal{X}, \mathcal{Y})$ είναι ένας Hausdorff, τοπικά κυρτός, τοπολογικός διανυσματικός χώρος.*

Πριν τη διατύπωση του βασικού θεωρήματος, θα διατυπώσουμε ένα Λήμμα που θα μας βοηθήσει στην μετέπειτα απόδειξη. Το λήμμα αυτό λέει ότι το sup συνεχών συναρτήσεων είναι κάτω ημισυνεχής συνάρτηση. Ο ορισμός της κάτω ημισυνεχούς συνάρτησης σε γενικούς τοπολογικούς χώρους γενικεύεται με φυσικό τρόπο. Κανείς μπορεί να τον βρει για παράδειγμα στο [17] και διατυπώνεται ως ακολούθως:

Ορισμός 6.2. *Εάν (X, σ) είναι ένας Hausdorff τοπολογικός χώρος, τότε η συνάρτηση $f : \mathcal{X} \to \mathbb{R}$ είναι κάτω ημισυνεχής στο $x \in \mathcal{X}$ ως προς την τοπολογία σ, εάν για κάθε $a \in \mathbb{R}$ με $f(x) > a$ υπάρχει ανοιχτή περιοχή U του x, ώστε $f(u) > a$ για κάθε $u \in U$.*

Με βάση αυτόν τον ορισμό μπορούμε τώρα να διατυπώσουμε και να αποδείξουμε το Λήμμα:

Λήμμα 7. *Έστω I ένα διατεταγμένο σύνολο δεικτών και $\{f_i\}_{i\in I}$ μια οικογένεια συνεχών συναρτήσεων ορισμένες σε ένα υποσύνολο $A \subseteq \mathcal{X}$. Τότε, η συνάρτηση $g(x) = \sup_{i\in I} f_i(x)$ είναι κάτω ημισυνεχής.*

Απόδειξη. Έστω $x \in \mathcal{X}$, ώστε $g(x) = \sup_{i\in I} f_i(x) > a$. Τότε, θα υπάρχει κάποιο $i \in I$, ώστε $f_i(x) > a$. Όμως αφού η f_i είναι συνεχής, τότε θα υπάρχει κάποια ανοιχτή περιοχή U του x, ώστε $f_i(y) > a$ για κάθε $y \in U$. Αλλά εξ ορισμού $g(y) \geq f_i(y)$ για όλα τα $y \in U$, οπότε $g(y) > a$ για κάθε $y \in U$. Άρα βρήκαμε μια ανοιχτή περιοχή που να ικανοποιεί τον παραπάνω ορισμό. Συνεπώς, η g είναι κάτω ημισυνεχής. □

Είμαστε λοιπόν σε θέση να διατυπώσουμε το βασικό θεώρημα, που θα μας δώσει το γενικό άνω φράγμα της LDP:

Θεώρημα 8. *Έστω $\mathcal{X}, \mathcal{Y}$ διανυσματικοί χώροι που ικανοποιούν την υπόθεση. Εφοδιάζουμε τον $\mathcal{X}$ με την ασθενή τοπολογία $\sigma(\mathcal{X}, \mathcal{Y})$. Θεωρούμε επίσης ένα κλειστό και κυρτό σύνολο $\mathcal{E} \subset \mathcal{X}$, στο οποίο είναι ορισμένα τα μέτρα πιθανότητας $\{\mu_n\}_{n\in\mathbb{N}}$. Ορίζουμε:*

$$\Lambda(\lambda) \triangleq \limsup_{n\to+\infty} \frac{1}{n} \log \int_{\mathcal{E}} e^{n\langle x,\lambda\rangle} d\mu_n(x), \quad \lambda \in \mathcal{Y}$$

και τον μετασχηματισμό Fenchel-Legendre:

$$\Lambda^*(x) = \sup_{\lambda\in\mathcal{Y}} \left(\langle x, \lambda\rangle - \Lambda(\lambda)\right), \quad x \in \mathcal{X}$$

Τότε:

1. *Η Λ^* είναι συνάρτηση ρυθμού*
2. *Για κάθε $K \subset \mathcal{E}$ συμπαγές:*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mu_n(K) \leq -\inf_{x\in K} \Lambda^*(x)$$

Απόδειξη. Αποδεικνύουμε αρχικά ότι η Λ^* είναι κυρτή συνάρτηση ρυθμού και έπειτα αποδεικνύουμε το άνω φράγμα:

1. Αφού $\Lambda(0) = 0$, τότε $\Lambda^*(x) \geq 0$.
   Επιπλέον, χρησιμοποιώντας το παραπάνω Λήμμα, η Λ^* είναι κάτω ημισυνεχής ως το sup συνεχών συναρτήσεων.

2. Θα αποκτήσουμε αρχικά ένα άνω φράγμα για μικρές περιοχές του x:

   $$A_{\lambda,\varepsilon}(x) = \{y \in \mathcal{E} : |\langle y, \lambda\rangle - \langle x, \lambda\rangle| < \varepsilon\}$$

   Ας παρατηρήσουμε ότι εάν $y \in A_{\lambda,\varepsilon}(x)$, τότε:

   $$\langle y, \lambda\rangle - \langle x, \lambda\rangle > -\varepsilon \Leftrightarrow n\langle y, \lambda\rangle + n\varepsilon > n\langle x, \lambda\rangle, \quad \forall n \in \mathbb{N}$$

   Οπότε:

   $$\begin{aligned} \mu_n(A_{\lambda,\varepsilon}(x)) &= \int_{A_{\lambda,\varepsilon}} d\mu_n(y) \\ &\leq \int_{A_{\lambda,\varepsilon}} e^{n\langle y,\lambda\rangle + n\varepsilon - n\langle x,\lambda\rangle} d\mu_n(y) \\ &= e^{-n\langle x,\lambda\rangle + n\varepsilon} \int_{A_{\lambda,\varepsilon}} e^{n\langle y,\lambda\rangle} d\mu_n(y) \end{aligned}$$

   Άρα:

   $$\limsup_{n\to+\infty} \frac{1}{n} \log \mu_n(A_{\lambda,\varepsilon}(x)) \leq -\langle x, \lambda\rangle + \varepsilon + \Lambda(\lambda) = \varepsilon - (\langle x, \lambda\rangle - \Lambda(\lambda))$$

   Από τον ορισμό του Λ^* έχουμε ότι για κάθε $\varepsilon > 0$ υπάρχει $\lambda = \lambda(x, \varepsilon)$:

   $$\begin{cases} \langle x, \lambda\rangle - \Lambda(\lambda) > \Lambda^* - \varepsilon \\ \langle x, \lambda\rangle - \Lambda(\lambda) > \dfrac{1}{\varepsilon} \end{cases} \Rightarrow \langle x, \lambda\rangle - \Lambda(\lambda) > (\Lambda^*(x) - \varepsilon) \wedge \frac{1}{\varepsilon}$$

   Οπότε:

   $$\limsup_{n\to+\infty} \frac{1}{n} \log \mu_n(A_{\lambda,\varepsilon}(x)) \leq \varepsilon - (\Lambda^*(x) - \varepsilon) \wedge \frac{1}{\varepsilon}$$

   Κατόπιν, καλύπτουμε το συμπαγές K με τέτοιες μικρές περιοχές. Δηλαδή, για $N \in \mathbb{N}$ και $x_i \in K$ για $i = 1, ..., N$ έχουμε: $K \subset \cup_{i=1}^{N} A_{\lambda,\varepsilon}(x_i)$

Και τότε:

$$
\begin{aligned}
\limsup_{n\to+\infty}\frac{1}{n}\log\mu_n(K) &\le \limsup_{n\to+\infty}\frac{1}{n}\log\mu_n(\cup_{i=1}^N A_{\lambda,\varepsilon}(x_i)) \\
&\le \limsup_{n\to+\infty}\frac{1}{n}\log\left(\sum_{i=1}^N \mu_n(A_{\lambda,\varepsilon}(x_i))\right) \\
&\le \limsup_{n\to+\infty}\frac{1}{n}\log\left(\max_{i=1,\dots,N}\mu_n(A_{\lambda,\varepsilon}(x_i))\right) \\
&= \max_{i=1,\dots,N}\left(\limsup_{n\to+\infty}\frac{1}{n}\log\mu_n(A_{\lambda,\varepsilon}(x_i))\right) \\
&\le \max_{i=1,\dots,N}\left(\varepsilon-(\Lambda^*(x_i)-\varepsilon)\wedge\frac{1}{\varepsilon}\right) \\
&\le \varepsilon-\inf_{x\in K}\left(\varepsilon-(\Lambda^*(x_i)-\varepsilon)\wedge\frac{1}{\varepsilon}\right)
\end{aligned}
$$

Οπότε στέλνοντας το $\varepsilon\to 0^+$ παίρνουμε:

$$
\limsup_{n\to+\infty}\frac{1}{n}\log\mu_n(K)\le -\inf_{x\in K}\Lambda^*(x)
$$

□

Ας παρατηρήσουμε ότι το άνω φράγμα αρκεί να το εκτιμήσουμε στα συμπαγή, καθώς η εκθετική σφιχτότητα θα μας δώσει εν τέλει την LDP.

Βρήκαμε άρα ένα άνω φράγμα για την ασθενή LDP και μια υποψήφια συνάρτηση ρυθμού:

$$
\Lambda^*(x)=\sup_{\lambda\in\mathcal{Y}}\{\langle x,\lambda\rangle-\Lambda(\lambda)\}
$$

.

Ένα φυσιολογικό και σημαντικό ερώτημα προκύπτει: Πότε είναι ικανοποιητικό αυτό το άνω φράγμα; Δηλαδή, μας δίνει μια σωστή συνάρτηση ρυθμού, αν ικανοποιείται η LDP;

Η απάντηση είναι καταφατική. Πριν το αποδείξουμε όμως, ας κάνουμε την εξής παρατήρηση:

Παρατήρηση. Η $\sigma(\mathcal{X},\mathcal{Y})$ είναι κανονική, δηλαδή αν $A\subset\mathcal{X}$ κλειστό και $x\notin A$, τότε υπάρχουν U,V ανοιχτά, με $U\cap V=\emptyset$, ώστε $x\in U$ και $A\subset V$. Αυτό προκύπτει από το μονοπάτι που διασχίσαμε για την απόδειξη του Λήμματος του Varadhan, όπου παρατηρήσαμε πως κάθε τοπολογικός διανυσματικός χώρος είναι κανονικός.
Ακόμα, θα διατυπώσουμε ένα σημαντικό θεώρημα, το Fenchel-Moreau, το οποίο θα μας βοηθήσει στη συνέχεια:

Θεώρημα 9 (Fenchel-Moreau). *Έστω $\mathcal{X}$ ένας τοπολογικός διανυσματικός χώρος, $d : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$. Θα συμβολίζουμε με X^* τον δυϊκό χώρο του $\mathcal{X}$. Ορίζουμε επίσης τον μετασχηματισμό Fenchel-Legendre της f ως το συναρτησιακό $f^* : \mathcal{X}^* \to R \cup \{\pm\infty\}$:*

$$f^*(x^*) = \sup_{x \in \mathcal{X}} \{x^*(x) - f(x)\}$$

Τότε, εάν η f είναι κάτω ημισυνεχής και κυρτή και $f(x) > -\infty$, ισχύει ότι:

$$f^{**} = f$$

όπου,

$$f^{**} = \sup_{x^* \in \mathcal{X}^*} \{x^* - f^*(x^*)\}$$

Δηλαδή ο μετασχηματισμός Fenchel-Legendre είναι ισομορφισμός.

Την απόδειξη του θεωρήματος αυτού την παραλείπουμε, καθώς είναι τεχνική και ξεφεύγει από τους σκοπούς της εργασίας. Τώρα, είμαστε σε θέση να διατυπώσουμε και να αποδείξουμε το εξής σημαντικό θεώρημα:

Θεώρημα 10. *Έστω ότι για τα μέτρα πιθανότητας $\{\mu_n\}_{n\in\mathbb{N}}$ ικανοποιείται η LDP με κυρτή συνάρτηση ρυθμού I σε κάποιο κλειστό, κυρτό σύνολο $\mathcal{E}$. Τότε:*

1. *Η LDP για τα $\{\mu_n\}_{n\in\mathbb{N}}$ ισχύει σε όλον τον $\mathcal{X}$ με κυρτή συνάρτηση ρυθμού την $\tilde{I}$:*

$$\tilde{I}(x) = \begin{cases} I(x), & x \in \mathcal{E} \\ \infty, & x \notin \mathcal{E} \end{cases}$$

2. *Αν επιπλέον $\Lambda(\lambda) < +\infty$ για κάθε $\lambda \in \mathcal{Y}$, τότε:*

$$\Lambda^*(x) = I(x), \quad \forall x \in \mathcal{E}$$

Απόδειξη. 1. Η $\tilde{I}$ είναι κυρτή, αφού για $\lambda \in [0,1]$ και $x, y \in \mathcal{X}$:

- Αν $x \in \mathcal{E}^c$ ή $y \in \mathcal{E}^c$, τότε $\tilde{I}(\lambda x + (1-\lambda)y) \leq \lambda\tilde{I}(x) + (1-\lambda)\tilde{I}(y)$, που ισχύει τετριμμένα
- Αν $x, y \in \mathcal{E}$, τότε αφού $\mathcal{E}$ κυρτό: $\lambda x + (1-\lambda)y \in \mathcal{E}$ και άρα:

  $$\tilde{I}(\lambda x+(1-\lambda)y) = I(\lambda x+(1-\lambda)y) \leq \lambda I(x)+(1-\lambda)I(y) = \lambda\tilde{I}(x)+(1-\lambda)\tilde{I}(y)$$

  λόγω της κυρτότητας της I στο $\mathcal{E}$.

2. Εάν $\Lambda(\lambda) < +\infty$, τότε από το Λήμμα του Varadhan για κάθε $\lambda \in \mathcal{Y}$ έχουμε:

$$\begin{aligned}\Lambda(\lambda) &= \limsup_{n\to+\infty} \frac{1}{n} \log \int_{\mathcal{E}} e^{n\langle x,\lambda\rangle} d\mu_n(x) \\ &= \lim_{n\to+\infty} \frac{1}{n} \log \int_{\mathcal{E}} e^{n\langle x,\lambda\rangle} d\mu_n(x) \\ &= \sup_{x\in\mathcal{E}} \{\langle x, \lambda\rangle - I(x)\}\end{aligned}$$

Άρα, $\Lambda(x) = I^*(x)$ για κάθε $x \in \mathcal{E}$. Αλλά αφού η I είναι κάτω ημισυνεχής, κυρτή και όχι ταυτοτικά ίση με $-\infty$ στο $\mathcal{E}$, από το θεώρημα Fenchel-Moreau, έχουμε ότι: $I^{**}(x) = \Lambda^*(x)$.

Αλλά αφού $\Lambda^*(x) = I^{**}(x)$, τότε $I(x) = \Lambda^*(x)$, για κάθε $x \in \mathcal{E}$. Δηλαδή, για κάθε $x \in \mathcal{E}$:

$$I(x) = \sup_{y \in \mathcal{Y}}\{\langle \lambda, x\rangle - \Lambda(\lambda)\}$$

□

Οπότε όπως βλέπουμε, προκύπτει όντως η σωστή συνάρτηση ρυθμού, αν τα μέτρα πιθανότητας $\{\mu_n\}_{n\in\mathbb{N}}$ ικανοποιούν την LDP.

2.3 Το θεώρημα του Sanov

Στο κεφάλαιο αυτό θα ορίσουμε δύο σημαντικές έννοιες στις μεγάλες αποκλίσεις: την εμπειρική κατανομή και τη σχετική εντροπία. Θα δούμε πως συνδέονται και θα δείξουμε επίσης ότι η εμπειρική κατανομή ικανοποιεί την LDP με συνάρτηση ρυθμού τη σχετική εντροπία. Αυτό ακριβώς μας λέει το θεώρημα του Sanov. Για να το κάνουμε αυτό όμως, θα πρέπει πρώτα να ενισχύσουμε τις γνώσεις μας σχετικά με την ασθενή σύγκλιση μέτρων πιθανότητας. Στη συνέχεια θα χρησιμοποιήσουμε γνώσεις απ' το προηγούμενο κεφάλαιο για να αποδείξουμε το σημαντικό αυτό θεώρημα.

2.3.1 Ασθενής σύγκλιση μέτρων πιθανότητας

Ένα βασικό εργαλείο που θα χρειαστούμε είναι αυτό της ασθενούς σύγκλισης μέτρων πιθανότητας. Το ζήτημα αυτό είναι εξαιρετικά εκτενές και μια λεπτομερής έκθεση γίνεται στα [16] και [1]. Εδώ θα παρουσιάσουμε τα βασικά στοιχεία αυτής της σύγκλισης, που θα μας χρησιμέψουν παρακάτω.

Έστω $\mathcal{S}$ ένας πολωνικός χώρος, δηλαδή ένας πλήρης, διαχωρίσιμος μετρικός χώρος, με μετρική d. Θα συμβολίζουμε με $C_b(\mathcal{S})$ τον χώρο των συνεχών και φραγμένων συναρτήσεων στον $\mathcal{S}$. Θεωρούμε τον χώρο $\mathcal{M}(\mathcal{S})$ όλων των προσημασμένων και πεπερασμένων μέτρων Borel στον $\mathcal{S}$. Ορίζουμε τη διγραμμική απεικόνιση:

$$\langle \cdot, \cdot \rangle : C_b(\mathcal{S}) \times \mathcal{M}(\mathcal{S}) \to \mathbb{R}$$
$$\text{με} \quad \langle f, \mu\rangle = \int_{\mathcal{S}} f d\mu$$

Όπως και στο προηγούμενο κεφάλαιο, θα εφοδιάσουμε τον χώρο $\mathcal{M}(\mathcal{S})$ με την ασθενή τοπολογία $\sigma(\mathcal{M}(\mathcal{S}), C_b(\mathcal{S}))$, δηλαδή τη μικρότερη τοπολογία που κάνει συνεχείς τις απεικονίσεις:

$$\mathcal{M}(\mathcal{S}) \ni \mu \mapsto \int_{\mathcal{S}} f d\mu, \quad \forall f \in C_b(\mathcal{S})$$

Με την ίδια λογική, μια βάση περιοχών για την τοπολογία αυτή είναι εκείνη που παράγεται από σύνολα της μορφής:

$$U_C(\mu_0) = \bigcap_{f \in C} \{\mu \in \mathcal{M}(\mathcal{S}) : \left| \int_{\mathcal{S}} f d\mu - \int_{\mathcal{S}} f d\mu_0 \right| < \varepsilon\}, \quad \varepsilon > 0, \mu_0 \in \mathcal{M}(\mathcal{S}), \dim(C) < \infty$$

Ακολουθώντας το ίδιο ακριβώς σκεπτικό με πριν, ο $\mathcal{M}(\mathcal{S})$ είναι ένας διανυσματικός, τοπικά κυρτός τοπολογικός χώρος Hausdorff. Το υποσύνολό του $\mathcal{M}_1(\mathcal{S})$ των Borel μέτρων πιθανότητας είναι αυτό που μας ενδιαφέρει. Το σύνολο αυτό το εφοδιάζουμε με την επαγόμενη τοπολογία του $\mathcal{M}(\mathcal{S})$. Το παρακάτω θεώρημα μας εξασφαλίζει ότι μπορούμε να μιλήσουμε για σύγκλιση μέτρων πιθανότητας στον χώρο αυτό:

Θεώρημα 11. *Αν $\mathcal{S}$ είναι πολωνικός χώρος με πλήρη μετρική ρ, τότε υπάρχει πλήρης μετρική, ώστε η ασθενής τοπολογία του $\mathcal{M}_1(\mathcal{S})$ που παράγεται από την μετρική αυτή να κάνει κάνει τον $\mathcal{M}_1(\mathcal{S})$ πολωνικό χώρο. Μία τέτοια μετρική είναι η μετρική του Lévy:*

$$d(\mu, \nu) = \inf\{\delta > 0 : \mu(F) \leq \nu(F^\delta) + \delta \ \ \textit{και} \ \ \nu(F) \leq \mu(F^\delta) + \delta, \ \ \forall F \subset \mathcal{S} \ \ \textit{closed}\}$$

όπου $F^\delta = \{x \in \mathcal{S} : \rho(x, F) \leq \delta\}$.

Επομένως, σε ό,τι ακολουθήσει στα επόμενα κεφάλαια μπορούμε να είμαστε σίγουροι ότι ο χώρος $\mathcal{M}_1(\mathcal{S})$ των μέτρων πιθανότητας είναι πολωνικός, αν ο $\mathcal{S}$ είναι πολωνικός. Η φύση της απόδειξης του θεωρήματος αυτού ξεφεύγει από τα πλαίσια και τους σκοπούς της παρούσας εργασίας και γι' αυτό δε θα την παρουσιάσουμε. Όμως, μία σαφή και κατανοητή απόδειξη μπορεί κανείς να βρει στο [8]. Η σύγκλιση μέτρων πιθανότητας ως προς την ασθενή τοπολογία ονομάζεται *ασθενής σύγκλιση μέτρων πιθανότητας*. Μάλιστα, στην απόδειξη του (11) κανείς αποδεικνύει ακριβώς ότι: $d(\mu_n, \mu) \to 0 \Leftrightarrow \mu_n \to \mu$ ασθενώς, καθώς $n \to +\infty$, όπου d η μετρική του Lévy. Πιο συγκεκριμένα, η ασθενής σύγκλιση μέτρων πιθανότητας ορίζεται με φυσιολογικό τρόπο ως εξής:

Ορισμός 11.1. *Μια ακολουθία Borel μέτρων πιθανότητας $\mu_n \in \mathcal{M}_1(\mathcal{S})$ θα συγκλίνει ασθενώς στο Borel μέτρο πιθανότητας $\mu \in \mathcal{M}_1(\mathcal{S})$ εάν για κάθε $f \in C_b(\mathcal{S})$:*

$$\lim_{n \to +\infty} \int_{\mathcal{S}} f d\mu_n = \int_{\mathcal{S}} f d\mu \tag{2.4}$$

Στο πνεύμα της συναρτησιακής ανάλυσης (βλέπε για παράδειγμα [17]):

$$\mu_n \Rightarrow \mu \Leftrightarrow \lim_{n \to +\infty} \langle f, \mu_n \rangle = \langle f, \mu \rangle, \quad \forall f \in C_b(\mathcal{S})$$

Φυσικά, δεν είναι πάντοτε εύκολο να ελέγχουμε την παραπάνω συνθήκη. Σε αυτό όμως μας βοηθάει το διάσημο θεώρημα Portmanteau, που δίνει ισοδύναμες συνθήκες για την ασθενή σύγκλιση μέτρων πιθανότητας [16].

Θεώρημα 12 (Portmanteau). *Έστω μ_n, μ μέτρα πιθανότητας στον μετρικό χώρο $(\mathcal{S}, d)$. Τα παρακάτω είναι ισοδύναμα:*

1. $\mu_n \Rightarrow \mu$

2. *Η συνθήκη (2.4) ισχύει για Lipschitz συνεχείς συναρτήσεις με τιμές στο* $[0,1]$.

3. *Για κάθε* $F \subset \mathcal{S}$ *κλειστό:* $\limsup_{n\to+\infty} \mu_n(F) \le \mu(F)$

4. *Για κάθε* $G \subset \mathcal{S}$ *ανοιχτό:* $\liminf_{n\to+\infty} \mu_n(G) \ge \mu(G)$

5. *Για κάθε* $B \in \mathcal{B}(\mathcal{S})$ *με* $\mu(\partial B) = 0$*:* $\lim_{n\to+\infty} \mu_n(B) = \mu(B)$

Απόδειξη. Θα αποδείξουμε το θεώρημα με τη σειρά: 1. $\Rightarrow$ 2. $\Rightarrow$ 3. $\Rightarrow$ 4. $\Rightarrow$ 5. $\Rightarrow$ 1.

- 1. $\Rightarrow$ 2. :

  Η κατεύθυνση αυτή είναι προφανής, αφού αν η (2.4) ισχύει για κάθε $f \in C_b(\mathcal{S})$, τότε θα ισχύει και για τις f εκεινες που είναι Lipschitz συνεχείς με τιμές στο $[0,1]$, αφού κι αυτές είναι συνεχείς και φραγμένες.

- 2. $\Rightarrow$ 3.

  Έστω ότι ισχύει η 2. Για $F \subset \mathcal{S}$ κλειστό και $x \in \mathcal{S}$ θέτουμε: $d(x,F) \triangleq \inf_{q\in F} d(x,q)$. Τότε προφανώς $|d(x,F) - d(y,F)| \le d(x,y)$ και άρα για $k \ge 1$ οι συναρτήσεις:

  $$f_k(x) = \frac{1}{1+kd(x,F)}$$

  είναι Lipschitz με σταθερά k, αφού:

  $$\begin{aligned}|f_k(x) - f_k(y)| &= \left|\frac{1}{1+kd(x,F)} - \frac{1}{1+kd(y,F)}\right| \\ &= k\left|\frac{d(y,F)-d(x,F)}{(1+kd(x,F))(1+kd(y,F))}\right| \\ &\le kd(x,y)\end{aligned}$$

  Αφού $d(x,F) \ge 0$ και άρα $1 + kd(x,F) \ge 1$ για όλα τα $x \in \mathcal{S}$. Επίσης, ισχύει ότι:

  $$\lim_{k\to\infty} f_k(x) = \begin{cases} 1, & x \in F \\ 0, & x \notin F \end{cases} = \mathbb{1}_{x\in F}(x)$$

  Άρα, οι συναρτήσεις f_k ικανοποιούν την 2., οπότε:

  $$\begin{aligned}\limsup_{n\to+\infty} \mu_n(F) &= \limsup_{n\to+\infty} \int_{\mathcal{S}} \mathbb{1}_{x\in F}(x) d\mu_n(x) \\ &\le \limsup_{n\to+\infty} \int_{\mathcal{S}} f_k(x) d\mu_n(x) \\ &= \lim_{n\to+\infty} \int_{\mathcal{S}} f_k(x) d\mu_n(x) \to \int_{\mathcal{S}} f_k(x) d\mu(x)\end{aligned}$$

Όμως οι f_k είναι φθίνουσες συναρτήσεις που συγκλίνουν κατά σημείο στην $\mathbb{1}_{x\in F}(x)$ και $\int_{\mathcal{S}} f_1 d\mu < +\infty$. Άρα μπορούμε να εφαρμόσουμε μια παραλλαγή του θεωρήματος μονότονης σύγκλισης, για να πάρουμε:

$$\lim_{k\to+\infty}\int_{\mathcal{S}} f_k(x)d\mu(x) = \int_{\mathcal{S}} \mathbb{1}_{x\in F}(x)d\mu = \mu(F)$$

Και άρα συνολικά:

$$\limsup_{n\to+\infty}\mu_n(F) \leq \mu(F)$$

Δηλαδή πήραμε την 3.

- 3. $\Rightarrow$ 4.

  Αυτή η κατεύθυνση είναι αρκετά απλή, αφού αν $F \subset \mathcal{S}$ ανοιχτό, τότε:

  $$\liminf_{n\to+\infty}\mu_n(G) = 1 - \limsup_{n\to+\infty}\mu_n(G^c) \geq 1 - \mu(G^c) = \mu(G)$$

  Αφού το G^c είναι κλειστό και εφαρμόσαμε την 3.

- 4. $\Rightarrow$ 5.

  Οι 3. και 4. μαζί μας δίνουν ότι για οποιοδήποτε $B \in \mathcal{B}(\mathcal{S})$ με $\mu(\partial B) = 0$ έχουμε:

  $$\begin{aligned}\limsup_{n\to+\infty}\mu_n(B) &\leq \limsup_{n\to+\infty}\mu_n(\bar{B})\\ &\leq \mu(\bar{B})\\ &\leq \mu(B^o) + \mu(\partial B)\\ &= \mu(B^o)\\ &\leq \liminf_{n\to+\infty}\mu_n(B^o)\\ &\leq \liminf_{n\to+\infty}\mu_n(B)\end{aligned}$$

  Και άρα $\lim_{n\to+\infty}\mu_n(B) = \mu(B)$, δηλαδή αποδείξαμε την 5.

- 5. $\Rightarrow$ 1.

  Αφού το ολοκλήρωμα είναι γραμμικό, τότε αρκεί να δείξουμε την (2.4) για συναρτήσεις με τιμές στο $[0,1]$. Έστω μια τέτοια $f \in C_b(\mathcal{S})$. Τότε, λόγω συνέχειας της f για κάθε $t \geq 0$:

  $$\partial\{f > t\} \subset \{f = t\} \tag{2.5}$$

  Για να το δεί κανείς αυτό, αρκεί να διαλέξει ένα $a \in \partial A = \partial\{x \in \mathcal{S} : f(x) > t\}$. Τότε, $f(a) \leq t$, αφού το $\{x \in \mathcal{S} : f(x) > t\}$ είναι ανοιχτό. Λόγω συνέχειας της f για κάθε $\varepsilon > 0$ υπάρχει $\delta > 0$, ώστε για κάθε $x \in B(a,\delta)$ έχουμε ότι:

$f(x) - t < \varepsilon$. Από την άλλη όμως από την ιδιότητα του συνόρου ενός συνόλου θα υπάρχει και κάποιο $y \in B(a, \delta) \cap A$. Οπότε, $0 < f(y) - t < \varepsilon$. Άρα, μπορούμε να βρούμε μια ακολουθία $\{x_n\}_{n\in\mathbb{N}} \subset A$, ώστε $x_n \to a$ και $f(x_n) - t = \dfrac{1}{n}$.

Οπότε περνώντας στο όριο και λόγω συνέχειας της f παίρνουμε: $f(a) = t$ και άρα ότι $a \in \{x \in \mathcal{S} : f(x) = t\}$, δηλαδή ικανοποιείται η (2.5).

Επιπλέον, είναι εύκολο να δει κανείς ότι τα σύνολα $\{f = t\}$ για κάθε $t \geq 0$ είναι ξένα μεταξύ τους. Οπότε, το πλήθος των t για τα οποία $\mu(\{f = t\}) > 0$ είναι αριθμήσιμο, δηλαδή έχει μέτρο 0. Αυτό είναι συνέπεια ενός γενικότερου αποτελέσματος στη θεωρία μέτρου, που μας λέει ότι αν έχουμε μια οικογένεια ξένων συνόλων θετικού μέτρου, τότε αυτή η οικογένεια είναι το πολύ αριθμήσιμη σε μέγεθος. Άρα:

$$\mu(\partial\{f > t\}) \leq \mu(\{f = t\}) = 0 \Rightarrow \mu(\partial\{f > t\}) = 0 \tag{2.6}$$

εκτός από αριθμήσιμα πολλά t. Οπότε κάνουμε τον υπολογισμό:

$$\begin{aligned}
\mathbb{E}^{\mu_n}[f] &= \int_0^{+\infty} \mu_n(\{f > t\})dt \\
&= \int_0^{1} \mu_n(\{f > t\})dt
\end{aligned}$$

Όμως, από το θεώρημα κυριαρχημένης σύγκλισης, αφού οι συναρτήσεις $\mu_n(\cdot)$ είναι φραγμένες στο $[0, 1]$:

$$\begin{aligned}
\lim_{n\to+\infty} \int_0^1 \mu_n(\{f > t\})dt &= \int_0^1 \lim_{n\to+\infty} \mu_n(\{f > t\})dt \\
&= \int_0^1 \mu(\{f > t\})dt
\end{aligned}$$

όπου η τελευταία ισότητα προέκυψε από την (2.6) και την υπόθεση 5. Και άρα:

$$\lim_{n\to+\infty} \mathbb{E}^{\mu_n}[f] = \int_0^1 \mu(\{f > t\})dt = \int_0^{+\infty} \mu(\{f > t\})dt = \mathbb{E}^{\mu}[f]$$

δηλαδή παίρνουμε την 1.

Η απόδειξη του θεωρήματος Portmanteau έχει ολοκληρωθεί. □

Παρατήρηση. Έστω μια ακολουθία τυχαίων μεταβλητών $\{X_n\}_{n\in\mathbb{N}}$ στον πολωνικό χώρο $\mathcal{S}$ με αντίστοιχες κατανομές $\{\mu_n\}_{n\in\mathbb{N}}$, δηλαδή: $\mu_n = \mathbb{P} \circ X_n^{-1}$. Έστω και η τυχαία μεταβλητή X, με κατανομή $\mu = \mathbb{P} \circ X^{-1}$, στον ίδιο χώρο $\mathcal{S}$. Αν οι τυχαίες μεταβλητές X_n συγκλίνουν κατά κατανομή στην X αυτό δε σημαίνει τίποτε άλλο παρά ότι οι κατανομές μ_n των X_n συγκλίνουν ως προς την ασθενή τοπολογία του $\mathcal{M}_1(\mathcal{S})$ (του συνόλου των μέτρων πιθανότητας) στην κατανομή της X, δηλαδή $\mu_n \Rightarrow \mu$.

Ολοκληρώνουμε το κεφάλαιο αυτό με την απόδειξη δύο σημαντικών ιδιοτήτων του υποσυνόλου $\mathcal{M}_1(\mathcal{S})$ του $\mathcal{M}(\mathcal{S})$:

Πρόταση 12.1. *Το $\mathcal{M}_1(\mathcal{S})$ είναι κλειστό και κυρτό υποσύνολο του $\mathcal{M}(\mathcal{S})$.*

Απόδειξη. Η κυρτότητα είναι απλή και δεν απαιτεί την ασθενή σύγκλιση. Έστω μ_1, μ_2 δύο μέτρα πιθανότητας στον $\mathcal{S}$ και $a \in [0,1]$. Έστω το μέτρο $\mu = a\mu_1 + (1-a)\mu_2$. Τότε:

1. $\mu(\emptyset) = a\mu_1(\emptyset) + (1-a)\mu_2(\emptyset) = a \cdot 0 + (1-a)\cdot 0 = 0$
2. $\forall A \in \mathcal{B}(\mathcal{S}) : \mu(A) = a\mu_1(A) + (1-a)\mu_2(A) \geq 0$
3. $\mu(\mathcal{S}) = a\mu_1(\mathcal{S}) + (1-a)\mu_2(\mathcal{S}) = a\cdot 1 + (1-a)\cdot 1 = 1$
4. Αν $A_1, ..., A_n \in \mathcal{B}(\mathcal{S})$ ξένα ανά δύο, τότε θέτω $A = \bigcup_{i\geq 1} A_i$ και άρα:

$$\mu(A) = a\mu_1(A) + (1-a)\mu_2(A) = \sum_{i=1}^{+\infty}(a\mu_1(A_i) + (1-a)\mu_2(A_i)) = \sum_{i=1}^{+\infty}\mu(A_i)$$

Για την κλειστότητα, θεωρούμε μια ακολουθία $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathcal{M}_1(\mathcal{S})$, η οποία συγκλίνει ασθενώς σε κάποιο μέτρο $\mu \in \mathcal{M}(\mathcal{S})$. Αρκεί να αποδείξουμε ότι $\mu(\mathcal{S}) = 1$. Εξ ορισμού της ασθενούς σύγκλισης έχουμε ότι:

$$\lim_{n\to+\infty}\int_{\mathcal{S}} f d\mu_n = \int_{\mathcal{S}} f d\mu, \quad \forall f \in C_b(\mathcal{S})$$

Διαλέγουμε $f(x) = 1$. Η f είναι συνεχής και φραγμένη, δηλαδή ικανοποιεί την παραπάνω σχέση, οπότε:

$$\mu(S) = \int_{\mathcal{S}} 1\cdot d\mu = \lim_{n\to+\infty}\int_{\mathcal{S}} 1\cdot d\mu_n = \lim_{n\to+\infty}\mu_n(\mathcal{S}) = \lim_{n\to+\infty} 1 = 1$$

Άρα $\mu \in \mathcal{M}_1(\mathcal{S})$, οπότε το $\mathcal{M}_1(\mathcal{S})$ είναι κλειστό. □

2.3.2 Εμπειρική κατανομή

Έστω $\mathcal{S}$ ένας πολωνικός χώρος. Το μέτρο Dirac δ_x στο $x \in \mathcal{S}$ ορίζεται ως:

$$\delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \quad , \quad A \in \mathcal{B}(\mathcal{S}) \tag{2.7}$$

Έστω $\{X_i\}_{n\in\mathbb{N}}$ ανεξάρτητες, ισόνομες τυχαίες μεταβλητές ορισμένες σε κάποιον χώρο πιθανότητας $(\Omega, \mathcal{F}, \mathbb{P})$ με τιμές στον $\mathcal{S}$ και με κατανομή $\mu \in \mathcal{M}_1(\mathcal{S})$. Τότε ορίζουμε την *εμπειρική κατανομή* των $\{X_i\}_{1\leq i\leq n}$ ως την εξής απεικόνιση:

$$L_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} \in \mathcal{M}_1(\mathcal{S}) \tag{2.8}$$

όπου δ_{X_i} είναι ο μέτρο Dirac στον $\mathcal{S}$, όπως ορίστηκε στην (2.7). Η L_n είναι ένα τυχαίο μέτρο πιθανότητας που για κάθε $\omega \in \Omega$ ορίζεται ως εξής:

$$\mathcal{B}(\mathcal{S}) \ni A \mapsto L_n(\omega)(A) = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(\omega)}(A) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_A(X_i(\omega))$$

Δίνει δηλαδή σε κάθε $A \in \mathcal{B}(\mathcal{S})$ βάρος ίσο με το πηλίκο των $\{X_i\}_{1\leq i\leq n}$ που ανήκουν στο A. Την ιδιότητα του μέτρου πιθανότητας δεν είναι δύσκολη να την αποδείξει κανείς, αφού χρησιμοποιεί απλώς το μέτρο Dirac. Εδώ θα δείξουμε ότι η L_n είναι τυχαία μεταβλητή, εάν εφοδιάσουμε τον $\mathcal{M}_1(\mathcal{S})$ με την σ-άλγεβρα $\mathcal{B}(\mathcal{M}_1(\mathcal{S}))$.

Έτσι, λοιπόν, εμείς θα θέλαμε να δείξουμε ότι:

$$\forall B \in \mathcal{B}(\mathcal{M}_1(\mathcal{S})) : \{\omega \in \Omega : L_n(\omega) \in B\} \in \mathcal{F}$$

Αφού όμως για κάθε $f \in C_b(\mathcal{S}), \varepsilon > 0, x \in \mathbb{R}$ τα σύνολα $\{\mu \in \mathcal{M}_1(\mathcal{S}) : |\int_{\mathcal{S}} f d\mu - x| < \varepsilon\}$ παράγουν τη σ-άλγεβρα $\mathcal{B}(\mathcal{M}_1(\mathcal{S}))$, τότε αρκεί να αποδείξουμε ότι:

$$\forall f \in C_b(\mathcal{S}), \varepsilon > 0, x \in \mathbb{R} : \{\omega \in \Omega : \left|\int_{\mathcal{S}} f dL_n - x\right| < \varepsilon\} \in \mathcal{F}$$

Όμως χρησιμοποιώντας την ιδιότητα του μέτρου Dirac παίρνουμε ότι:

$$\int_{\mathcal{S}} f dL_n = \int_{\mathcal{S}} f d\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}\right) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{S}} f d\delta_{X_i} = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$$

και άρα αφού οι X_i είναι τυχαίες μεταβλητές και η f συνεχής, θα έχουμε:

$$\forall f \in C_b(\mathcal{S}), \varepsilon > 0, x \in \mathbb{R} : \{\omega \in \Omega : \left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - x\right| < \varepsilon\} \in \mathcal{F}$$

οπότε η $L_n : \Omega \to \mathcal{M}_1(\mathcal{S})$ είναι μια τυχαία μεταβλητή. Άρα, δικαιωματικά τελικά, η L_n είναι ένα τυχαίο μέτρο πιθανότητας.

Αφού εξετάσαμε τη φύση της εμπειρικής κατανομής, θα θέλαμε να εξετάσουμε και την οριακή της συμπεριφορά. Συγκεκριμένα, που συγκλίνει ασθενώς ως μέτρο πιθανότητας; Για να το απαντήσουμε αυτό, θα χρειαστεί να ορίσουμε μια βοηθητική έννοια:

Ορισμός 12.1. *Μια ακολουθία συναρτήσεων $\{g_k\}_{k\in\mathbb{N}} \subseteq C_b(\mathcal{S})$ θα λέμε ότι καθορίζει την ασθενή σύγκλιση στον $\mathcal{M}_1(\mathcal{S})$, εάν η ασθενής σύγκλιση $\mu_n \Rightarrow \mu$ είναι ισοδύναμη με την:*

$$\int_{\mathcal{S}} g_k d\mu_n = \int_{\mathcal{S}} g_k d\mu, \quad k = 1, 2, ...$$

Όταν ο $\mathcal{S}$ είναι πολωνικός χώρος, υπάρχει πάντοτε μια ακολουθία συναρτήσεων που να καθορίζει την ασθενή σύγκλιση (βλέπε για παράδειγμα [20]). Στο ερώτημα που μας απασχολεί την απάντηση μας τη δίνει το ακόλουθο σημαντικό θεώρημα (στην ουσία, ένας νόμος μεγάλων αριθμών):

Θεώρημα 13. *Έστω $\{X_n\}_{n\in\mathbb{N}}$ μια ακολουθία ανεξάρτητων, ισόνομων τυχαίων μεταβλητών με κοινή κατανομή μ. Τότε:*

$$\mathbb{P}[\{\omega \in \Omega : L_n(\omega) \Rightarrow \mu\}] = 1$$

Δηλαδή:

$$L_n \Rightarrow \mu, \quad \mathbb{P} - \sigma.\beta.$$

Απόδειξη. Έχουμε ότι:

$$\lim_{n\to+\infty}\int_{\mathcal{S}} f dL_n = \lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n} f(X_i) = \mathbb{E}^{\mu}[f(X_i)] = \int_{\mathcal{S}} f d\mu, \quad \mathbb{P}-\sigma.\beta. \tag{2.9}$$

Η παραπάνω σύγκλιση είναι σύγκλιση τυχαίων μεταβλητών και ισχύει χάρη στον ισχυρό νόμο των μεγάλων αριθμών, αφού η f είναι φραγμένη.

Επειδή ο $\mathcal{S}$ είναι πολωνικός, από την προηγούμενη συζήτηση θα υπάρχει μια ακολουθία συναρτήσεων $\{f_k\} \subseteq C_b(\mathcal{S})$ που θα καθορίζει την ασθενή σύγκλιση, δηλαδή:

$$\int_{\mathcal{S}} f_k d\mu_n = \int_{\mathcal{S}} f_k d\mu, \quad k = 1, 2, ...$$

Και αφού η σχέση (2.9) ισχύει για όλες τις $f \in C_b(\mathcal{S})$, τότε θα ισχύει και για όλες τις f_k. Αλλά αυτό εξ ορισμού των f_k είναι ισοδύναμο με τον ισχυρισμό του Θεωρήματος. □

Το παραπάνω θεώρημα είναι σημαντικό, καθώς αποτελεί ένα ανάλογο του Νόμου των Μεγάλων Αριθμών, δηλαδή μιας τυπικής συμπεριφοράς. Όπως πολύ εύστοχα επισημαίνουν και οι συγγραφείς στο [20], το θεώρημα του Sanov, του οποίου η απόδειξη σταδιακά καταφθάνει, μας λέει με ποιον τρόπο πραγματοποιείται η απόκλιση από την τυπική συμπεριφορά του παραπάνω θεωρήματος.

2.3.3 Σχετική Εντροπία

Η επόμενη σημαντική έννοια που θα μας απασχολήσει είναι η σχετική εντροπία. Η σχετική εντροπία ενός μέτρου $\nu \in \mathcal{M}_1(\mathcal{S})$ (μιας κατανομής) ως προς το μέτρο $\mu \in \mathcal{M}_1(\mathcal{S})$ (κατανομή) ορίζεται ως:

$$H(\nu|\mu) = \begin{cases} \int_{\mathcal{S}} f \log f d\mu, & \nu \ll \mu \text{ και } f = \dfrac{d\nu}{d\mu} \\ \infty, & \nu \not\ll \mu \end{cases} \tag{2.10}$$

και αποτελεί ένα μέτρο του πόσο διαφέρουν οι δύο κατανομές. Αποτελεί βασική έννοια σε πολλούς κλάδους επιστημών, ειδικά στην θεωρία πληροφοριών. Εκεί τη βρίσκει κανείς υπο την ονομασία *Απόκλιση (ή απόσταση) Kullback-Leibler* [7] και διαθέτει μάλιστα και ένα διακριτό ανάλογο:

$$H(P||Q) = \sum_{x\in\mathcal{S}} P(x)\frac{P(x)}{Q(x)}$$

που όμως δε θα μας απασχολήσει στην παρούσα εργασία. Στη θερμοδυναμική, η σχετική εντροπία διαφέρει από την 2.10 κατά ένα πρόσημο "-" στο ολοκλήρωμα. Άρα, όταν θα λέμε *ελαχιστοποίηση* της σχετικής εντροπίας, θα εννοούμε *μεγιστοποίηση* της φυσικής σχετικής εντροπίας.

Φυσικά, η σχετική εντροπία δεν είναι μετρική, αφού σίγουρα δεν είναι συμμετρική, δηλαδή $H(\nu|\mu) \neq H(\mu|\nu)$. Όμως, είναι μη-αρνητική, όπως θα διαπιστώσουμε πολύ σύντομα. Επίσης, είναι καλώς ορισμένη. Όμως, πριν παρουσιάσουμε τις βασικές ιδιότητες της σχετικής εντροπίας είναι σημαντικό να διατυπώσουμε εδώ δύο θεωρήματα που θα μας φανούν χρήσιμα.

Το θεώρημα του Prokhorov συνδέει την έννοια της σχετικής συμπάγειας με τη σφιχτότητα των μέτρων πιθανότητας. Υπενθυμίζουμε ότι ένα σύνολο K είναι σχετικά συμπαγές, εάν το $\bar{K}$ είναι συμπαγές. Το θεώρημα του Ulam μας δίνει έναν χαρακτηρισμό σφιχτότητας για τα μέτρα πιθανότητας. Η απόδειξη του θεωρήματος του Prokhorov ξεφεύγει από τους σκοπούς της παρούσας εργασίας. Το θεώρημα του Ulam αποδεικνύεται πολύ εύκολα με χρήση του θεωρήματος του Prokhorov.

Θεώρημα 14 (Prokhorov). *Έστω $\mathcal{S}$ πολωνικός χώρος και ένα σύνολο $K \subset \mathcal{M}_1(\mathcal{S})$. Τότε, το K είναι σχετικά συμπαγές αν και μόνο αν για κάθε $\varepsilon > 0$ υπάρχει $F_\varepsilon \subset \mathcal{S}$ συμπαγές, ώστε $\sup\limits_{\mu \in K} \mu(F_\varepsilon^c) < \varepsilon$.*

Η ισοδύναμη συνθήκη δεν είναι τίποτε άλλο από τη συνθήκη σφιχτότητας των μέτρων πιθανότητας.

Θεώρημα 15 (Ulam). *Εάν $\mathcal{S}$ είναι πολωνικός χώρος και $\mu \in \mathcal{M}_1(\mathcal{S})$, τότε για κάθε $\varepsilon > 0$ υπάρχει $F_\varepsilon \subset \mathcal{S}$ συμπαγές, ώστε $\mu(F_\varepsilon^c) < \varepsilon$*

Απόδειξη. Έστω $K = \{\mu\} \subset \mathcal{M}_1(\mathcal{S})$. Τότε,το K είναι συμπαγές και $K = \bar{K}$, δηλαδή το K είναι σχετικά συμπαγές. Άρα, από το θεώρημα του Prokhorov για κάθε $\varepsilon > 0$ υπάρχει $F_\varepsilon \subset \mathcal{S}$ συμπαγές, ώστε $\mu(F_\varepsilon^c) < \varepsilon$. $\square$

Τώρα είμαστε σε θέση να διατυπώσουμε τις βασικότερες ιδιότητες της σχετικής εντροπίας:

Πρόταση 15.1 (Ιδιότητες της $H(\nu|\mu)$). *Έστω $\nu, \mu \in \mathcal{M}_1(\mathcal{S})$ δύο κατανομές. Τότε:*

1. *Η $H(\nu|\mu)$ είναι καλώς ορισμένη.*

2. *$H(\nu|\mu) \geq 0$ και $H(\nu|\mu) = 0 \Leftrightarrow \nu = \mu$.*

3. ***(Ανισότητα της Εντροπίας)** Για κάθε $g : \mathcal{S} \to \mathbb{R}$ με $g \in b\mathcal{B}(\mathcal{S})$ (φραγμένη και Borel μετρήσιμη) ισχύει:*

$$H(\nu|\mu) \geq \int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu \tag{2.11}$$

4. *Ειδικότερα, η (2.11) γίνεται ακόμα πιο ισχυρή:*

$$H(\nu|\mu) = \sup_{f \in b\mathcal{B}(\mathcal{S})} \{ \int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu \} \tag{2.12}$$

Από αυτό συμπεραίνουμε ότι η $H(\nu|\mu)$ είναι κυρτή στον $\mathcal{M}_1(\mathcal{S})$.

5. *Αν $\mathcal{S}$ πολωνικός, τότε για κάθε $a \in \mathbb{R}$ τα $\Psi_a = \{\nu \in \mathcal{M}_1(\mathcal{S}) : H(\nu|\mu) \leq a\}$ είναι συμπαγή ως προς την ασθενή τοπολογία.*

Απόδειξη. 1. Εύκολα διαπιστώνουμε ότι: $x \log x \geq -\frac{1}{e}$. Συνεπώς, $(f \log f)^- = -\min(0, f \log f) \leq \frac{1}{e}$ και άρα το ολοκλήρωμα $\int_{\mathcal{S}} f \log f d\mu$ είναι καλώς ορισμένο.

2. Εάν $\nu \not\ll \mu$, τότε $H(\nu|\mu) = +\infty > 0$. Άρα αρκεί να θεωρήσουμε ότι $\nu \not\ll \mu$. Έστω τότε $f = \frac{d\nu}{d\mu}$. Εύκολα διαπιστώνει κανείς ότι η συνάρτηση:

$$g(x) = x \log x$$

είναι (αυστηρά) κυρτή στο $[0, +\infty]$. Άρα, εφαρμόζουμε την ανισότητα Jensen:

$$\begin{aligned} H(\nu|\mu) &= \int_{\mathcal{S}} g(f(x)) d\mu(x) = \int_{\mathcal{S}} g d(\mu \circ f^{-1}) \\ &\geq g\left(\int_{\mathcal{S}} x d(\mu \circ f^{-1})\right) = g\left(\int_{\mathcal{S}} f(x) d\mu(x)\right) \\ &= g(1) = 0 \end{aligned}$$

Επίσης, λόγω της αυστηρής κυρτότητας της g: $H(\nu|\mu) = 0 \Leftrightarrow f = c$ μ-σ.β. $\xRightarrow{\int_{\mathcal{S}} f d\mu = 1} f(x) = 1$, μ-σ.β., άρα $\nu = \mu$.

3. Επίσης δεν είναι δύσκολο να διαπιστώσει κανείς ότι:

$$\sup_x \{xy - (x \log x - x + 1)\} = e^y - 1$$

Και άρα αν $g \in b\mathcal{B}(\mathcal{S})$ παίρνουμε:

$$\begin{aligned} \int_{\mathcal{S}} g d\nu &= \int_{\mathcal{S}} g f d\mu \\ &\leq \int_{\mathcal{S}} (f \log f - f + 1 + e^g - 1)\, d\mu \\ &= H(\nu|\mu) - \int_{\mathcal{S}} f d\mu + \int_{\mathcal{S}} d\mu + \int_{\mathcal{S}} (e^g - 1)\, d\mu \\ &= H(\nu|\mu) - 1 + 1 + \int_{\mathcal{S}} (e^g - 1)\, d\mu \\ &= H(\nu|\mu) + \int_{\mathcal{S}} (e^g - 1)\, d\mu \end{aligned}$$

Κάνουμε τον μετασχηματισμό $g(x) \mapsto g(x) - c, \;\; c \in \mathbb{R}$:

$$\int_{\mathcal{S}} g d\nu \leq H(\nu|\mu) + c + \int_{\mathcal{S}} \left(e^{g-c} - 1\right) d\mu, \quad \forall c \in \mathbb{R}$$

Διαλέγουμε ως τιμή του c εκείνη που θα μας δώσει την κατάλληλη εκτίμηση. Εν προκειμένω, επιλέγουμε:

$$c = \log \int_{\mathcal{S}} e^g d\mu$$

Και άρα:

$$\int_{\mathcal{S}} g d\nu \leq H(\nu|\mu) + \log \int_{\mathcal{S}} e^g d\mu$$

Ή ισοδύναμα:

$$H(\nu|\mu) \geq \int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu$$

4. Μόλις αποδειξαμε ότι για κάθε $g \in b\mathcal{B}(\mathcal{S})$:

$$H(\nu|\mu) \geq \int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu$$

Και άρα:

$$H(\nu|\mu) \geq \sup_{g \in b\mathcal{B}(\mathcal{S})} \{\int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu\}$$

Χρειάζεται να βρούμε μια συνάρτηση στο $b\mathcal{B}(\mathcal{S})$, η οποία να πιάνει το sup. Έστω $\nu \ll \mu$ και $f = \dfrac{d\nu}{d\mu}$ η Radon-Nikodym παράγωγος. Θεωρούμε: $g = \log f$. Αν ίσχυε ότι $g \in b\mathcal{B}(\mathcal{S})$, τότε:

$$\int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu = \int_{\mathcal{S}} \log f d\nu - \log \int_{\mathcal{S}} f d\mu = \int_{\mathcal{S}} f \log f d\mu - \log 1 = H(\nu|\mu)$$

και άρα το sup θα πιανόταν πάνω στις φραγμένες, Borel μετρήσιμες συναρτήσεις. Όμως, επειδή η f δεν είναι απαραίτητα φραγμένη μακριά από το 0 και το $+\infty$, τότε ούτε και η g είναι απαραίτητα φραγμένη. Γι' αυτό, κάνουμε αποκοπή (truncation). Συγκεκριμένα, εάν ορίσουμε τις φ_n ως:

$$\varphi_n(x) = \begin{cases} \frac{1}{n}, & x \leq \frac{1}{n} \\ x, & \frac{1}{n} \leq x \leq n \\ n, & x \geq n \end{cases}$$

Τότε, ορίζουμε ως g_n τις:

$$g_n = \log \varphi_n(f) = \begin{cases} -\log n, & x \leq \frac{1}{n} \\ \log f, & \frac{1}{n} \leq x \leq n \\ \log n, & x \geq n \end{cases}$$

Και τότε $-\log n \leq g_n \leq \log n$, οπότε οι g_n είναι φραγμένες και μετρήσιμες, άρα, από το θεώρημα μονότονης σύγκλισης:

$$\int_{\mathcal{S}} g_n d\nu - \log \int_{\mathcal{S}} e^{g_n} d\mu \longrightarrow \int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu = H(\nu|\mu) - 0 = H(\nu|\mu)$$

Εάν $\nu \not\ll \mu$, τότε εξ ορισμού της απόλυτης συνέχειας θα υπάρχει $A \subset \mathcal{S}$, ώστε $\mu(A) = 0$ και $\nu(A) > 0$. Επίσης, εξ ορισμού της σχετικής εντροπίας: $H(\nu|\mu) = +\infty$. Άρα, ψάχνουμε φραγμένες, μετρήσιμες συναρτήσεις g_n, ώστε:

$$\int_{\mathcal{S}} g_n d\nu - \log \int_{\mathcal{S}} e^{g_n} d\mu \longrightarrow +\infty = H(\nu|\mu)$$

Διαλέγουμε $g_n = n\mathbb{1}_A$. Τότε, αφενός:

$$\int_{\mathcal{S}} g_n d\nu = n \int_{\mathcal{S}} \mathbb{1}_A d\nu = n\nu(A) > 0$$

Αφετέρου:

$$\log \int_{\mathcal{S}} e^{g_n} d\mu = \int_A e^n d\nu - \log \int_A 1 d\mu = \mu(A)e^n + 1 \cdot \mu(A^c) = \mu(A^c) = 1$$

Αφού το σύνολο A το διαλέξαμε έτσι ώστε $\mu(A) = 0$. Συνολικά τα παραπάνω μας δίνουν:

$$\int_{\mathcal{S}} g_n d\nu - \log \int_{\mathcal{S}} e^{g_n} d\mu = n\nu(A) \longrightarrow +\infty = H(\nu|\mu)$$

Και άρα συμπεραίνουμε την ισότητα στην 2.12. Η κυρτότητα της $H(\nu|\mu)$ προκύπτει τώρα εύκολα, όπως και στην προεργασία του θεωρήματος του Cramér.

5. Για την ιδιότητα αυτή θα χρησιμοποιήσουμε τα θεωρήματα των Prokhorov και Ulam. Η απόδειξη του [9] είναι πολύ σαφής και διαισθητική, οπότε από αυτή θα εμπνευστούμε για τα παρακάτω. Συγκεκριμένα, επειδή $\bar{\Psi}_a = \Psi_a$, τότε τα Ψ_a θα είναι σχετικά συμπαγή αν και μόνο αν είναι συμπαγή. Αρκεί να δείξουμε ότι για κάθε $\varepsilon > 0$ υπάρχει $F_\varepsilon \subset \mathcal{S}$ συμπαγές, ώστε $\sup_{\mu \in \Psi_a} \mu(F_\varepsilon^c) < \varepsilon$ και τότε από το θεώρημα του Prokhorov τα Ψ_a θα είναι συμπαγή.

   Έστω λοιπόν $\varepsilon > 0$ και $\delta > 0$. Για $\mu \in \mathcal{M}_1(\mathcal{S})$ από το θεώρημα του Ulam θα υπάρχει $K_\varepsilon \subset \mathcal{S}$ συμπαγές, ώστε $\mu(K_\varepsilon^c) < \delta$.

   Επίσης, από την 2.11 ξέρουμε ότι για κάθε $g \in b\mathcal{B}(\mathcal{S})$:

$$H(\nu|\mu) \geq \int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu, \quad \forall \nu \in \mathcal{M}_1(\mathcal{S})$$

   Έστω τώρα ένα $\nu \in \Psi_a$. Εάν $a < 0$, τότε $\Psi_a = \emptyset$, άρα μπορούμε να θεωρήσουμε ότι $a > 0$. Διαλέγουμε $g = \log\left(1 + \frac{1}{\delta}\right) \mathbb{1}_{K_\varepsilon^c}$. Τότε:

$$\begin{aligned} H(\nu|\mu) &\geq \int_{\mathcal{S}} \log\left(1 + \frac{1}{\delta}\right) \mathbb{1}_{K_\varepsilon^c} d\nu - \log \int_{\mathcal{S}} e^{\log\left(1+\frac{1}{\delta}\right)\mathbb{1}_{K_\varepsilon^c}} d\mu \\ &= \log\left(1 + \frac{1}{\delta}\right) \nu(K_\varepsilon^c) - \log\left(\mu(K_\varepsilon) + \left(1 + \frac{1}{\delta}\right)\mu(K_\varepsilon^c)\right) \\ &= \log\left(1 + \frac{1}{\delta}\right) \nu(K_\varepsilon^c) - \log\left(1 + \frac{1}{\delta}\mu(K_\varepsilon^c)\right) \end{aligned}$$

   Και άρα λύνοντας την παραπάνω ανισότητα παίρνουμε ότι:

$$\nu(K_\varepsilon^c) \leq \frac{H(\nu|\mu) + \log\left(1 + \frac{1}{\delta}\mu(K_\varepsilon^c)\right)}{\log\left(1 + \frac{1}{\delta}\right)}$$

Και άρα αφού λόγω υπόθεσης $\mu(K_\varepsilon^c) < \delta$ και $H(\nu|\mu) \leq a$:

$$\sup_{\nu \in \Psi_a} \nu(K_\varepsilon^c) \leq \frac{a + \log 2}{\log\left(1 + \dfrac{1}{\delta}\right)}$$

Εάν διαλέξουμε το $\delta = \delta(\varepsilon) > 0$ έτσι ώστε:

$$\frac{a + \log 2}{\log\left(1 + \dfrac{1}{\delta}\right)} < \frac{\varepsilon}{2}$$

Τότε παίρνουμε:

$$\sup_{\nu \in \Psi_a} \nu(K_\varepsilon^c) < \varepsilon, \quad \forall \varepsilon > 0$$

Οπότε από το θεώρημα του Prokhorov τα Ψ_a είναι συμπαγή για κάθε $a \in \mathbb{R}$.

$\square$

Επέκταση στον $C_b(\mathcal{S})$. Τελικά, το sup στην (2.12) πιάνεται στις συνεχείς και φραγμένες συναρτήσεις στον $\mathcal{S}$, δηλαδή ισχύει ότι:

$$H(\nu|\mu) = \sup_{f \in C_b(\mathcal{S})} \{\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu\} \tag{2.13}$$

Συγκεκριμένα, ως προς την ασθενή τοπολογία του $\mathcal{M}_1(\mathcal{S})$ η σχετική εντροπία είναι κάτω ημισυνεχής, ως το supremum συνεχών συναρτήσεων.

Απόδειξη. Ας θυμηθούμε ότι κάθε συνεχής συνάρτηση είναι (Borel) μετρήσιμη, αφού εν γένει αν f είναι μια συνεχής συνάρτηση και U ένα ανοιχτό σύνολο, τότε $f^{-1}(U)$ ανοιχτό, δηλαδή ανήκει στην κλάση Borel. Άρα, ισχύει ο εγκλεισμός $C_b(\mathcal{S}) \subseteq b\mathcal{B}(\mathcal{S})$. Συνεπώς:

$$\sup_{f \in C_b(\mathcal{S})} \{\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu\} \leq \sup_{f \in b\mathcal{B}(\mathcal{S})} \{\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu\} = H(\nu|\mu)$$

Άρα, αρκεί να αποδείξουμε και την ανάποδη ανισότητα στην παραπάνω σχέση. Για τον σκοπό αυτό ορίζουμε :

$$a \triangleq \sup_{f \in C_b(\mathcal{S})} \{\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu\}$$

και χωρίς βλάβη της γενικότητας υποθέτουμε ότι $a < +\infty$ (ειδάλλως η ανισότητα που προσπαθούμε να αποδείξουμε θα ισχύει τετριμμένα). Έστω $\mathcal{H}$ το σύνολο:

$$\mathcal{H} = \{f \in b\mathcal{B}(\mathcal{S}) : \int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu \leq a\}$$

Δείχνουμε αρχικά ότι $C_b(\mathcal{S}) \subseteq \mathcal{H}$. Πράγματι, εάν $g \in C_b(\mathcal{S})$, τότε:

$$\int_{\mathcal{S}} g d\nu - \log \int_{\mathcal{S}} e^g d\mu \leq \sup_{f \in C_b(\mathcal{S})} \{\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu\} = a$$

Και άρα $g \in \mathcal{H}$. Οπότε όντως $C_b(\mathcal{S}) \subseteq \mathcal{H}$.

Στη συνέχεια, θέλουμε να δείξουμε ότι $b\mathcal{B}(\mathcal{S}) \subset \mathcal{H}$. Γι'αυτό, θεωρούμε την ακολουθία $\{f_n\}_{n\in\mathbb{N}} \subseteq \mathcal{H}$ με $|f_n(x)| \leq M$ για κάποιο $M > 0$ και $f_n(x) \to f(x)$ για κάθε $x \in \mathcal{S}$. Αφού $f_n \in \mathcal{H}$, τότε από το θεώρημα κυριαρχημένης σύγκλισης:

$$a \geq \int_{\mathcal{S}} f_n d\nu - \log \int_{\mathcal{S}} e^{f_n} d\mu \to \int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu$$

Και άρα για το όριο: $f \in \mathcal{H}$, δηλαδή το σύνολο $\mathcal{H}$ είναι κλειστό. Από το Λήμμα A.11. στο [20] $b\mathcal{B}(\mathcal{S}) \subseteq \mathcal{H}$. Το ανωτέρω Λήμμα προκύπτει ως πόρισμα του θεωρήματος μονότονης κλάσης, όμως η απόδειξή του είναι αρκετά τεχνική και δεν εξυπηρετεί τους σκοπούς της παρούσας εργασίας, οπότε την παραλείπουμε.

Εν τέλει, παίρνουμε ότι εάν $f \in b\mathcal{B}(\mathcal{S})$ τότε $f \in \mathcal{H}$, δηλαδή:

$$\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu \leq a$$

Και συνεπώς:

$$a \leq H(\nu|\mu) = \sup_{f\in b\mathcal{B}(\mathcal{S})} \{\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu \leq a\} \leq a$$

Και άρα:

$$H(\nu|\mu) = a = \sup_{f\in C_b(\mathcal{S})} \{\int_{\mathcal{S}} f d\nu - \log \int_{\mathcal{S}} e^f d\mu\}$$

□

Συνοψίζοντας, η $H(\nu|\mu) : \mathcal{M}_1(\mathcal{S}) \to \mathbb{R}$, όπου $\mathcal{S}$ πολωνικός και $\mu \in \mathcal{M}_1(\mathcal{S})$ (σταθερό), είναι μια καλή, κυρτή συνάρτηση ρυθμού, αφού πέρα από κυρτή είναι και κάτω ημισυνεχής με συμπαγή level sets. Και τώρα είμαστε σε θέση να διατυπώσουμε και να αποδείξουμε το θεώρημα του Sanov.

2.3.4 Η απόδειξη του θεωρήματος του Sanov

Το θεώρημα του Sanov μας δίνει ότι η εμπειρική κατανομή ικανοποιεί την LDP με συνάρτηση ρυθμού την σχετική εντροπία. Συγκεκριμένα, αφού η L_n, που δίνεται από την 2.8 είναι μια τυχαία μεταβλητή, μπορούμε να ορίσουμε την κατανομή της ρ_n ως:

$$\forall A \in \mathcal{B}(\mathcal{M}_1(\mathcal{S})) : \quad \rho_n(A) = \mathbb{P}[L_n \in A]$$

Δηλαδή είναι ένα μέτρο πιθανότητας που παίρνει ενδεχόμενα του χώρου $M_1(\mathcal{S})$.

Θεώρημα 16 (**Θεώρημα του Sanov**). *Έστω $\mathcal{S}$ πολωνικός χώρος και $\{X_n\}_{n\in\mathbb{N}}$ μια ακολουθία ανεξάρτητων, ισόνομων τυχαίων μεταβλητών με τιμές $\mathcal{S}$ και κοινή κατανομή μ. Έστω επίσης ρ_n η κατανομή της εμπειρικής κατανομής τους L_n. Τότε, η $\{\rho_n\}_{n\in\mathbb{N}}$ ικανοποιεί την LDP στον χώρο $\mathcal{M}_1(\mathcal{S})$ με καλή, κυρτή συνάρτηση ρυθμού:*

$$I(\nu) = H(\nu|\mu), \quad \forall \nu \in \mathcal{M}_1(\mathcal{S})$$

Απόδειξη. Την απόδειξη θα την παρουσιάσουμε σε 3 βήματα. Πρώτα θα αποδείξουμε το άνω φράγμα της ασθενούς LDP για συμπαγή σύνολα. Έπειτα θα αποδείξουμε την εκθετική σφιχτότητα των $\{\rho_n\}_{n\in\mathbb{N}}$, οπότε η ασθενής LDP θα συνεπάγεται την πλήρη LDP. Τέλος, θα αποδείξουμε το κάτω φράγμα της LDP για ανοιχτά σύνολα.

- **Άνω φράγμα για συμπαγή σύνολα**

  Εδώ θα χρησιμοποιήσουμε τη γενική μεθοδολογία για το άνω φράγμα σε τοπικά κυρτούς, διανυσματικούς τοπολογικούς χώρους. Θέτουμε:

  $$\mathcal{X} = \mathcal{M}(\mathcal{S}),\ \ \mathcal{Y} = C_b(\mathcal{S}),\ \ \mathcal{E} = \mathcal{M}_1(\mathcal{S})$$

  όπου έχουμε ήδη αποδείξει ότι το $\mathcal{M}_1(\mathcal{S})$ είναι κυρτό και κλειστό. Επίσης, θέτουμε τη διγραμμική μορφή $\langle\cdot,\cdot\rangle : \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ να είναι η:

  $$\langle \mu, f\rangle = \int_{\mathcal{S}} f d\mu$$

  Παρατηρούμε επιπλέον ότι για κάθε $g\in C_b(\mathcal{S})$:

  $$\langle L_n, g\rangle = \int_{\mathcal{S}} g dL_n = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{S}} g d\delta_{X_i} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

  Συνεπώς:

  $$\begin{aligned}\Lambda(g) &= \limsup_{n\to+\infty}\frac{1}{n}\log\int_{\mathcal{S}} e^{n\langle x,g\rangle}d\rho_n(x)\\ &= \limsup_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{n\langle L_n,g\rangle}]\\ &= \limsup_{n\to+\infty}\frac{1}{n}\log\mathbb{E}[e^{g(X_i)}]^n\\ &= \limsup_{n\to+\infty}\log\mathbb{E}[e^{g(X_i)}]\\ &= \log\mathbb{E}[e^{g(X_i)}]\\ &= \log\int_{\mathcal{S}} e^g d\mu\end{aligned}$$

  Άρα:

  $$\begin{aligned}\Lambda^*(\nu) &= \sup_{g\in C_b(\mathcal{S})}\{\langle \nu, g\rangle - \log\int_{\mathcal{S}} e^g d\mu\}\\ &= \sup_{g\in C_b(\mathcal{S})}\{\int_{\mathcal{S}} g d\nu - \log\int_{\mathcal{S}} e^g d\mu\}\\ &= H(\nu|\mu)\end{aligned}$$

  Όπου την τελευταία ισότητα την αποδείξαμε προηγουμένως. Οπότε, από το γενικό άνω φράγμα για κάθε $K\subset\mathcal{M}_1(\mathcal{S})$ συμπαγές:

  $$\limsup_{n\to+\infty}\frac{1}{n}\log\rho_n(L)\le -\inf_{\nu\in K} H(\nu|\mu)$$

- **Εκθετική σφιχτότητα**

Από το θεώρημα του Ulam για το $\mu \in \mathcal{M}_1(\mathcal{S})$ θα έχουμε ότι για κάθε $\delta > 0$ υπάρχει $\Gamma_\delta \subset \mathcal{S}$ συμπαγές, ώστε $\mu(\Gamma_\delta^c) < e^{-2\delta^2}$.

Επίσης, αφού $\mathcal{S}$ πολωνικός, τότε και ο $\mathcal{M}_1(\mathcal{S})$ με την ασθενή τοπολογία είναι πολωνικός από θεώρημα που έχουμε ήδη διατυπώσει. Άρα, η κλειστότητα των υποσυνόλων του μπορεί να αποφασιστεί με ακολουθίες, μέσω της ασθενούς σύγκλισης. Έστω λοιπόν τα σύνολα:

$$A_\delta = \{\nu \in M_1(\mathcal{S}) : \nu(\Gamma_\delta) \geq 1 - \frac{1}{\delta}\} = \{\nu \in M_1(\mathcal{S}) : \nu(\Gamma_\delta^c) \leq \frac{1}{\delta}\}$$

Τότε, εάν $\{\nu_n\}_{n\in\mathbb{N}} \subseteq A_\delta$ με $\nu_n \Rightarrow \nu \in \mathcal{M}_1(\mathcal{S})$ θα έχουμε από το θεώρημα Portmanteau:

$$\nu(\Gamma_\delta) \geq \limsup_{n\to+\infty} \nu_n(\Gamma_\delta) \geq 1 - \frac{1}{\delta}$$

Και άρα $\nu \in A_\delta$, δηλαδή τα A_δ είναι κλειστά.

Έστω τώρα ότι $\Delta \in \mathbb{N}$ και $K_\Delta = \bigcap_{\delta\geq\Delta} A_\delta$. Τότε, το K_Δ είναι κλειστό ως τομή κλειστών (ας θυμηθούμε ότι η κλειστότητα διατηρείται ως προς αυθαίρετο πλήθος τομών). Άρα, $\bar{K}_\Delta = K_\Delta$. Επίσης, $\nu \in K_\Delta \Leftrightarrow \nu \in A_\delta$ για κάθε $\delta \geq \Delta$, οπότε $\nu(\Gamma_\delta^c) \leq \dfrac{1}{\delta}$ για κάθε $\delta \geq \Delta$.

Άρα, για κάθε $\tilde{\varepsilon} > 0$ διαλέγω $\delta = \delta(\tilde{\varepsilon})$ με $\dfrac{1}{\delta} < \tilde{\varepsilon}$ και $\delta \geq \Delta$, οπότε:

$$\nu(\Gamma_\delta) < \tilde{\varepsilon} \Rightarrow \sup_{\nu\in K_\Delta} \nu(\Gamma_\delta^c) < \tilde{\varepsilon}$$

Άρα, από το θεώρημα Prokhorov τα K_Δ είναι συμπαγή.

Υπολογίζουμε:

$$\rho_n(A_\delta^c) = \mathbb{P}\left(L_n(\Gamma_\delta^c) > \frac{1}{\delta}\right) = \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\Gamma_\delta^c}(X_i) > \frac{1}{\delta}\right) = \mathbb{P}\left(2\delta^2\sum_{i=1}^n \mathbb{1}_{\Gamma_\delta^c}(X_i) > 2n\delta\right)$$

Από την ανισότητα Markov, καθώς και από τη συνθήκη του θεωρήματος του Ulam θα πάρουμε ότι:

$$\begin{aligned}
\mathbb{P}\left(2\delta^2\sum_{i=1}^n \mathbb{1}_{\Gamma_\delta^c}(X_i)\right) &\leq e^{-2n\delta}\mathbb{E}\left(e^{\sum_{i=1}^n \mathbb{1}_{\Gamma_\delta^c}(X_i)}\right) \\
&= e^{-2n\delta}\mathbb{E}\left(e^{\mathbb{1}_{\Gamma_\delta^c}(X_i)}\right)^n \\
&= e^{-2n\delta}\left(e^{2\delta^2}\mu(\Gamma_\delta^c) + \mu(\Gamma_\delta)\right)^n \\
&\leq e^{-2n\delta}\left(e^{2\delta^2}e^{-2\delta^2} + 1\right)^n
\end{aligned}$$

$$
\begin{aligned}
&= 2^n e^{-2n\delta} \\
&= e^{-n\delta}\left(2e^{-\delta}\right)^n
\end{aligned}
$$

Αφού $K_\Delta^c = \bigcup_{\delta \geq \Delta} A_\delta^c$, τότε:

$$
\begin{aligned}
\rho_n(K_\Delta^c) &= \rho_n\left(\bigcup_{\delta \geq \Delta} A_\delta^c\right) \\
&\leq \sum_{\delta \geq \Delta} \rho_n(A_\delta^c) \\
&\leq \sum_{\delta \geq \Delta} e^{-n\delta}(2e^{-\delta})^n \\
&\leq \sum_{\delta \geq \Delta} e^{-n\delta} \\
&= \frac{e^{-n\Delta}}{1 - e^{-n}}
\end{aligned}
$$

όπου η προτελευταία ανισότητα προέκυψε καθώς για κάθε $\Delta \in \mathbb{N}$ θα ισχύει ότι: $\delta \geq 1 = \log e > \log 2$, οπότε $2e^{-\delta} \leq 1$. Η τελευταία ισότητα προκύπτει ως το αποτέλεσμα μιας γεωμετρικής σειράς. Συνεπώς:

$$
\limsup_{n \to +\infty} \frac{1}{n} \log \rho_n(K_\Delta^c) \leq -\Delta + \limsup_{n \to +\infty} \frac{1}{n} \log(1 - e^{-n}) = -\Delta < \varepsilon
$$

αφού $\Delta \in \mathbb{N}$ και $\varepsilon > 0$. Άρα, η $\{\rho_n\}_{n \in \mathbb{N}}$ είναι εκθετικά σφιχτή. Επομένως, αν αποδείξουμε την ασθενή LDP, θα έχουμε κατευθείαν την πλήρη LDP.

- **Κάτω φράγμα για ανοιχτά σύνολα**

Εάν $\nu \not\ll \mu$, τότε $H(\nu|\mu) = +\infty$ και το κάτω φράγμα ισχύει τετριμμένα.

Έστω ότι $\nu \ll \mu$ και $\varphi = \dfrac{d\nu}{d\mu}$ η παράγωγος Radon-Nikodym του ν ως προς μ. Έστω επίσης $U_\nu \subset \mathcal{M}_1(\mathcal{S})$ μια γειτονιά του ν. Για να αποδείξουμε το κάτω φράγμα, θα κάνουμε εκθετικό πλάγιασμα. Έχουμε:

$$
\begin{aligned}
\rho_n(U_\nu) &= \mathbb{P}[L_n \in U_\nu] \\
&= \int_{\{L_n \in U_\nu\}} d\mu(x_1) \ldots d\mu(x_n) \\
&\geq \int_{\{L_n \in U_\nu\}} \frac{1}{\varphi(x_1)} \cdots \frac{1}{\varphi(x_n)} d\nu(x_1) \ldots d\nu(x_n)
\end{aligned}
$$

$$= \int\limits_{\{L_n \in U_\nu\}} e^{-\sum\limits_{i=1}^{n} \log \varphi(x_i)} d\nu(x_1) \dots d\nu(x_n)$$

$$\geq \int\limits_{\{L_n \in U_\nu\} \cap \{|\int \log \varphi dL_n - H(\nu|\mu)| \leq \varepsilon\}} e^{-\sum\limits_{i=1}^{n} \log \varphi(x_i)} d\nu(x_1) \dots d\nu(x_n)$$

$$\geq e^{-n(\varepsilon + H(\nu|\mu))} \nu^{(n)} \left(\{L_n \in U_\nu\} \bigcap \{ \left| \int \log \varphi dL_n - H(\nu|\mu) \right| \leq \varepsilon \} \right)$$
$$= e^{-n(\varepsilon + H(\nu|\mu))} \mathbb{P} \left(L_n^Y \in U_\nu, \left| \frac{1}{n} \sum_{i=1}^{n} \log \varphi(Y_i) - H(\nu|\mu) \right| \leq \varepsilon \right)$$

όπου οι $\{Y_i\}_{i=1,\dots,n}$ είναι ανεξάρτητες, ισόνομες με κατανομή ν και η L_n^Y είναι η εμπειρική κατανομή τους. Η προτελευταία ανισότητα προκύπτει, αφού ο επιπλέον περιορισμός στο ολοκλήρωμα μας έδωσε:

$$\frac{1}{n} \sum_{i=1}^{n} \log \varphi(X_i) \leq \varepsilon + H(\nu|\mu)$$

Ας δούμε τώρα πως μας βοήθησε το εκθετικό πλάγιασμα, δηλαδή πως μπορούμε να εφαρμόσουμε τον νόμο των μεγάλων αριθμών. Αφού $L_n^Y \in U_\nu$, τότε για κάθε $\varepsilon > 0$, $m = 1, ..., k$ και $g_m \in C_b(\mathcal{S})$ θα έχουμε ότι:

$$|\langle L_n^Y, g_m \rangle - \langle \nu, g_m \rangle| < \varepsilon \Leftrightarrow \left| \frac{1}{n} \sum_{i=1}^{n} g_m(Y_i) - \int_{\mathcal{S}} g_m d\nu \right| < \varepsilon$$

όμως υπό το μέτρο ν:

$$\int_{\mathcal{S}} g_m d\nu = \mathbb{E}^\nu[g_m(Y_1)]$$

Ενώ για τη σχετική εντροπία έχουμε:

$$H(\nu|\mu) = \int_{\mathcal{S}} \log \varphi d\nu = \mathbb{E}^\nu[\log g(Y_1)]$$

Θέτουμε $g_0 = \log \varphi$. Τότε, για $\tilde{\varepsilon} > 0$ με $\tilde{\varepsilon} < \varepsilon$ θα πάρουμε:

$$\mathbb{P} \left(L_n^Y \in U_\nu, \left| \frac{1}{n} \sum_{i=1}^{n} \log \varphi(Y_i) - H(\nu|\mu) \right| \leq \varepsilon \right) =$$
$$= \mathbb{P} \left(\bigcap_{m=1}^{k} \{ \left| \frac{1}{n} \sum_{i=1}^{n} g_m(Y_i) - \mathbb{E}^\nu[g_m(Y_1)] \right| < \varepsilon \}, \{ \left| \frac{1}{n} \sum_{i=1}^{n} \log \varphi(Y_i) - H(\nu|\mu) \right| \leq \varepsilon \} \right) \geq$$
$$\geq \mathbb{P} \left(\bigcap_{m=0}^{k} \{ \left| \frac{1}{n} \sum_{i=1}^{n} g_m(Y_i) - \mathbb{E}^\nu[g_m(Y_1)] \right| < \tilde{\varepsilon} \} \right) \longrightarrow 1$$

Η τελευταία σύγκλιση ισχύει χάρη στο Νόμο των Μεγάλων Αριθμών. Πράγματι, αφού οι $\{Y_i\}$ είναι ανεξάρτητες και ισόνομες, θα έχουμε:

$$\begin{aligned}\mathbb{P}\left(\bigcap_{m=0}^{k}\{\left|\frac{1}{n}\sum_{i=1}^{n}g_m(Y_i)-\mathbb{E}^{\nu}[g_m(Y_1)]\right|<\tilde{\varepsilon}\}\right) &= \prod_{m=0}^{k}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}g_m(Y_i)-\mathbb{E}^{\nu}[g_m(Y_1)]\right|<\tilde{\varepsilon}\right)\\ &= \mathbb{P}^{m}\left(\left|\frac{1}{n}\sum_{i=1}^{n}g_m(Y_i)-\mathbb{E}^{\nu}[g_m(Y_1)]\right|<\tilde{\varepsilon}\right)\end{aligned}$$

Αφού οι g_m είναι φραγμένες, τότε από τον νόμο των μεγάλων αριθμών θα έχουμε ότι:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}g_m(Y_i)-\mathbb{E}^{\nu}[g_m(Y_1)]\right|<\tilde{\varepsilon}\right)\longrightarrow 1$$

Και άρα:

$$\mathbb{P}\left(\bigcap_{m=0}^{k}\{\left|\frac{1}{n}\sum_{i=1}^{n}g_m(Y_i)-\mathbb{E}^{\nu}[g_m(Y_1)]\right|<\tilde{\varepsilon}\}\right)\longrightarrow 1$$

Τελικά, για το κάτω φράγμα θα ισχύει:

$$\liminf_{n\to+\infty}\frac{1}{n}\log\mathbb{P}[L_n\in U_\nu]\geq -H(\nu|\mu)-\tilde{\varepsilon}$$

για κάθε $\tilde{\varepsilon}>0$ με $\tilde{\varepsilon}<\varepsilon$. Στέλνοντας το $\tilde{\varepsilon}\to 0$:

$$\liminf_{n\to+\infty}\frac{1}{n}\log\mathbb{P}[L_n\in U_\nu]\geq -H(\nu|\mu)$$

Και συνεπώς:

$$\liminf_{n\to+\infty}\frac{1}{n}\log\mathbb{P}[L_n\in U_\nu]\geq -\inf_{\nu\in U_\nu}H(\nu|\mu)$$

Η απόδειξη του κάτω φράγματος έχει στην ουσία ολοκληρωθεί, αφού αν $G\subset\mathcal{M}_1(\mathcal{S})$ ανοιχτό, τότε $G=\bigcup_{\nu\in\mathcal{M}_1(\mathcal{S})}U_\nu$, δηλαδή γράφεται σαν ένωση ανοιχτών περιοχών του ν. Άρα, εάν $\nu\in G$, τότε $U_\nu\subset G$, οπότε:

$$\liminf_{n\to+\infty}\frac{1}{n}\log\rho_n(G)\geq\liminf_{n\to+\infty}\frac{1}{n}\log\rho_n(U_\nu)\geq -H(\nu|\mu)$$

για κάθε $\nu\in G$ και συνεπώς:

$$\liminf_{n\to+\infty}\frac{1}{n}\log\rho_n(G)\geq -\inf_{\nu\in G}H(\nu|\mu)$$

Η απόδειξη του θεωρήματος του Sanov έχει ολοκληρωθεί. □

Κεφάλαιο 3

Η αρχή του Gibbs

Έστω ότι σε ένα σμήνος σωματιδίων μας δίνεται μια παρατήρηση για την συμπεριφορά της εμπειρικής τους κατανομής, για παράδειγμα στη μορφή ενός ενεργειακού περιορισμού. Θα θέλαμε να εξάγουμε κάποιο συμπέρασμα για την κατανομή των σωματιδίων, δεδομένου αυτού του περιορισμού. Η αρχή του Gibbs (Gibbs conditioning principle) μας βοηθάει με το ερώτημα αυτό. Στην ουσία, η αρχή αυτή μας λέει το εξής: Έστω ότι $\{X_i\}_{1\leq i\leq n}$ είναι μια ακολουθία ανεξάρτητων, ισόνομων τυχαίων μεταβλητών. Τότε, δοθέντος ότι η εμπειρική τους κατανομή ανήκει σε κάποιο "σπάνιο σύνολο" A, η κατανομή της X_1 συγκλίνει στην κατανομή που ελαχιστοποιεί τη σχετική εντροπία στο σύνολο A. Το αποτέλεσμα αυτό είναι εντυπωσιακό και η σημασία του είναι μεγάλη στη στατιστική φυσική: Υπό έναν ενεργειακό περιορισμό, η φύση μεγιστοποιεί τη εντροπία.

Η αρχή του Gibbs αποτελεί μια γενικότερη δήλωση για τον τρόπο που ο περιορισμός της εμπειρικής κατανομής αλλάζει την κατανομή των τυχαίων μεταβλητών. Εμείς εδώ θα αποδείξουμε δύο σημαντικά αποτελέσματα: την αρχή της μέγιστης εντροπίας, που αποτελεί στην ουσία την αρχή του Gibbs και ένα πόρισμά της τόσο σημαντικό για τη στατιστική μηχανική, που αποκαλείται πολλές φορές από μόνο του αρχή του Gibbs.

3.1 Αρχή της μέγιστης εντροπίας

Το 1975 ο Csiszár βρήκε μια ταυτότητα πολύ χρήσιμη, καθώς μελετούσε το θεώρημα του Sanov και την αρχή του Gibbs [10].

Πρόταση 16.1 (Νόμος Παραλληλογράμμου του Csiszár). *Εάν $\pi, \nu, \mu \in \mathcal{M}_1(\mathcal{S})$, τότε ισχύει ο Νόμος Παραλληλογράμμου του Csiszár:*

$$H(\pi|\mu) + H(\nu|\mu) = 2H\left(\frac{\pi+\nu}{2}|\mu\right) + H\left(\pi|\frac{\pi+\nu}{2}\right) + H\left(\nu|\frac{\pi+\nu}{2}\right) \tag{3.1}$$

Απόδειξη. Έστω ότι $\pi \ll \mu$ και $\nu \ll \mu$. Εάν $f_1 = \dfrac{d\pi}{d\mu}$ και $f_2 = \dfrac{d\nu}{d\mu}$, τότε:

$$\frac{d\left(\dfrac{\pi+\nu}{2}\right)}{d\mu} = \frac{1}{2}f_1 + \frac{1}{2}f_2, \quad \frac{d\nu}{d\left(\dfrac{\pi+\nu}{2}\right)} = \frac{d\nu}{d\mu}\frac{2}{f_1+f_2} = \frac{2f_2}{f_1+f_2}$$

$$\frac{d\pi}{d\left(\dfrac{\pi+\nu}{2}\right)} = \frac{d\pi}{d\mu}\frac{2}{f_1+f_2} = \frac{2f_1}{f_1+f_2}$$

Και άρα:

$$\begin{aligned}
&2H(\frac{\pi+\nu}{2}|\mu) + H(\pi|\frac{\pi+\nu}{2}) + H(\nu|\frac{\pi+\nu}{2}) = \\
&= \int_{\mathcal{S}}(f_1+f_2)\log\left(\frac{f_1+f_2}{2}\right)d\mu + \int_{\mathcal{S}}\frac{2f_1}{f_1+f_2}\log\left(\frac{2f_1}{f_1+f_2}\right)d\left(\frac{\pi+\nu}{2}\right) \\
&+ \int_{\mathcal{S}}\frac{2f_2}{f_1+f_2}\log\left(\frac{2f_2}{f_1+f_2}\right)d\left(\frac{\pi+\nu}{2}\right) \\
&= \int_{\mathcal{S}}(f_1+f_2)\log\left(\frac{f_1+f_2}{2}\right)d\mu + \int_{\mathcal{S}}\frac{2f_1}{f_1+f_2}\log\left(\frac{2f_1}{f_1+f_2}\right)\frac{f_1+f_2}{2}d\mu \\
&+ \int_{\mathcal{S}}\frac{2f_2}{f_1+f_2}\log\left(\frac{2f_2}{f_1+f_2}\right)\frac{f_1+f_2}{2}d\mu \\
&= \int_{\mathcal{S}}[(f_1+f_2)\log\left(\frac{f_1+f_2}{2}\right) + f_1\log\left(\frac{2f_1}{f_1+f_2}\right) + f_2\log\left(\frac{2f_2}{f_1+f_2}\right)]d\mu \\
&= \int_{\mathcal{S}}f_1\log f_1 d\mu + \int_{\mathcal{S}}f_2\log f_2 d\mu \\
&= H(\pi|\mu) + H(\nu|\mu)
\end{aligned}$$

□

Η παραπάνω ταυτότητα θα μας βοηθήσει να αποδείξουμε την μοναδικότητα στην αρχή της μέγιστης εντροπίας. Παρακάτω θα συμβολίζουμε με ρ_n^C τη δεσμευμένη κατανομή της εμπειρικής κατανομής L_n, δεδομένου ότι $L_n \in C$, όπου $C \subset \mathcal{M}_1(\mathcal{S})$ είναι ένα κλειστό και κυρτό σύνολο. Δηλαδή:

$$\rho_n^C \in \mathcal{M}_1(\mathcal{M}_1(\mathcal{S})) : \quad \rho_n^C(A) = \mathbb{P}[L_n \in A | L_n \in C], \ \forall A \in \mathcal{B}(\mathcal{M}_1(\mathcal{S}))$$

Θα διατυπώσουμε την αρχή της μέγιστης εντροπίας σε δύο μέρη.

Πρόταση 16.2 (Αρχή της Μέγιστης Εντροπίας: Μέρος 1). *Έστω $C \subset \mathcal{M}_1(\mathcal{S})$ κλειστό και κυρτό, ώστε:*

$$\inf_{\nu \in C} H(\nu|\mu) = \inf_{\nu \in C^o} H(\nu|\mu) < +\infty$$

Τοτε, υπάρχει μοναδικό $\nu_* \in C$ που ελαχιστοποιεί τη σχετική εντροπία $H(\nu|\mu)$ στο C και $\rho_n^C \Rightarrow \delta_{\nu_*}$, δηλαδή:

$$\lim_{n\to+\infty}\mathbb{P}[L_n \in A | L_n \in C] = \delta_{\nu_*}(A), \ \forall A \in \mathcal{B}(\mathcal{M}_1(\mathcal{S}))$$

στην ασθενή τοπολογία του $\mathcal{M}_1(\mathcal{M}_1(\mathcal{S}))$, όπου δ_ν είναι το μέτρο Dirac στον $\mathcal{M}_1(\mathcal{S})$.

Απόδειξη. Για την ύπαρξη, η $H(\nu|\mu)$ είναι καλή συνάρτηση ρυθμού και άρα λαμβάνει το ελάχιστό της σε κλειστά σύνολα. Αφού το C είναι κλειστό, τότε υπάρχει ν_* που ελαχιστοποιεί την $H(\cdot|\mu)$ στο C.

Για τη μοναδικότητα, έστω ότι υπάρχουν δύο μέτρα $\nu_1, \nu_2 \in C$, ώστε,

$$\inf_{\nu \in C} H(\nu|\mu) = H(\nu_1|\mu), \quad \inf_{\nu \in C} H(\nu|\mu) = H(\nu_2|\mu)$$

Ονομάζουμε $\inf\limits_{\nu \in C} H(\nu|\mu) = k$. Από τον νόμο παραλληλογράμμου του Csiszár:

$$H(\nu_1|\mu) + H(\nu_2|\mu) = 2H\left(\frac{\nu_1+\nu_2}{2}|\mu\right) + H\left(\nu_1|\frac{\nu_1+\nu_2}{2}\right) + H\left(\nu_2|\frac{\nu_1+\nu_2}{2}\right)$$

Αφού το C είναι κυρτό, τότε $\dfrac{\nu_1+\nu_2}{2} \in C$, οπότε,

$$H\left(\frac{\nu_1+\nu_2}{2}|\mu\right) \geq \inf_{\nu \in C} H(\nu|\mu) = k$$

Άρα θα πάρουμε,

$$k + k \geq 2k + H\left(\nu_1|\frac{\nu_1+\nu_2}{2}\right) + H\left(\nu_2|\frac{\nu_1+\nu_2}{2}\right)$$

Αφού εξ υποθέσεως $k < \infty$ θα έχουμε:

$$H\left(\nu_1|\frac{\nu_1+\nu_2}{2}\right) + H\left(\nu_2|\frac{\nu_1+\nu_2}{2}\right) \leq 0$$

Αλλά, όπως έχουμε δει, η σχετική εντροπία είναι πάντοτε μη-αρνητική, οπότε θα πρέπει:

$$\begin{cases} H\left(\nu_1|\frac{\nu_1+\nu_2}{2}\right) = 0 \Leftrightarrow \nu_1 = \frac{\nu_1+\nu_2}{2} \\ H\left(\nu_2|\frac{\nu_1+\nu_2}{2}\right) = 0 \Leftrightarrow \nu_2 = \frac{\nu_1+\nu_2}{2} \end{cases}$$

Και άρα $\nu_1 = \nu_2$, δηλαδή το ελάχιστο είναι μοναδικό.

Για να δείξουμε την ασθενή σύγκλιση θεωρούμε μια ανοιχτή περιοχή U του ν_*. Έστω ρ_n η κατανομή της εμπειρικής κατανομής, όπως ορίστηκε στο προηγούμενο κεφάλαιο. Αφού $\nu_* \notin U^c$ και U^c κλειστό, χρησιμοποιώντας το θεώρημα του Sanov και τον ορισμό της δεσμευμένης πιθανότητας, θα πάρουμε ότι:

$$\begin{aligned}
\limsup_{n\to+\infty} \frac{1}{n} \log \rho_n^C(U^c) &= \limsup_{n\to+\infty} \frac{1}{n} \log \rho_n(U^c \cap C) - \liminf_{n\to+\infty} \frac{1}{n} \log \rho_n(C) \\
&\leq - \inf_{\nu \in U^c \cap C} H(\nu|\mu) - \liminf_{n\to+\infty} \frac{1}{n} \log \rho_n(C^o) \\
&\leq - \inf_{\nu \in U^c \cap C} H(\nu|\mu) + \inf_{\nu \in C^o} H(\nu|\mu) \\
&= - \inf_{\nu \in U^c \cap C} H(\nu|\mu) + H(\nu_*|\mu) < 0
\end{aligned}$$

Αφού $H(\nu_*|\mu) = \inf\limits_{\nu \in C} H(\nu|\mu) < \inf\limits_{\nu \in U^c \cap C} H(\nu|\mu)$.

Έτσι, λοιπόν, εάν $A \subset \mathcal{M}_1(\mathcal{S})$ είναι κλειστό, θα έχουμε ότι:

- Αν $\nu_* \in A$: $\limsup_{n\to+\infty} \rho_n^C(A) \leq 1 = \delta_{\nu_*}(A)$, αφού το $\rho_n^C(\cdot)$ είναι μέτρο πιθανότητας.
- Αν $\nu_* \notin A$: Παίρνω για $U = A^c$. Τότε, το U είναι ανοιχτό και $\nu_* \in U$. Οπότε,
$$\limsup_{n\to+\infty} \rho_n^C(A) = 0 = \delta_{\nu_*}(A)$$

Άρα, από το θεώρημα Portmanteau θα έχουμε: $\rho_n^C \Rightarrow \delta_{\nu_*}$. □

Η επόμενη Πρόταση ολοκληρώνει την παραπάνω:

Πρόταση 16.3 (Αρχή της Μέγιστης Εντροπίας: Μέρος 2). *Έστω ότι $\rho_n^C \Rightarrow \delta_{\nu_*}$. Εάν $k \in \mathbb{N}$ και $\mu_n^C(\cdot) = \mathbb{P}[X_k \in \cdot | L_c \in C]$, τότε $\mu_n^C \Rightarrow \nu_*$, δηλαδή για κάθε $f \in C_b(\mathcal{S})$:*
$$\lim_{n\to+\infty} \mathbb{E}[f(X_k)|L_n \in C] = \int f d\nu_*$$
όπου το ν_* είναι το μοναδικό στοιχείο του $C \subset \mathcal{M}_1(\mathcal{S})$ που ελαχιστοποιεί τη σχετική εντροπία στο C.

Τέλος, εάν $X^k = (X_1, ..., X_k)$ και $\mu_n^C(\cdot) = \mathbb{P}[X^k \in \cdot | L_n \in C]$, τότε: $\mu_n^C \Rightarrow \nu_*^k$, όπου το ν_*^k είναι ομοίως το μοναδικό στοιχείο του $C \subset \mathcal{M}_1(\mathcal{S}^k)$ που ελαχιστοποιεί τη σχετική εντροπία στο C.

Απόδειξη. Δείχνουμε πρώτα ότι $\mu_n^C \Rightarrow \nu_*$. Έστω $f \in C_b(\mathcal{S})$ και $F(v) = \int f d\nu$ με $v \in \mathcal{M}_1(\mathcal{S})$. Η F είναι προφανώς συνεχής και φραγμένη από τον τρόπο που έχουμε ορίσει τη σύγκλιση στην ασθενή τοπολογία του $\mathcal{M}_1(\mathcal{S})$. Επίσης, αφού οι $\{X_i\}$ είναι ανεξάρτητες και ισόνομες:
$$\begin{aligned}\int f d\mu_n^C &= \mathbb{E}[f(X_k)|L_n \in C] \\ &= \mathbb{E}[\frac{1}{n}\sum_{k=1}^{n} f(X_k)|L_n \in C] \\ &= \mathbb{E}[\int f dL_n | L_n \in C] \\ &= \mathbb{E}[F(L_n)|L_n \in C] \\ &= \int F d\rho_n^C \longrightarrow \int F d\delta_{\nu_*} = F(\nu_*) = \int f d\nu_*\end{aligned}$$
όπου η σύγκλιση προέκυψε από την Πρόταση 1. Άρα, $\mu_n^C \Rightarrow \nu_*$.

Θα δείξουμε τώρα ότι $\mu_n^C \Rightarrow \nu_*^k$. Έστω $\varphi_j \in C_b(\mathcal{S})$, $j = 1, ..., k$. Θέλουμε να δείξουμε ότι:
$$\int \varphi_1 \cdots \varphi_k d\mu_n^C \longrightarrow \int \varphi_1 \cdots \varphi_k d\nu_*^k$$
Αρχικά παρατηρούμε ότι η κατανομή μ_n^C είναι αναλλοίωτη σε μεταθέσεις των $X_1, ..., X_k$. Ισχύει ότι:
$$\int \varphi_1 \cdots \varphi_k d\mu_n^C = \int \prod_{j=1}^{k} \varphi_j d\mu_n^C = \mathbb{E}[\prod_{j=1}^{k} \varphi_j(X_j)|L_n \in C]$$

Σχηματίζουμε τον μέσο όρο, όπως και πριν, αλλά τώρα πάνω σε όλες τις διατεταγμένες k-άδες $(i_1, ..., i_k) \in 1, ..., n^k$ χωρίς επαναλήψεις:

$$\langle \varphi_1 \cdots \varphi_k, \mu_n^C \rangle = \int \varphi_1 \cdots \varphi_k d\mu_n^C = \frac{(n-k)!}{k!} \sum_{i_1 \neq \ldots \neq i_k} \int_{\mathcal{S}^k} \varphi_1(y_{i_1}) \cdots \varphi_k(y_{i_k}) d\mu_n^C(\vec{y})$$

Όμως,

$$\left(\frac{\varphi_1(X_1) + \cdots + \varphi_1(X_n)}{n} \right) \cdots \left(\frac{\varphi_k(X_1) + \cdots + \varphi_k(X_n)}{n} \right) = \frac{1}{n^k} \sum_{i_1, \ldots, i_k} \varphi_1(X_{i_1}) \cdots \varphi_k(X_{i_k})$$

Οπότε:

$$\begin{aligned}
\mathbb{E}[\langle \varphi_1, L_n \rangle \cdots \langle \varphi_k, L_n \rangle | L_n \in C] &= \frac{1}{n^k} \sum_{i_1, \ldots, i_k} \mathbb{E}[\varphi_1(X_{i_1}) \cdots \varphi_k(X_{i_k}) | L_n \in C] \\
&= \frac{1}{n^k} \sum_{i_1, \ldots, i_k} \int_{\mathcal{S}^k} \varphi_1(y_{i_1}) \cdots \varphi_k(y_{i_k}) d\mu_n^C(\vec{y})
\end{aligned}$$

Αφού $\varphi_j \in C_b(\mathcal{S})$ για κάθε $j = 1, ..., k$, τότε υπάρχουν σταθερές $M_1, ..., M_k > 0$, ώστε $|\varphi_j| \leq M_j$ για όλα τα $j = 1, .., k$, Οπότε:

$$|\varphi_1 \cdots \varphi_k| \leq M_1 \cdots M_k = M$$

Οπότε τελικά έχουμε:

$$\begin{aligned}
&|\langle \varphi_1 \cdots \varphi_k, \mu_n^C \rangle - \mathbb{E}[\langle \varphi_1, L_n \rangle \cdots \langle \varphi_k, L_n \rangle | L_n \in C]| = \\
&= \left| \frac{(n-k)!}{n!} \sum_{i_1 \neq \ldots \neq i_k} \int \varphi_1 \cdots \varphi_k d\mu_n^C - \frac{1}{n^k} \sum_{i_1, \ldots, i_k} \int \varphi_1 \cdots \varphi_k d\mu_n^C \right| \\
&= \left| \frac{(n-k)!}{n!} \sum_{i \in \Pi_1^k} \int \varphi_1 \cdots \varphi_k d\mu_n^C - \frac{1}{n^k} \sum_{i \in \Pi_2^k} \int \varphi_1 \cdots \varphi_k d\mu_n^C \right| \\
&= \left| \left(\frac{(n-k)!}{n!} - \frac{1}{n^k} \right) \sum_{i \in \Pi_1^k} \int \varphi_1 \cdots \varphi_k d\mu_n^C - \frac{1}{n^k} \sum_{i \in \Pi_2^k \setminus \Pi_1^k} \int \varphi_1 \cdots \varphi_k d\mu_n^C \right| \\
&\leq \frac{(n-k)! n^k - n!}{n^k n!} M \frac{n!}{(n-k)!} + \frac{1}{n^k} M \left(n^k - \frac{n!}{(n-k)!} \right) \\
&= \left(1 - \frac{n!}{n^k (n-k)!} \right) M + \left(1 - \frac{n!}{n^k (n-k)!} \right) M
\end{aligned}$$

$$= 2M \left(1 - \frac{n!}{n^k (n-k)!}\right)$$
$$= 2M \left(1 - \frac{n(n-1)\cdots(n-k+1)}{n^k}\right)$$
$$= 2M \left(1 - \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}\right) \longrightarrow 0, \quad n \to +\infty$$

αφού όλα τα κλάσματα εντός της παρένθεσης τείνουν στο 1. Άρα:

$$\left| \int \prod_{j=1}^{k} \varphi_j d\mu_n^C - \mathbb{E}[\prod_{j=1}^{k} \frac{\varphi_j(X_1) + \cdots + \varphi_j(X_n)}{n} | L_n \in C] \right| \longrightarrow 0, \quad n \to +\infty$$

Η απόδειξη έχει σχεδόν τελειώσει. Οι $F_j(\nu) = \int \varphi_j d\nu$ για $\nu \in \mathcal{M}_1(\mathcal{S})$ είναι φραγμένες και συνεχείς για όλα τα $j = 1, ..., k$, άρα και η $F = \prod\limits_{j=1}^{k} F_j$ είναι φραγμένη και συνεχής, δηλαδή $F \in C_b(\mathcal{M}_1(\mathcal{S}^k))$. Οπότε:

$$\begin{aligned}
\mathbb{E}[\prod_{j=1}^{k} \frac{\varphi_j(X_1) + \cdots + \varphi_j(X_n)}{n} | L_n \in C] &= \mathbb{E}[\langle \varphi_1, L_n \rangle \cdots \langle \varphi_k, L_n \rangle | L_n \in C] \\
&= \mathbb{E}[F_1(L_n) \cdots F_k(L_n) | L_n \in C] \\
&= \mathbb{E}[F(L_n) | L_n \in C] \\
&= \int F d\rho_n^C \longrightarrow \int F d\delta_{\nu_*} \\
&= F(\nu_*) \\
&= \prod_{j=1}^{k} \int \varphi_j d\nu_* \\
&= \int_{\mathcal{S}} \varphi_1(y_1) d\nu_*(y_1) \cdots \int_{\mathcal{S}} \varphi_k(y_k) d\nu_*(y_k) \\
&= \int_{\mathcal{S}^k} \varphi_1(y_1) \cdots \varphi_k(y_k) d\nu_*(y_1) \cdots d\nu_*(y_k) \\
&= \int_{\mathcal{S}^k} \prod_{j=1}^{k} \varphi_j d\nu_*^k
\end{aligned}$$

Και άρα:

$$\left| \int_{\mathcal{S}^k} \prod_{j=1}^{k} \varphi_j d\mu_n^C - \int_{\mathcal{S}^k} \prod_{j=1}^{k} \varphi_j d\nu_*^k \right| \longrightarrow 0, \quad n \to +\infty$$

Δείξαμε δηλαδή ότι: $\mu_n^C \Rightarrow \nu_*^k$. □

Συνεπώς, εάν $\rho_n^C \Rightarrow \delta_{\nu_*}$ τότε υπό τον περιορισμό $L_n \in C$ οι $X_1, ..., X_k$ αλλάζουν αποκτώντας κατανομή ν_* και παραμένοντας ανεξάρτητες. Μπορούμε να συνδυάσουμε τις παραπάνω δύο Προτάσεις σε ένα Θεώρημα:

Θεώρημα 17 (Η Αρχή της Μέγιστης Εντροπίας). *Αν $C \subset \mathcal{M}_1(\mathcal{S})$ είναι ένα κλειστό, κυρτό σύνολο τέτοιο ώστε:*

$$\inf_{\nu \in C} H(\nu|\mu) = \inf_{\nu \in C^o} H(\nu|\mu) < +\infty$$

τότε υπάρχει μοναδικό ν_* που ελαχιστοποιεί την $H(\nu|\mu)$ στο C και:

$$\mu_n^C(\cdot) = \mathbb{P}[(X_1, ..., X_k) \in \cdot | L_n \in C] \Rightarrow \nu_*^k, \quad \forall k \in \mathbb{N}$$

Η Αρχή της Μέγιστης Εντροπίας έχει διαισθητική σημασία στη στατιστική. Έστω πως έχουμε την πεποίθηση, ότι μια τυχαία μεταβλητή X ακολουθεί κατανομή μ. Όμως, έπειτα από κάποιο πείραμα ανακαλύπτουμε ότι η μ δεν είναι πλέον η πιθανότερη κατανομή της X, δηλαδή μπορεί να υπάρχει κάποια άλλη κατανομή που πιθανώς να την περιγράφει καλύτερα. Ποια πρέπει λοιπόν να είναι η νέα βέλτιστη μαντεψιά μας για την κατανομή της X απ' όλες τις ταιριαστές κατανομές που βρίσκονται στο C; Την απάντηση μας τη δίνει το παραπάνω Θεώρημα (η αρχή της μέγιστης εντροπίας): Η νέα βέλτιστη μαντεψιά μας είναι η "κοντινότερη" στη μ με την έννοια της σχετικής εντροπίας. Δηλαδή θα είναι η κατανομή εκείνη που ελαχιστοποιεί τη σχετική εντροπία ως προς την αρχική κατανομή μ.

3.1.1 Η αρχή του Gibbs υπό τη συνθήκη του Cramér

Θα παρουσιάσουμε τώρα ένα σημαντικό Πόρισμα της αρχής της μέγιστης εντροπίας. Όμως πρώτα θα χρειαστούμε μια Πρόταση, η οποία επιλύει το πρόβλημα ελαχιστοποίησης της σχετικής εντροπίας υπό περιορισμούς.

Πρόταση 17.1. *Έστω $g \in C_b(\mathcal{S})$ και $z \in \mathbb{R}$. Τότε το πρόβλημα:*

$$\inf_{\nu} H(\nu|\mu)$$

υπό τους περιορισμούς:

$$\int d\nu = 1, \quad \int g d\nu = z$$

Έχει μοναδική λύση το μέτρο Gibbs:

$$\frac{d\nu_{\beta_*}}{d\mu} = \frac{e^{-\beta_* g}}{Z_{\beta_*,g}} \tag{3.2}$$

όπου β_* η μοναδική τιμή του β τέτοια ώστε:

$$\int g d\nu_{\beta_*} = z$$

Και $Z_{\beta_*,g} = \mathbb{E}[e^{-\beta_* g}] < \infty$

Απόδειξη. Στην ουσία ψάχνουμε το ελάχιστο της $H(\cdot|\mu)$ στο σύνολο:

$$M_0 = \{\nu \in \mathcal{M}_1(\mathcal{S}) : \int g d\nu = z\}$$

Το οποίο είναι προφανώς κλειστό. Άρα, αν βρούμε ένα ελάχιστο, τότε αυτό θα είναι μοναδικό, όπως αποδείξαμε νωρίτερα με τον νόμο παραλληλογράμμου του Csiszár.

Φυσικά, έχει νόημα να αναζητάμε ένα $\nu \in \mathcal{M}_1(\mathcal{S})$ ώστε $\nu \ll \mu$. Έστω $f = \dfrac{d\nu}{d\mu}$ η παράγωγος Radon-Nikodym. Το πρόβλημα είναι ισοδύναμο με το:

$$\inf_f \int f \log f d\mu$$

Υπό τους περιορισμούς:

$$\int f d\mu = 1, \quad \int f g d\mu = z$$

Για να το επιλύσουμε, θα χρησιμοποιήσουμε λογισμό των μεταβολών σε άπειρες διαστάσεις (βλ. [17]). Σχηματίζουμε αρχικά την Λαγκραντζιανή:

$$\mathcal{L}(f) = \int f \log f d\mu + \lambda_1 \Big(\int f g d\mu - z \Big) + \lambda_2 \Big(\int f d\mu - 1 \Big)$$

και σχηματίζουμε την παράγωγο Gateaux της $\mathcal{L}$ θεωρώντας h τυχαία μετρήσιμη συνάρτηση στον $\mathcal{S}$ και $\varepsilon > 0$:

$$\mathcal{L}(f + \varepsilon h) = \int (f + \varepsilon h) \log(f + \varepsilon h) d\mu + \lambda_1 \Big(\int (f + \varepsilon h) g d\mu - z \Big) + \lambda_2 \Big(\int (f + \varepsilon h) d\mu - 1 \Big)$$
$$= \int f \log(f + \varepsilon h) d\mu + \varepsilon \int h \log(f + \varepsilon h) d\mu + \lambda_1 \Big(\int f g d\mu - z \Big) + \varepsilon \lambda_1 \int h g d\mu + \varepsilon \lambda_2 \int h d\mu$$
$$+ \lambda_2 \Big(\int f d\mu - 1 \Big)$$

Άρα,

$$\begin{aligned}
\mathcal{L}(f + \varepsilon h) - \mathcal{L}(f) &= \int f \log(f + \varepsilon h) d\mu + \varepsilon \int h \log(f + \varepsilon h) d\mu + \varepsilon \lambda_1 \int h g d\mu + \varepsilon \lambda_2 \int h d\mu \\
&\quad - \int f \log f d\mu \\
&= \int f \log \left(1 + \varepsilon \frac{h}{f} \right) d\mu + \varepsilon \int h \log(f + \varepsilon h) d\mu + \varepsilon \lambda_1 \int h g d\mu + \varepsilon \lambda_2 \int h d\mu
\end{aligned}$$

Για $\varepsilon > 0$ πολύ μικρό θα έχουμε τα αναπτύγματα:

$$\log \left(1 + \varepsilon \frac{h}{f} \right) = \varepsilon \frac{h}{f} - \varepsilon^2 \frac{h^2}{f^2} + O(\varepsilon^3)$$
$$\log(f + \varepsilon h) = \log f + \varepsilon h \frac{1}{f} + O(\varepsilon^2)$$

Οπότε αντικαθιστώντας τα αναπτύγματα και διαιρώντας με $\varepsilon > 0$ παίρνουμε:

$$\begin{aligned}
\frac{\mathcal{L}(f + \varepsilon h) - \mathcal{L}(f)}{\varepsilon} &= \int \left(h - \varepsilon \frac{h^2}{f} + h \log f + \varepsilon \frac{h^2}{f} + \lambda_1 h g + \lambda_2 h + O(\varepsilon^2) \right) d\mu \\
&= \int (1 + \log f + \lambda_1 g + \lambda_2 + O(\varepsilon^2)) h d\mu
\end{aligned}$$

Οπότε ζητάμε την f, ώστε:

$$\lim_{\varepsilon \to 0} \frac{\mathcal{L}(f+\varepsilon h) - \mathcal{L}(f)}{\varepsilon} = 0 \Leftrightarrow \int (1 + \log f + \lambda_1 g + \lambda_2) h d\mu = 0, \quad \forall h$$

Για να ισχύει αυτό για όλες τις μετρήσιμες h στον $\mathcal{S}$ θα πρέπει από γνωστό θεώρημα του λογισμού μεταβολών να ισχύει:

$$1 + \log f + \lambda_1 g + \lambda_2 = 0 \Leftrightarrow f = e^{-\lambda_1 g - \lambda_2 - 1}$$

Λόγω των περιορισμών παίρνουμε:

$$\int f d\mu = 1 \Leftrightarrow \int e^{-\lambda_1 g} d\mu = e^{\lambda_1 + 1}$$

Οπότε:

$$f = \frac{e^{-\lambda_1 g}}{\int e^{-\lambda_1 g} d\mu}$$

Ονομάζουμε $\beta = \lambda_1$ και τότε η f ανήκει στην οικογένεια των κατανομών Gibbs που ορίζονται ως:

$$\frac{d\nu}{d\mu}(x) = \frac{e^{-\beta g(x)}}{Z_{\beta,g}} \tag{3.3}$$

όπου το $Z_{\beta,g} = \int f d\mu = \mathbb{E}^{\mu}[e^{-\beta g}]$.

Εμείς αναζητούμε το β ώστε $\int g d\nu_\beta = z$. Επειδή όμως:

$$H(\nu_\beta|\mu) = \int \log \frac{d\nu_\beta}{d\mu} d\nu_\beta = \int \log \frac{e^{-\beta g}}{Z_{\beta,g}} d\nu_\beta = -\beta \int g d\nu_\beta - \log Z_{\beta,g} < +\infty$$

τότε: $\inf_{\nu \in A_0} H(\nu|\mu) < +\infty$ και άρα όπως είδαμε ο νόμος παραλληλογράμμου του Csiszár μας εξασφαλίζει τη μοναδικότητα του ν_β και άρα του β, αφού τα μέτρα Gibbs χαρακτηρίζονται από αυτή την παράμετρο. Αυτό το μοναδικό β είναι που αποκαλούμε β_*. □

Μια δημοφιλής συνέπεια της παραπάνω Πρότασης είναι η ακόλουθη πρόταση, που συμπυκνώνει την ιδέα ότι εάν επιβάλλουμε τον περιορισμό η L_n να ανήκει σε κάποιο "σπάνιο" σύνολο, τότε η κατανομή του $(X_1, ..., X_k)$ θα τείνει στην κατανομή εκείνη που ελαχιστοποιεί τη σχετική εντροπία ως προς την αρχική κατανομή, ενώ οι μεταβλητές θα παραμείνουν ανεξάρτητες. Η κατανομή αυτή είναι ακριβώς η κατανομή του Gibbs που λύνει το πρόβλημα ελαχιστοποίησης της παραπάνω Πρότασης

Πρόταση 17.2 (Η Αρχή του Gibbs). *Έστω S πολωνικός χώρος, $f \in C_b(S), z \in \mathbb{R}$ και $\delta > 0$. Για κάθε $\delta > 0$ ορίζουμε τα σύνολα M_δ ως εξής:*

$$M_\delta = \{\nu \in \mathcal{M}_1(S) : \left| \int f d\nu - z \right| \leq \delta\}$$

Έστω επίσης μια ακολουθία από ανεξάρτητες, ισόνομες τυχαίες μεταβλητές $\{X_n\}_{n\in\mathbb{N}}$ στον S με κατανομή μ. Αν $k \in \mathbb{N}$, συμβολίζω με $\nu^k_{\delta,n}$ τη δεσμευμένη κατανομή του διανύσματος $(X_1, ..., X_k)$ δεδομένου ότι $L_n \in M_\delta$. Οπότε:

$$\nu^k_{\delta,n}(A) = \mathbb{P}[(X_1, ..., X_k) \in A | L_n \in M_\delta], \quad \forall A \in \mathcal{B}(S^k)$$

Τότε:

$$\lim_{\delta\to 0}\lim_{n\to+\infty} \nu^k_{\delta,n} = \nu^k_{\beta_*}$$

όπου το ν_β είναι το μέτρο Gibbs που ορίζεται ως:

$$\frac{d\nu_\beta}{d\mu} = \frac{e^{-\beta f}}{Z_\beta} \tag{3.4}$$

όπου $Z_\beta = \mathbb{E}^\mu[e^{-\beta f}] < +\infty$ και β_* είναι η μοναδική τιμή του β, για την οποία ισχύει ότι $\int f d\nu_\beta = z$

Απόδειξη. Για $\nu \in \mathcal{M}_1(S)$ θεωρούμε τη συνάρτηση $F(\nu) = \int f d\nu$. Για να εφαρμόσουμε την αρχή της μέγιστης εντροπίας, πρέπει να δείξουμε αρχικά ότι τα M_δ είναι κυρτά και κλειστά. Η κυρτότητα είναι εύκολη, καθώς η F είναι μια γραμμική συνάρτηση. Πράγματι, εάν $\nu_1, \nu_2 \in M_\delta$ και $a \in [0, 1]$, τότε:

$$\begin{aligned}
|F(a\nu_1 + (1-a)\nu_2) - z| &= |aF(\nu_1) + (1-a)F(\nu_2) - az - (1-a)z| \\
&= |a(F(\nu_1) - z) + (1-a)(F(\nu_2) - z)| \\
&\le a|F(\nu_1) - z| + (1-a)|F(\nu_2) - z| \\
&\le a\delta + (1-a)\delta = \delta
\end{aligned}$$

Άρα $a\nu_1 + (1-a)\nu_2 \in M_\delta$, οπότε τα M_δ είναι κύρτα.

Η κλειστότητα επίσης δεν είναι δύσκολη, εάν θυμηθούμε τη σχέση μεταξύ της ασθενούς τοπολογίας και της ασθενούς σύγκλισης μέτρων πιθανότητας. Συγκεκριμένα, η F είναι συνεχής συνάρτηση, αφού έτσι έχουμε ορίσει την σύγκλιση στην ασθενή τοπολογία. Οπότε τα M_δ είναι κλειστά ως αντίστροφη εικόνα κλειστού συνόλου μέσω συνεχούς συνάρτησης. Τέλος, απομένει να αποδείξουμε ότι:

$$\inf_{\nu\in M_\delta} H(\nu|\mu) < \infty \tag{3.5}$$

Όμως εάν θεωρήσουμε το σύνολο M_0, που ορίζεται ως:

$$M_0 = \{\nu \in \mathcal{M}_1(S) : \int f d\nu = z\}$$

Τότε $\forall\delta > 0 : M_0 \subseteq M_\delta$ και $\nu_{\beta_*} \in M_0$, άρα $\forall\delta > 0 : \nu_{\beta_*} \in M_\delta$. Όμως λόγω της συνθήκης του Cramér έχουμε ότι $H(\nu_{\beta_*}|\mu) = -\beta_* z - \log Z_{\beta_*} < \infty$, οπότε ικανοποιείται η (2).

Άρα ικανοποιούνται όλες οι υποθέσεις της αρχής της μέγιστης εντροπίας, επομένως για κάθε $\delta > 0$ υπάρχει μοναδικό μέτρο πιθανότητας $\nu^k_{\delta,*}$, το οποίο να ελαχιστοποιεί τη σχετική εντροπία $H(\nu|\mu)$ στο M_δ και να ισχύει ότι:

$$\lim_{n\to+\infty} \nu^k_{\delta,n} = \nu^k_{\delta,*}$$

Οπότε θα έχουμε ότι:

$$\lim_{\delta\to 0} \lim_{n\to+\infty} \nu^k_{\delta,n} = \lim_{\delta\to 0} \nu^k_{\delta,*} = \nu^k_{\beta_*}$$

όπου το $\nu^k_{\beta_*}$ ελαχιστοποιεί τη σχετική εντροπία ως προς μ στο σύνολο M_0. Αυτό όμως το μέτρο το έχουμε υπολογίσει σε προηγούμενη Πρόταση και δεν είναι τίποτε άλλο από το μέτρο Gibbs (1) με το β_* να είναι όντως η μοναδική τιμή του β που ικανοποιεί τον περιορισμό $\int f d\nu_\beta = z$. $\square$

Συνεπώς, εάν επιβάλλουμε τον περιορισμό $\left|\dfrac{g(X_1)+\cdots+g(X_n)}{n} - z\right| \leq \delta$, τότε η κατανομή του $X^k = (X_1, ..., X_k)$ καθώς $n \to +\infty$ και $\delta \to 0$ θα τείνει στην:

$$\nu^k_{\beta_*} = \frac{1}{Z^k_{\beta_*,g}} \int e^{\beta_*(g(x_1)+\cdots+g(x_k))} d\mu^k(\vec{x})$$

Πριν κλείσουμε την ενότητα αυτή, ας σημειώσουμε δύο πράγματα. Πρώτον, στις παραπάνω αποδείξεις μπορεί κανείς να καταλάβει το σκεπτικό υπό το οποίο επιλέξαμε το νέο μέτρο στο θεώρημα του Cramér κατά τον τρόπο αυτό: Ήταν το μέτρο Gibbs που ελαχιστοποιούσε τη σχετική εντροπία υπό τον περιορισμό οι τυχαίες μεταβλητές να ικανοποιούν κάποιου είδους τυπική συμπεριφορά, όπως το Κεντρικό Οριακό Θεώρημα. Δεύτερον, είναι σημαντικό να επαναλάβουμε ότι η Αρχή του Gibbs, όπως τη διατυπώσαμε, ικανοποιείται για κατανομές που έχουν πεπερασμένη ροπογεννήτρια, δηλαδή που ικανοποιούν τη συνθήκη του Cramér. Τι γίνεται αν οι τυχαίες μεταβλητές ακολουθούν κάποια κατανομή της οποίας η ροπογεννήτρια δεν ορίζεται ή είναι άπειρη; Το κομμάτι αυτό δεν είναι τετριμμένο και θα δούμε πως θεραπεύεται στο επόμενο κεφάλαιο.

Ένα σημαντικό σχόλιο που πρέπει να κάνουμε είναι να εξετάσουμε το τι συμβαίνει για $g(x) = x$, δηλαδή εάν θέλουμε να μελετήσουμε τον μέσο όρο $\dfrac{S_n}{n}$. Η απάντηση είναι ότι η συνάρτηση $\nu \mapsto \int x d\nu(x)$ δεν είναι συνεχής, εάν εφοδιάσουμε τον $\mathcal{M}_1$ με την ασθενή τοπολογία, αφού η $g(x) = x$ δεν είναι φραγμένη. Αυτό στη γλώσσα των πιθανοτήτων σημαίνει ότι **δεν** ισχύει κατ' ανάγκη η συνεπαγωγή:

$$X_n \xrightarrow{\mathrm{d}} X \Rightarrow \mathbb{E}[X_n] \to \mathbb{E}[X]$$

Η τοπολογία, λοιπόν, που θα κάνει την εν λόγω συνάρτηση συνεχή είναι η Wasserstein-1 και είναι ισχυρότερη της ασθενούς τοπολογίας. Για να την ορίσουμε, εφοδιάζουμε τον $\mathcal{M}_1(\mathcal{S})$ με τη μετρική:

$$W_1(\mu,\nu) = \inf_{\pi\in\Gamma(\mu,\nu)} \int_{\mathcal{S}\times\mathcal{S}} d(x,y) d\pi(x,y)$$

όπου $d(x,y)$ είναι η μετρική που επάγει την τοπολογία του πολωνικού χώρου $\mathcal{S}$ και $\Gamma(\mu,\nu) = \{\pi \in \mathcal{M}_1(\mathcal{S}\times\mathcal{S}) : \pi(A\times\mathcal{S}) = \mu(A), \pi(\mathcal{S}\times A) = \nu(A), \forall A \in \mathcal{B}(\mathcal{S})\}$, δηλαδή ο $\Gamma(\mu,\nu)$ είναι ο χώρος των μέτρων πιθανότητας που ανήκουν στον χώρο γινόμενο $\mathcal{S}\times\mathcal{S}$ και έχουν ως περιθώριες τις μ,ν. Η μετρική αυτή επάγει την τοπολογία Wasserstein-1 και, με αυτήν την τοπολογία, η απεικόνιση $\nu \mapsto \int x d\nu(x)$ είναι συνεχής (για περισσότερες λεπτομέρειες μπορεί κανείς να συμβουλευτεί το κεφάλαιο 6 στο [5]).

Κλείνουμε αυτή την παράγραφο κάνοντας μια αναφορά στην αρχή του Gibbs για την περίπτωση που το k μεγαλώνει με το n. Συγκεκριμένα, παραπάνω αποδείξαμε την αρχή του Gibbs για την περίπτωση όπου το $k \in \mathbb{N}$, δηλαδή το μέγεθος του δείγματος, ήταν σταθερό στο όριο καθώς $n \to +\infty$. Ποια είναι η επίδραση στην αρχή του Gibbs σε υποσύνολα του συστήματός μας, των οποίων το μέγεθος μεγαλώνει όσο αυξάνεται το μέγεθος του συστήματος, δηλαδή όταν: $k = k(n) \to +\infty$, όσο $n \to +\infty$; Την απάντηση θα μας την δώσει μια σημαντική συνέπεια του παρακάτω θεωρήματος (7.3.21 στο [15]):

Θεώρημα 18. *Εάν έχουμε ένα $C \subset \mathcal{M}_1(\mathcal{S})$ κλειστό, κυρτό σύνολο για το οποίο ισχύει η γνωστή συνθήκη:*

$$\inf_{\nu\in C} H(\nu|\mu) = \inf_{\nu\in C^o} H(\nu|\mu) = +\infty$$

Τότε υπό την υπόθεση ότι:

$$\rho_n(C) e^{n \inf_{\nu\in C} H(\nu|\mu)} \geq g_n > 0$$

για κάθε $k = k(n)$ θα έχουμε:

$$H(\mu_n^C|\nu_*^k) \leq \frac{1}{\lfloor \frac{n}{k(n)} \rfloor} \log\left(\frac{1}{g_n}\right) \tag{3.6}$$

όπου:

$$\mu_n^C(\cdot) = \mathbb{P}[(X_1,...,X_{k(n)}) \in \cdot | L_n \in C]$$

και ν_* είναι το μέτρο στο C, στο οποίο πιάνει το ελάχιστο η σχετική εντροπία.

Για να διατυπώσουμε και να αποδείξουμε τη σημαντική συνέπεια της (3.6) πρέπει πρώτα να εισάγουμε την ανισότητα του Pinsker, που θα τη χρειαστούμε και παρακάτω.

Πρόταση 18.1 (Ανισότητα του Pinsker).

$$||\nu-\mu||_{t.v}^2 \leq \frac{1}{2} H(\nu|\mu), \quad \forall \nu,\mu \in \mathcal{M}_1(\mathcal{S}^k) \tag{3.7}$$

όπου $||\nu||_{t.v} = \sup_{A\in\mathcal{B}(\mathcal{S}^k)} |\nu(A)|$ είναι η total variation νόρμα.

Απόδειξη. Έστω $A \in \mathcal{B}(\mathcal{S}^k)$ και χωρίς βλάβη της γενικότητας θεωρούμε ότι $H(\nu|\mu) < +\infty$. Αν $f = \dfrac{d\nu}{d\mu}$, τότε: $\nu(A) = \int_A f d\mu$. Οπότε εκτιμούμε τη διαφορά:

$$\begin{aligned}
|\nu(A) - \mu(A)| &= \left|\int_A f d\mu - \int_A d\mu\right| \\
&= \left|\int_A (f-1) d\mu\right| \\
&= \left|\int_A (f-1)\mathbb{1}_{\{f\geq 1\}} d\mu + \int_A (f-1)\mathbb{1}_{\{f\leq 1\}} d\mu\right| \\
&= \left|\int_{A\cap\{f\geq 1\}} (f-1) d\mu - \int_{A\cap\{f\leq 1\}} (1-f) d\mu\right| \\
&\leq \max\{\int_{A\cap\{f\geq 1\}} (f-1) d\mu, \int_{A\cap\{f\leq 1\}} (1-f) d\mu\} \\
&\leq \max\{\int_{\{f\geq 1\}} (f-1) d\mu, \int_{\{f\leq 1\}} (1-f) d\mu\} \\
&= \int_{\{f\geq 1\}} (f-1) d\mu \\
&= \int_{\{f\geq 1\}} f d\mu - \int_{\{f\geq 1\}} d\mu \\
&= \nu(\{f \geq 1\}) - \mu(\{f \geq 1\})
\end{aligned}$$

Άρα,

$$||\nu - \mu||_{t.v.} \triangleq \sup_{A\in\mathcal{B}(\mathcal{S}^k)} |\nu(A) - \mu(A)| \leq \nu(\{f \geq 1\}) - \mu(\{f \geq 1\})$$

Η ανισότητα είναι ισότητα, αφού $\{f \geq 1\} \in \mathcal{B}(\mathcal{S}^k)$ (καθώς η f είναι μετρήσιμη). Άρα αρκεί να δείξουμε ότι:

$$|\nu(\{f \geq 1\}) - \mu(\{f \geq 1\})|^2 \leq \frac{1}{2} H(\nu|\mu)$$

Γι'αυτό, θέτουμε $B = \{f \geq 1\}$. Τότε, αφού $\lambda\nu(B) = \int \lambda\mathbb{1}_B d\nu$, για $\lambda > 0$, από τη γνωσή ανισότητα που ικανοποιεί η σχετική εντροπία (ανισότητα της εντροπίας) θα πάρουμε:

$$\int \lambda\mathbb{1}_B d\nu \leq H(\nu|\mu) + \log\int e^{\lambda\mathbb{1}_B} d\mu$$

Οπότε,

$$\int \mathbb{1}_B d\nu \leq \frac{1}{\lambda} H(\nu|\mu) + \frac{1}{\lambda}\log\{e^\lambda\mu(B) + 1\cdot\mu(B^c)\}$$

Άρα,

$$\nu(B) \leq \frac{1}{\lambda} H(\nu|\mu) + \frac{1}{\lambda}\log\{e^\lambda\mu(B) + \mu(B^c)\}$$

Έτσι παίρνουμε,

$$\begin{aligned}
\nu(B) - \mu(B) &\leq \frac{1}{\lambda} H(\nu|\mu) + \frac{1}{\lambda}\log\{e^\lambda\mu(B) + \mu(B^C)\} - \mu(B) \\
&= \frac{1}{\lambda} H(\nu|\mu) + \frac{1}{\lambda}\Big(\log\{e^\lambda\mu(B) + \mu(B^c)\} - \lambda\mu(B)\Big)
\end{aligned}$$

Όμως, έστω η $h(\lambda) = \log\{e^{\lambda}\mu(B) + \mu(B^c)\} - \lambda\mu(B)$. Τότε,

$$\nu(B) - \mu(B) \leq \frac{1}{\lambda} H(\nu|\mu) + \frac{1}{\lambda} h(\lambda), \quad \forall \lambda > 0$$

Αλλά, ισχύουν τα εξής:

$$\begin{aligned} h(0) &= 0 \\ h'(\lambda) &= \frac{e^{\lambda}\mu(B)}{e^{\lambda}\mu(B) + \mu(B^c)} - \mu(B) \Rightarrow h'(0) = 0 \\ h''(\lambda) &= \frac{e^{\lambda}\mu(B)\mu(B^c)}{(e^{\lambda}\mu(B) + \mu(B^c))^2} \leq \frac{1}{4}, \quad \left(\text{αφού } \frac{\alpha\beta}{(\alpha+\beta)^2} \leq \frac{1}{4}\right) \end{aligned}$$

Οπότε,

$$h(\lambda) = \int_0^{\lambda}\int_0^{u} h''(t)dtdu \leq \int_0^{\lambda}\int_0^{u} \frac{1}{4}dtdu = \int_0^{\lambda} \frac{u}{4}du = \frac{\lambda^2}{8}$$

Έτσι λοιπόν,

$$\nu(B) - \mu(B) \leq \frac{1}{\lambda} H(\nu|\mu) + \frac{\lambda}{8}, \quad \forall \lambda > 0 \Rightarrow \nu(B) - \mu(B) \leq \min_{\lambda > 0} g(\lambda)$$

όπου:

$$g(\lambda) = \frac{1}{\lambda} H(\nu|\mu) + \frac{\lambda}{8} \Rightarrow g'(\lambda) = -\frac{1}{\lambda^2} H(\nu|\mu) + \frac{1}{8} = 0$$

Αφου $H(\nu|\mu) \geq 0$ και $\lambda > 0$ θα πάρουμε: $\lambda = \sqrt{8H(\nu|\mu)}$. Τέλος, αφού:

$$g''(\lambda) = \frac{2H(\nu|\mu)}{\lambda^3} > 0, \quad \forall \lambda > 0$$

Τότε το $\lambda = \sqrt{8H(\nu|\mu)}$ είναι ολικό ελάχιστο της g και συνεπώς:

$$\min_{\lambda > 0} g(\lambda) = \frac{1}{\sqrt{8H(\nu|\mu)}} + \frac{\sqrt{8H(\nu|\mu)}}{8} = \frac{1}{\sqrt{2}}\sqrt{H(\nu|\mu)}$$

Άρα,

$$\nu(B) - \mu(B) \leq \frac{1}{\sqrt{2}}\sqrt{H(\nu|\mu)} \Rightarrow ||\nu - \mu||_{t.v} \leq \frac{1}{\sqrt{2}}\sqrt{H(\nu|\mu)}$$

Οπότε,

$$||\nu - \mu||_{t.v}^2 \leq \frac{1}{2} H(\nu|\mu)$$

□

Η σημαντική συνέπεια δίνεται στην παρακάτω Πρόταση:

Πρόταση 18.2 (Η αρχή του Gibbs για k=k(n)). *Εάν* $\lim_{n\to+\infty} \frac{1}{n} k(n) \log\left(\frac{1}{g_n}\right) = 0$, *τότε:*

$$\lim_{n\to+\infty} ||\mu_n^C - \nu_*^{k(n)}||_{t.v.} = 0$$

Όπου: $\mu_n^C(\cdot) = \mathbb{P}[(X_1, ..., X_{k(n)}) \in \cdot | L_n \in C]$

Απόδειξη. Η ανισότητα του Pinsker για $\nu = \mu_n^C$ και $\mu = \nu_*^{k(n)}$ θα μας δώσει:

$$||\mu_n^C - \nu_*^{k(n)}||_{t.v.}^2 \leq \frac{1}{2} H(\mu_n^C | \nu_*^k)$$

Όμως η (3.6) δίνει ότι:

$$H(\mu_n^C | \nu_*^k) \leq \frac{1}{\lfloor \frac{n}{k(n)} \rfloor} \log\left(\frac{1}{g_n}\right)$$

Δηλαδή συνολικά θα έχουμε:

$$||\mu_n^C - \nu_*^{k(n)}||_{t.v.}^2 \leq \frac{1}{2} \frac{1}{\lfloor \frac{n}{k(n)} \rfloor} \log\left(\frac{1}{g_n}\right)$$

Οπότε εάν ισχύει η συνθήκη $\lim_{n\to+\infty} \frac{1}{n} k(n) \log\left(\frac{1}{g_n}\right) = 0$ παίρνουμε το ζητούμενο:

$$\lim_{n\to+\infty} ||\mu_n^C - \nu_*^{k(n)}||_{t.v.} = 0$$

□

Ας παρατηρήσουμε ότι εάν $k(n) = O(n)$, η συνθήκη της παραπάνω Πρότασης δε θα ισχύει κατ' ανάγκη, Οπότε δεν είναι εν γένει σωστό το αποτέλεσμα της αρχής του Gibbs για $k(n) = O(n)$.

Κλείνοντας την ενότητα αυτή, ας προσέξουμε ότι εάν $k = k(n) \to +\infty$ με ρυθμό τέτοιον ώστε να ισχύει η συνθήκη της Πρότασης, τότε η ανεξαρτησία των $X_1, ..., X_{k(n)}$ δεν παύει να ισχύει καθώς $n \to +\infty$. Είναι μονάχα η κατανομή τους που αλλάζει.

3.2 Κατανομές με βαριές ουρές

Ήρθε η στιγμή να δούμε τι συμβαίνει όταν παραβιάζεται η συνθήκη του Cramér. Θεωρούμε ως συνήθως έναν χώρο πιθανότητας $(\Omega, \mathcal{F}, \mathbb{P})$. Έστω επίσης μια τυχαία μεταβλητή $X : \Omega \to \mathbb{R}$. Θα συμβολίζουμε με $\bar{F}(x)$ την πιθανότητα που η X αφήνει δεξιά του $x \in \mathbb{R}$, δηλαδή:

$$\bar{F}(x) = \mathbb{P}[X > x] = 1 - F(x)$$

όπου $F(x)$ είναι η συνάρτηση κατανομής πιθανότητας της X.

Θα αποκαλούμε την κατανομή της X ως *κατανομή με βαριά ούρα* (heavy-tailed) [18] εάν για $t > 0$:

$$\mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} dF(x) = +\infty$$

Δηλαδή οι θετικές εκθετικές ροπές τους είναι άπειρες. Προφανώς, για τέτοιες κατανομές δεν ικανοποιείται η συνθήκη του Cramér. Οπότε, πολλά από τα αποτελέσματα που αποδείξαμε (για παράδειγμα, η αρχή του Gibbs) παύουν πια να ισχύουν. Οπότε έχει νόημα να αναρωτηθούμε αν αλλάζει και με ποιον τρόπο η συμπεριφορά των μεγάλων αποκλίσεων στην περίπτωση αυτή. Συγκεκριμένα, θα μελετήσουμε την αρχή του Gibbs για μια κατηγορία κατανομών με βαριές ουρές που ονομάζονται υπο-εκθετικές. Όμως, ας δούμε πρώτα τι το ιδιαίτερο έχουν οι κατανομές με βαριές ουρές και τι συνέπεια έχει η παραβίαση της συνθήκης του Cramér.

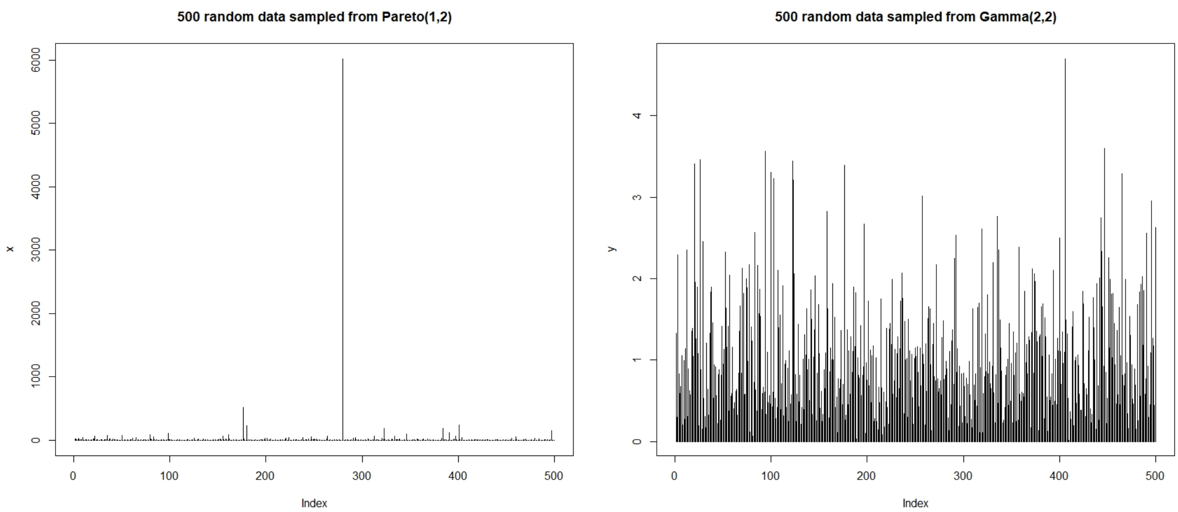

Figure 3.1: Αριστερά: Τυχαίο δείγμα μεγέθους 500 από την κατανομή Pareto με παραμέτρους 1 και 2. Δεξιά: Τυχαίο δείγμα μεγέθους 500 από την κατανομή Γάμμα με παραμέτρους 2 και 2

Η σημασία των κατανομών με βαριές ουρές έγκειται στο ότι βιώνουν ένα τυχαίο μεγάλο άλμα. Τι σημαίνει όμως αυτό; Στην εικόνα (3.1) φαίνονται δύο Διαγράμματα. Έχουμε κάνει τυχαία δειγματοληψία 500 σημείων από την κατανομή Pareto(1,2) και από την κατανομή Gamma(2,2). Η πρώτη κατανομή είναι με βαριά ουρά, ενώ η δεύτερη δεν έχει βαριά ουρά (είναι όπως λέμε "light-tailed", δηλαδή με "ελαφριά ουρά"). Η συμπεριφορά της πρώτης φαίνεται να είναι η εξής: σε ένα τυχαίο σημείο παίρνει μια πολύ μεγάλη τιμή (βιώνει ένα πολύ μεγάλο άλμα), ενώ συγκριτικά οι υπόλοιπες τιμές που παίρνει είναι αρκετά μικρές. Στη κατανομή Γάμμα φαίνεται ότι δεν υπάρχει η ίδια συμπεριφορά.

Αυτή ακριβώς είναι και η συμπεριφορά των κατανομών με βαριές ουρές. Συγκεκριμένα, εάν διαλέξουμε ένα τυχαίο δείγμα από κάποια κατανομή με τυχαία ουρά, τότε μπορεί κάποια στιγμή να εμφανιστεί μια τιμή τόσο μεγάλη που να αντιπροσωπεύει από μόνη της το μέγεθος του δείγματος. Ακριβώς το γεγονός αυτό είναι που περιγράφουν οι κατανομές αυτές. Για παράδειγμα, υπάρχει η δημοφιλής φράση ότι "η Pareto κατανομή μας δείχνει ότι το 80% του δείγματος το αντιπροσωπεύει το 20% των τιμών του". Άλλο παράδειγμα εμφάνισης κατανομών με βαριές ουρές είναι για την περιγραφή των οικονομικών ανισοτήτων στις Ηνωμένες Πολιτείες σήμερα [4].

Εν γένει, οι κατανομές με βαριές ουρές που χρησιμοποιούνται στις εφαρμογές είναι τριών κατηγοριών:

- *Κατανομές με Κανονικά Μεταβαλλόμενες Ουρές:*

  $$\bar{F}(x) \sim x^{-a}L(x), \quad a > 0$$

  όπου η $L(x)$ είναι μια κανονικά μεταβαλλόμενη συνάρτηση, δηλαδή

  $$\lim_{n\to+\infty} \frac{L(tx)}{L(x)} = t^b, \quad b > 0, \forall t > 0,$$

- *Κατανομές με ουρές τύπου Log-Normal:*

  $$\bar{F}(x) \sim x^{-\beta} e^{-a(\log x)^\gamma}, \quad \gamma > 1$$

- *Κατανομές με ουρές τύπου Weibull:*

  $$\bar{F}(x) \sim x^{-\gamma} e^{-ax^\lambda}, \quad \lambda < 1$$

Διάσημα παραδείγματα τέτοιων κατανομών είναι οι: Lognormal, Weibull, Pareto, Zipf, Laplace, Cauchy, Lévy και πολλές άλλες.

Στην παρούσα εργασία θα επικεντρωθούμε στις υπο-εκθετικές κατανομές με βαριές ουρές και θα δούμε πως αλλάζει η αρχή του Gibbs στην περίπτωση αυτή. Ένα καλό ερώτημα ως κίνητρο για τη μελέτη μας είναι το εξής: Η συμπεριφορά που βλέπουμε στην εικόνα (3.1) θα εξακολουθεί να ισχύει για την αλλαγμένη κατανομή των μεταβλητών, αν δεσμεύσουμε το άθροισμά τους σε μια μεγάλη απόκλιση;

3.2.1 Υποεκθετικές κατανομές

Έστω $\{X_n\}_{n\in\mathbb{N}}$ μη-αρνητικές, ανεξάρτητες, ισόνομες τυχαίες μεταβλητές με κοινή κατανομή μ σε έναν χώρο πιθανότητας $(\Omega, \mathcal{F}, \mathbb{P})$. Θα συμβολίζουμε με S_n το άθροισμα των n πρώτων τυχαίων μεταβλητών, δηλαδή: $S_n = X_1 + \ldots + X_n$. Υιοθετούμε επίσης τους συμβολισμούς:

$$\bar{F}(x) = \mathbb{P}[X_k > x], \quad \bar{F}_n(x) = \mathbb{P}[S_n > x]$$

Η κλάση των κατανομών με βαριές ουρές που θα ασχοληθούμε εδώ είναι οι *υποεκθετικές*. Συγκεκριμένα, η κατανομή των X_i είναι υποεκθετική, εάν:

$$\lim_{x\to+\infty} \frac{\bar{F}_n(x)}{n\bar{F}(x)} = 1, \quad \forall n \in \mathbb{N}, n \geq 2 \tag{3.8}$$

Η σχέση (3.8) μας εγγυάται την ύπαρξη ακολουθίας $\{d_n\}_{n\in\mathbb{N}}$, $d_n \to +\infty$, ώστε:

$$\lim_{n\to+\infty} \sup_{x\geq d_n} \left| \frac{\bar{F}_n(x)}{n\bar{F}(x)} - 1 \right| = 0 \tag{3.9}$$

Ας κάνουμε όμως λίγες πράξεις πρώτα για να καταλάβουμε τον τρόπο που λειτουργούν οι υποεκθετικές κατανομές. Συγκεκριμένα, για κάθε $n \geq 1$ κάνοντας απλές πράξεις μπορούμε να πάρουμε το εξής:

$$\begin{aligned}
\mathbb{P}[\max_{1\leq k\leq n} X_k > x] &= 1 - \mathbb{P}[\max_{1\leq k\leq n} X_k \leq x] \\
&= 1 - \mathbb{P}[X_1 \leq x, \ldots, X_n \leq x] \\
&= 1 - \mathbb{P}[X_1 \leq x]^n \\
&= 1 - F^n(x) \\
&= F^0(x) + F^1(x) + \ldots + F^{n-1}(x) \\
&- F^1(x) - \ldots - F^{n-1}(x) \\
&- F^n(x) \\
&= \sum_{k=0}^{n-1} F^k(x) - \sum_{k=1}^{n-1} F^{k+1}(x) \\
&= (1 - F(x)) \sum_{k=0}^{n-1} F^k(x) \\
&= \bar{F}(x) \sum_{k=1}^{n-1} F^k(x) \sim n\bar{F}(x), \quad \text{όσο } x \to +\infty
\end{aligned}$$

Άρα,

$$\lim_{x\to+\infty} \frac{\mathbb{P}[\max_{1\leq k\leq n} X_k > x]}{n\bar{F}(x)} = 1 \tag{3.10}$$

Συνδυάζοντας τις (3.8) και (3.10) παίρνουμε:

$$\lim_{n\to+\infty} \frac{\mathbb{P}[X_1 + \ldots + X_n > x]}{\mathbb{P}[\max_{1\leq k\leq n} X_k > x]} = 1 \tag{3.11}$$

Η διαισθητική ερμηνεία της (3.11) είναι ότι μεγάλες τιμές του αθροίσματος n μη-αρνητικών, ανεξάρτητων, μ-κατανεμημένων υποεκθετικών τυχαίων μεταβλητών τυπικά πραγματοποιείται με τη μέγιστη τιμή τους να παίρνει μια μεγάλη τιμή. Ισοδύναμα, η υπέρβαση κάποιου κατωφλιού του αθροίσματος μη-αρνητικών, υποεκθετικών μ-κατανεμημένων, τυχαίων μεταβλητών συμβαίνει εξαιτίας της υπέρβασης αυτού του κατωφλιού από τη μεγαλύτερη μεταβλητή στο δείγμα. Η εικόνα στα Διαγράμματα (3.1) επαληθεύεται. Ο λόγος που οι υποεκθετικές κατανομές αποτελούν από μόνες τους μια υποκατηγορία των κατανομών με βαριές ουρές είναι επειδή για κάθε $t > 0$,

$$e^{tx}\bar{F}(x) \to 0, \quad x \to +\infty$$

Δηλαδή η $\bar{F}(x)$ φθίνει στο 0 πιο αργά από οποιοδήποτε εκθετικό. Έτσι, δικαιώνεται και η ονομασία των υποεκθετικών κατανομών. Για την απόδειξη του αποτελέσματος αυτού παραπέμπουμε στο [2], συγκεκριμένα στο Λήμμα 1.3.5.

3.2.2 Η αρχή του Gibbs υπό την παραβίαση της συνθήκης του Cramér

Ήρθε η ώρα να αποδείξουμε και μια μορφή της αρχής του Gibbs, όταν παραβιάζεται η συνθήκη του Cramér. Συγκεκριμένα, θα θεωρήσουμε ότι οι τυχαίες μεταβλητές ακολουθούν κάποια υποεκθετική κατανομή, αξιοποιώντας τα παραπάνω αποτελέσματα. Πρώτα όμως θα ξεκινήσουμε με μια Πρόταση:

Πρόταση 18.3. *Εάν m είναι ένα μέτρο πιθανότητας στον $(\Omega, \mathcal{F}, \mathbb{P})$ και $A \in \mathcal{F}$ ώστε $m[A] > 0$ και m_A είναι το μέτρο:*

$$m_A[B] = m[B|A], \quad \forall B \in \mathcal{F}$$

Τότε, το m_A λύνει το πρόβλημα ελαχιστοποίησης:

$$\min_{\nu[A]=1} H(\nu|m) \tag{3.12}$$

Απόδειξη. Έστω ένα μέτρο ν, ώστε $\nu[A] = 1$ και φυσικά $\nu \ll m$, ώστε $f = \dfrac{d\nu}{dm}$ να είναι η αντίστοιχη παράγωγος Radon-Nikodym. Εύκολα διαπιστώνει κανείς την ισχύ της ανισότητας:

$$f \log f \geq cf - e^{c-1}, \quad \forall c \in \mathbb{R}$$

Ολοκληρώνοντας και κάνοντας πράξεις,

$$\begin{aligned}
\int_A f \log f dm &\geq \int_A (cf - e^{c-1}) dm \\
&= c\int_A f dm - e^{c-1}\int_A dm \\
&= c\nu[A] - m[A]e^{c-1} \\
&= c - m[A]e^{c-1}
\end{aligned}$$

Όμως για τη συνάρτηση $g(c) = c - m[A]e^{c-1}$ έχουμε ότι,

$$\begin{aligned}
&g'(c) = 1 - m[A]e^{c-1} \Leftrightarrow c = 1 - \log m[A] \\
&g''(c) = -m[A]e^{c-1} < 0
\end{aligned}$$

Άρα βρήκαμε το μέγιστο της g. Βελτιστοποιώντας,

$$H(\nu|m) = \int_A f \log f dm \geq 1 - \log m[A] - m[A] \cdot \frac{1}{m[A]} = -\log m[A]$$

Όμως,

$$H(m_A|m) = \int_A \frac{1}{m[A]} \log\left(\frac{1}{m[A]}\right) dm = -\log m[A]$$

Και άρα, $H(\nu|m) \geq H(m_A|m)$ για όλα τα ν τέτοια ώστε $\nu[A] = 1$ και $H(\nu|m) = H(m_A|m) \Leftrightarrow \nu = m_A$, οπότε το m_A λύνει το πρόβλημα ελαχιστοποίησης (3.12) $\square$

Παρακάτω υιοθετούμε τον εξής συμβολισμό, ακολουθώντας το [12]:

Για ένα διάνυσμα $\vec{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ συμβολίζουμε σ^j τον τελεστή που εναλλάσσει την j-οστή και την τελευταία συνιστώσα του $\vec{x}$. Ακόμα, για $A \in \mathcal{B}(\mathbb{R}^n)$ ορίζουμε τα μέτρα πιθανότητας $\mu_{n,x}$ και ν_x ως εξής:

$$\mu_{n,x}[A] = \mathbb{P}[(X_1, ..., X_n) \in A | S_n > x], \quad \nu_x[A] = \mathbb{P}[X_i \in A | X_i > x] = \frac{\mu[A \bigcap (x, +\infty)]}{\bar{F}(x)}$$

Δηλαδή, $\mu_{n,x}$ είναι η δεσμευμένη κατανομή του τυχαίου διανύσματος $\vec{X} = (X_1, ..., X_n)$, δεδομένου ότι το άθροισμα των $\{X_i\}_{1 \leq i \leq n}$ ξεπερνάει το κατώφλι x. Η ν_x είναι η δεσμευμένη κατανομή μιας από της X_i, δεδομένου ότι αυτή ξεπερνάει κάποιο κατώφλι x. Επίσης, παρακάτω θα συμβολίζουμε ως $\vec{X}^{n-1} = (X_1, ..., X_{n-1})$.

Ακόμα, θεωρούμε τον τελεστή $\mathrm{T} : \bigcup_{n \in \mathbb{N}} \mathbb{R}^n \to \bigcup_{n \in \mathbb{N}} \mathbb{R}^n$ που ανταλλάσσει την τελευταία και τη μέγιστη τιμή μιας πεπερασμένης ακολουθίας. Ο τελεστής αυτός δεν είναι δύσκολο να οριστεί. Πράγματι, εάν $\vec{x} = (x_1, ..., x_n)$ είναι μια ακολουθία στον $\mathbb{R}^n$, τότε η k- οστή συνιστώσα του $\mathrm{T}\vec{x}$ θα δίνεται από τη σχέση:

$$\mathrm{T}(x_1, ..., x_n)_k = \begin{cases} \max\limits_{1 \leq i \leq n} x_i, & \text{εάν } k = n \\ x_n, & \text{εάν } x_k > \max\limits_{1 \leq i < k} x_i \text{ και } x_k = \max\limits_{k \leq i} x_i \\ x_k, & \text{διαφορετικά} \end{cases}$$

Έτσι, εάν $M_{\vec{x}}$ είναι η μέγιστη τιμή του διανύσματος $\vec{x}$ και $m_{\vec{x}}$ η θέση της, θα έχουμε ότι: $\mathrm{T}\vec{x} = \sigma^{m_{\vec{x}}} \vec{x}$.

Είμαστε έτοιμοι να διατυπώσουμε το βασικό θεώρημα,

Θεώρημα 19. *Έστω μ μια υποεκθετική κατανομή με $\operatorname{supp} \mu \subseteq \mathbb{R}_+$. Έστω επίσης μια ακολουθία $\{d_n\}_{n \in \mathbb{N}}$, με $d_n \to +\infty$, ώστε να ικανοποιείται η (3.9). Τότε,*

$$\lim_{y \to +\infty} \sup_{n \in \mathbb{N}} \sup_{x \geq d_n \vee y} ||\mu_{n,x} \circ \mathrm{T}^{-1} - (\mu^{n-1} \times \nu_x)||_{t.v} = 0 \tag{3.13}$$

Απόδειξη. Από την ταυτότητα του Pinsker (3.7):

$$H(\pi|\nu) \geq 2||\pi - \nu||_{t.v.}^2$$

Μπορούμε να πάρουμε:

$$H\left(\pi|\frac{\pi+\nu}{2}\right) \geq 2\left\|\pi - \frac{\pi+\nu}{2}\right\|_{t.v}^2 = \frac{1}{2}||\pi - \nu||_{t.v.}^2$$

καθώς και την παρακάτω σχέση,

$$H\left(\nu|\frac{\pi+\nu}{2}\right) \geq 2\left\|\nu - \frac{\pi+\nu}{2}\right\|_{t.v}^2 = \frac{1}{2}||\pi - \nu||_{t.v.}^2$$

Συνδυάζοντας τις παραπάνω με το νόμο παραλληλογράμμου του Csiszár (3.1):

$$H(\pi|m) + H(\nu|m) = 2H\left(\frac{\pi+\nu}{2}|m\right) + H\left(\pi|\frac{\pi+\nu}{2}\right) + H\left(\nu|\frac{\pi+\nu}{2}\right)$$

Παίρνουμε ότι:

$$H(\pi|m) + H(\nu|m) - 2H\left(\frac{\pi+\nu}{2}|m\right) \geq ||\pi - \nu||_{t.v.}^2 \tag{3.14}$$

Εάν το μέτρο πιθανότητας π λύνει το μεταβολικό πρόβλημα (3.12) και ν είναι ένα άλλο μέτρο πιθανότητας με $\nu[A] = 1$, τότε αφού:

$$\left(\frac{\pi+\nu}{2}\right)[A] = \frac{1}{2}(\pi[A] + \nu[A]) = 1$$

Θα έχουμε ότι:

$$H\left(\frac{\pi+\nu}{2}|m\right) \geq H(\pi|m) \Leftrightarrow H(\pi|m) - 2H\left(\frac{\pi+\nu}{2}|m\right) \leq -H(\pi|m)$$

Και άρα από την (3.14):

$$||\pi - \nu||_{t.v.}^2 \leq H(\nu|m) - H(\pi|m) \tag{3.15}$$

Εφαρμόζουμε την (3.15) για:

- $A = \{\sum_{i=1}^{n} X_i > x\} \in \mathcal{B}(\mathbb{R}^n)$
- $m = \mu^n$ $(m[A] = \mu^n[A] > 0)$
- $\pi = \mu_{n,x}$ $(\pi[A] = \mathbb{P}[\sum_{i=1}^{n} X_i > x| \sum_{i=1}^{n} X_i > x] = 1$ και όπως είδαμε λύνει το μεταβολικό πρόβλημα)
- $\nu = \mu_{n,x}^* = \frac{1}{n}\sum_{j=1}^{n} \sigma^j(\mu^{n-1} \times \nu_x)$

Επίσης ας παρατηρήσουμε ότι αφού $\operatorname{supp}\mu \subseteq \mathbb{R}_+$ και $\operatorname{supp}\nu_x = (x, +\infty)$, τότε $\mu^*_{n,x}[S_n > x] = 1$. Ακόμα, τα μέτρα $\mu_{n,x}$ και μ^n είναι απόλυτα συνεχή, αφού:

$$\mu_{n,x} = \int d\mu_{n,x} = \int_{\{\sum\limits_{i=1}^{n} x_i > x\}} \frac{1}{\bar{F}_n(x)} d\mu^n = \int \frac{\mathbb{1}\{\sum\limits_{i=1}^{n} x_i > x\}}{\bar{F}_n(x)} d\mu^n$$

Οπότε $\mu_{n,x} \ll \mu^n$ με πυκνότητα Radon-Nikodym: $\dfrac{d\mu_{n,x}}{d\mu^n} = \dfrac{\mathbb{1}\{\sum\limits_{i=1}^{n} x_i > x\}}{\bar{F}_n(x)}$. Ισχύει επίσης ότι $\mu^*_{n,x} \ll \mu^n$, αφού:

$$\mu^{n-1} \times \nu_x = \int d\mu^{n-1} d\nu_x = \int \frac{d\nu_x}{d\mu} d\mu^n \Leftrightarrow \frac{d(\mu^{n-1} \times \nu_x)}{d\mu} = \frac{d\nu_x}{d\mu}$$

Όμως ισχύει ακόμα ότι,

$$\nu_x = \int d\nu_x = \int_{\{\sum\limits_{i=1}^{n} x_i > x\}} \frac{1}{\bar{F}(x)} d\mu = \int \frac{\mathbb{1}_{\{\sum_{j=1}^{n} x_j > x\}}}{\bar{F}(x)} d\mu \Leftrightarrow \frac{d\nu_x}{d\mu} = \frac{\mathbb{1}_{\{\sum_{j=1}^{n} x_j > x\}}}{\bar{F}(x)}$$

Συνεπώς παίρνουμε,

$$\frac{d(\mu^{n-1} \times \nu_x)}{d\mu^n} = \frac{\mathbb{1}_{\{\sum_{j=1}^{n} x_j > x\}}}{\bar{F}(x)}$$

Ο τελεστής σ^j αλλάζει μονάχα τη θέση των κατανομών στο διάνυσμα $\mu^{n-1} \times \nu_x$. Όμως αυτό δε θα αλλάξει την παράγωγο Radon-Nikodym, αφού το διάνυσμα αυτό είναι εναλλάξιμο. Συνεπώς παίρνουμε ότι:

$$\frac{d\mu^*_{n,x}}{d\mu^n} = \frac{1}{n} \sum_{j=1}^{n} \frac{\mathbb{1}_{\{\sum_{j=1}^{n} x_j > x\}}}{\bar{F}(x)} = \frac{N_n(\vec{x})}{n\bar{F}(x)}$$

Όπου $N_n(\vec{x})$ είναι το πλήθος των συνιστωσών του $\vec{x}$, που είναι μεγαλύτερες του x. Συνδυάζοντας τις δύο παραπάνω παραγώγους Radon-Nikodym και την (3.15) παίρνουμε ότι:

$$\begin{aligned}
||\mu^*_{n,x} - \mu_{n,x}||^2_{t.v} &\leq H(\mu^*_{n,x}|\mu^n) - H(\mu_{n,x}|\mu^n) \\
&= \int \log\left(\frac{d\mu^*_{n,x}}{d\mu^n}\right) d\mu^*_{n,x} - \int \log\left(\frac{d\mu_{n,x}}{d\mu^n}\right) d\mu_{n,x} \\
&= \int \log\left(\frac{N_n(\vec{x})}{n\bar{F}(x)}\right) d\mu^*_{n,x} - \int_{\{\sum\limits_{i=1}^{n} x_i > x\}} \log\left(\frac{1}{\bar{F}_n(x)}\right) d\mu_{n,x}
\end{aligned}$$

$$
\begin{aligned}
&= -\log(n\bar{F}(x)) + \int \log N_n(\vec{x}) d\mu^*_{n,x} + \log \bar{F}_n(x) \\
&= \log\left(\frac{\bar{F}_n(x)}{n\bar{F}(x)}\right) + \int \log N_n(\vec{x}) d\mu^*_{n,x} \\
&= \log\left(\frac{\bar{F}_n(x)}{n\bar{F}(x)}\right) + \int \log N_n(\vec{x}) d\mu^{n-1} d\nu_x \\
&= \log\left(\frac{\bar{F}_n(x)}{n\bar{F}(x)}\right) + \int \log(1 + N_{n-1}(\vec{x}^{n-1})) d\mu^{n-1} \\
&\leq \log\left(\frac{\bar{F}_n(x)}{n\bar{F}(x)}\right) + \int N_{n-1}(\vec{x}^{n-1}) d\mu^{n-1} \\
&= \log\left(\frac{\bar{F}_n(x)}{n\bar{F}(x)}\right) + (n-1)\bar{F}(x)
\end{aligned}
$$

Όπου η τελευταία ισότητα προκύπτει από την πυκνότητα Radon-Nikodym,

$$\frac{d\mu^*_{n-1,x}}{d\mu^{n-1}} = \frac{N_{n-1}(\vec{x}^{n-1})}{(n-1)\bar{F}(x)}$$

που υπολογίζεται όπως προηγουμένως. Τώρα, είτε αφήνοντας το $x \to +\infty$ και χρησιμοποιώντας την (3.8), είτε κρατώντας $x \geq d_n$ (με $n\bar{F}(d_n) \to 0$), αφήνοντας $n \to +\infty$ και αξιοποιώντας την (3.9), παίρνουμε ότι:

$$||\mu_{n,x} - \mu^*_{n,x}||_{t.v} \to 0$$

$\square$

Ας αναλογιστούμε το παραπάνω αποτέλεσμα. Καταρχάς, είναι σαφές ότι αφού η κατανομή του διανύσματος $\vec{X} = (X_1, ..., X_n)$ είναι αναλλοίωτη σε μεταθέσεις υπό τον περιορισμό $\{S_n > x\}$, τότε η θέση της μέγιστης από αυτές είναι ομοιόμορφα κατανεμημένη στο $\{1, .., n\}$. Το θεώρημα αυτό μας λέει ότι η υπό συνθήκη κατανομή της μέγιστης από τις τυχαίες μεταβλητές γίνεται ασυμπτωτικά μία τυχαία τοποθετημένη ν_x, ενώ η κατανομή των $n-1$ υπόλοιπων μεταβλητών είναι το γινόμενο μ^{n-1}. Δηλαδή, οι υπόλοιπες μεταβλητές παραμένουν ανεξάρτητες με κατανομή που δεν έχει αλλάξει.

Άρα, η συνθήκη $\{S_n > x\}$ επηρεάζει μόνο τη μέγιστη συνιστώσα στο όριο, ενώ οι $n-1$ μικρότερες μεταβλητές γίνονται ασυμπτωτικά ανεξάρτητες. Το αποτέλεσμα αυτό είναι εντελώς διαφορετικό από αυτό που πήραμε στην αρχή του Gibbs, όταν δεν παραβιαζόταν η συνθήκη του Cramér! Πράγματι, όπως είδαμε ήδη, στην περίπτωση των πεπερασμένων εκθετικών ροπών οι τυχαίες μεταβλητές στο όριο παρέμεναν ανεξάρτητες και η κατανομή τους άλλαζε σε εκείνη που ελαχιστοποιούσε τη σχετική εντροπία ως προς τον εκάστοτε περιορισμό.

Τώρα, στην περίπτωση των υποεκθετικών κατανομών, η μέγιστη από τις τυχαίες μεταβλητές απορροφάει όλες τις συσχετίσεις που προκύπτουν από την επιβολή της συνθήκης $\{S_n > x\}$ και οι υπόλοιπες τυχαίες μεταβλητές παραμένουν ασυμπτωτικά ανεξάρτητες. Το αποτέλεσμα αυτό παρουσιάζει μια τυπική συμπεριφορά των υποεκθετικών κατανομών, όπως την περιγράψαμε παραπάνω. Συγκεκριμένα, από την (3.11) βλέπουμε ότι η μέγιστη από τις τυχαίες μεταβλητές καθορίζει την τιμή του αθροίσματος.

Τέλος, από το παραπάνω Θεώρημα εύκολα διαπιστώνουμε ότι το βασικό αποτέλεσμα διατηρεί την ορθότητά του, εάν κρατήσουμε το n σταθερό και στείλουμε μόνο το $x \to +\infty$. Το αποτέλεσμα αυτό το διατυπώνουμε σαν Πόρισμα:

Πόρισμα 19.1. *Εάν η μ είναι υποεκθετική και* $\operatorname{supp}\mu \subseteq \mathbb{R}_+$*, τότε για κάθε $n \in \mathbb{N}$:*

$$\lim_{x\to+\infty} ||\mu_{n,x} \circ \mathrm{T}^{-1} - (\mu^{n-1} \times \nu_x)||_{t.v} = 0$$

Κεφάλαιο 4

Εφαρμογή της αρχής του Gibbs στην χρηματοοικονομία

4.1 Μεγάλες απώλειες χαρτοφυλακίου

Οι πρώτοι που διατύπωσαν μια ασυμπτωτική εκτίμηση για τις μεγάλες απώλειες χαρτοφυλακίου ήταν οι Dembo, Deutschel και Duffie στο [6]. Εδώ θα δούμε πως οι συγγραφείς υλοποίησαν την αρχή του Gibbs για την μελέτη αυτών των απωλειών, καθώς και τα συμπεράσματα που προκύπτουν από την εφαρμογή αυτή.

Ως πρότυπο για τη μελέτη των μεγάλων απωλειών χαρτοφυλακίου ας θεωρήσουμε ένα τραπεζικό χαρτοφυλάκιο. Συγκεκριμένα, ας θεωρήσουμε ότι μια τράπεζα έχει n το πλήθος δανειστές, k συνολικά τύπων. Για παράδειγμα, στην περίπτωσή μας μπορούμε να διαχωρίσουμε τους n δανειστές σε $k = 2$ κατηγορίες/τύπους: αυτούς που έχουν υψηλή πιθανότητα να ξεπληρώσουν την οφειλή τους και αυτούς που έχουν μικρή πιθανότητα. Για κάθε θέση $i = 1, ..., n$, τη στιγμή της ανατίμησης του χαρτοφυλακίου ορίζουμε τις τυχαίες μεταβλητές:

- $U_i \geq 0$: Δηλώνει την έκθεση (το ποσό που θα χαθεί στο ενδεχόμενο απώλειας)
- Z_i: Δηλώνει την ύπαρξη ή μη απώλειας. Ορίζεται ως:

$$Z_i = \begin{cases} 1, & \text{απώλεια} \\ 0, & \text{όχι απώλεια} \end{cases}$$

Με βάση τον παραπάνω συμβολισμό, κάθε θέση i του χαρτοφυλακίου βιώνει απώλεια $Z_i U_i$, κατά την ανατίμησή του.
Συνυπολογίζοντας κάθε θέση i, η συνολική απώλεια του χαρτοφυλακίου θα δίνεται από τη σχέση:

$$S_n = \sum_{i=1}^{n} Z_i U_i$$

Στόχος είναι ο χαρακτηρισμός για $x \gg 0$ της δεσμευμένης κατανομής των U_i και Z_i, δεδομένου ότι $\dfrac{S_n}{n} \geq x$.

Προφανώς, για να εκτιμήσουμε τη συμπεριφορά του χαρτοφυλακίου απαιτείται η γνώση του γενικότερου κοινωνικού και οικονομικού περιβάλλοντος από το οποίο επηρεάζονται οι αποφάσεις ανθρώπων και εταιριών. Απαιτείται δηλαδή η γνώση των εξωγενών εκείνων δυνάμεων που επηρεάζουν την αγορά και έχουν αντίκτυπο στην απόδοσή της. Το περιβάλλον αυτό αποτελεί μείγμα ποικίλων παραγόντων, όπως δημογραφικών, τεχνολογικών, πολιτισμικών, και πολλών άλλων, και είναι προφανώς στοχαστικής φύσης. Συμπεριλαμβάνουμε, λοιπόν, αυτή τη γνώση της αγοράς σε μια τυχαία μεταβλητή που αποκαλούμε Y.

Ας επανέλθουμε προς στιγμήν στο παράδειγμα της τράπεζας και των δανειστών. Στην περίπτωση αυτή, κατά την κατασκευή ενός μοντέλου για την αξιολόγηση της πιστοληπτικής ικανότητας των δανειστών, η τυχαία μεταβλητή Y μπορεί να περιέχει κεντρικής σημασίας πληροφορίες για την κατάσταση των εταιριών και της βιομηχανίας. Υποθέτουμε ότι μπορούμε να διαλέξουμε την τυχαία μεταβλητή Y με τέτοιον τρόπο, ώστε δεσμεύοντας ως προς την Y, οι μεταβλητές που δηλώνουν απώλεια $Z_1, U_1, ..., Z_n, U_n$ να είναι ανεξάρτητες. Για απλότητα θεωρούμε επίσης ότι για κάθε θέση i οι U_i και Y είναι ανεξάρτητες.

Το γεγονός ότι το χαρτοφυλάκιο περιέχει k τύπους θέσεων σημαίνει ουσιαστικά ότι η κατανομή των Y, U_i, Z_i είναι η ίδια για κάθε θέση i τύπου a. Οι τύποι θέσης θεωρούμε ότι είναι σταθεροί. Στο τραπεζικό μας παράδειγμα αυτό μπορεί να σημαίνει ότι η πιστοληπτική ικανότητα ενός δανειστή δεν αλλάζει στη διάρκεια μέχρι την ανατίμηση του χαρτοφυλακίου. Παρακάτω υιοθετούμε τον εξής συμβολισμό:

1. U_a: Μια γενική μεταβλητή U_i που ανταποκρίνεται στον τύπο a.
2. Z_a: Μια γενική μεταβλητή Z_i που ανταποκρίνεται στον τύπο a.
3. $\delta_a(Y) = \mathbb{P}[Z_a = 1|Y]$: Η Y-δεσμευμένη πιθανότητα απώλειας για τη θέση τύπου a.

Θεωρούμε ακόμα ότι ο αριθμός των θέσεων τύπου a είναι ένα σταθερό κλάσμα $q_a > 0$ του συνολικού αριθμού n των θέσεων. Τέλος, θεωρούμε ότι οι τυχαίες μεταβλητές U_a έχουν πεπερασμένες εκθετικές ροπές, ώστε να ισχύει η αρχή του Gibbs. Για τον υπολογισμό της πιθανότητας μιας αναπάντεχα μεγάλης απώλειας, είναι σημαντικό να μελετήσουμε τη λογαριθμοποιημένη ροπογεννήτρια συνάρτηση $\Lambda(\cdot|Y)$:

$$\begin{aligned}
\Lambda(t|Y) &= \frac{1}{n}\log \mathbb{E}[e^{tS_n}|Y] \\
&= \frac{1}{n}\log \mathbb{E}[\prod e^{tU_iZ_i}|Y] \\
&= \frac{1}{n}\log \prod \mathbb{E}[e^{tU_iZ_i}|Y], \qquad \text{Ανεξαρτησία των } Z_i, U_i, \text{ δεδομένης της } Y \\
&= \frac{1}{n}\sum_{i=1}^{n}\log \mathbb{E}[e^{tU_iZ_i}|Y]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_a q_a \log \mathbb{E}[e^{tU_a Z_a}|Y] \\
&= \sum_a q_a \log \left(\mathbb{E}[e^{tU_a}|Y]\mathbb{P}[Z_a = 1|Y] + \mathbb{P}[Z_a = 0|Y]\right) \\
&= \sum_a q_a \log \left(\mathbb{E}[e^{tU_a}]\delta_a(Y) + 1 - \delta_a(Y)\right), \qquad \text{Ανεξαρτησία των } U_a, Y \\
&= \sum_a q_a \log \left(M_a(t)\delta_a(Y) + 1 - \delta_a(Y)\right)
\end{aligned}
$$

Όπου $M_a(t) = \mathbb{E}[e^{tU_a}] < +\infty$ είναι η ροπογεννήτρια της απώλειας στην έκθεση U_a τύπου a.

Για μεγάλο n, δεσμεύοντας στο ενδεχόμενο $\{S_n > nx\}$ και στην Y, παίρνουμε τις ακόλουθες εκτιμήσεις, σύμφωνα με την αρχή του Gibbs, όπως τη διατυπώσαμε στο προηγούμενο κεφάλαιο:

- Οι απώλειες $Z_i U_i$ παραμένουν ανεξάρτητες κι ισόνομες σε κάθε τύπο a, μολονότι οι Z_i και U_i μεταξύ τους πλέον δεν είναι κατ' ανάγκη ανεξάρτητες (όπως είχαμε υποθέσει αρχικά).

- Για εκείνες τις θέσεις i τύπου a, για τις οποίες $Z_i = 1$, η δεσμευμένη κατανομή της U_i γίνεται η:
$$d\nu_a^s = \frac{e^{su}}{M_a(s)} dQ_a(u)$$
όπου Q_a είναι η αρχική τους κατανομή και το s δίνεται από τη λύση της εξίσωσης:
$$x = \int y d\nu_a^s(y) = \int \frac{y e^{sy}}{M_a(s)} dQ_a(y) = \mathbb{E}[U_a e^{sU_a}] \frac{1}{M_a(s)} = \Lambda'(s|Y)$$
Η παραπάνω εξίσωση $x = \Lambda'(s|Y)$ έχει μοναδική λύση s, εάν $x_1(Y) < x < x_*(Y)$, όπου:
$$x_1(Y) = L'(0|Y) = \frac{1}{n}\mathbb{E}[L_n|Y] = \sum_a q_a \delta_a(Y) \mathbb{E}[U_a]$$
και,
$$x_*(Y) = \sum_{\{a:\, \delta_a(Y) > 0\}} q_a \operatorname{ess\,sup}\{U_a\}$$
δηλαδή $x_1(Y)$ είναι η μέση απώλεια ανά θέση, δεδομένης της Y, και $x_*(Y)$ είναι η μέγιστη τιμή της μέσης απώλειας για κάθε θέση, δεδομένης της Y.

- Προφανώς, για εκείνα τα i τύπου a για τα οποία $Z_i = 0$, η δεσμευμένη κατανομή των U_i παραμένει η ίδια με την αρχική τους κατανομή, Q_a, αφού στην περίπτωση αυτή δεν είχαμε απώλεια.

Παράδειγμα: Εάν υποθέσουμε ότι η U_a ακολουθεί εκθετική κατανομή με παράμετρο β_a, τότε:

$$M_a(t) = \frac{\beta_a}{\beta_a - t}$$

όταν $t < \min_a \beta_a = t_0$ και τότε:

$$\Lambda(t|Y) = \sum_a q_a \log\left(1 - \delta_a(Y) + \delta_a(Y)\frac{\beta_a}{\beta_a - t}\right)$$

όταν $t < t_0$ και $\Lambda(t|Y)$ διαφορετικά. Επίσης, σύμφωνα με τα παραπάνω, θα έχουμε μοναδική λύση $s = s(x, Y)$ της εξίσωσης $x = \Lambda'(s|Y)$ για $x > x_1(Y) = \sum_a q_a \delta_a(Y) \cdot \frac{1}{\beta_a}$. Επίσης, αν υποθέσουμε ότι $\delta_a(Y) > 0$, τότε $x_*(Y) = +\infty$. Άρα, τελικά εάν η U_i πριν τη δέσμευση είχε εκθετική κατανομή με παράμετρο β_a, τότε, μια θέση i, για την οποία $Z_i = 1$, θα έχει απώλεια έκθεσης U_i, που μετά τη δέσμευση θα έχει εκθετική κατανομή με παράμετρο $\beta_a - s(x, Y)$.

Κλείνοντας με ένα ερώτημα: Ως τροφή για σκέψη μπορούμε να θέσουμε στους εαυτούς μας το εξής ερώτημα: πως θα συμπεριφέρονταν οι U_i, εάν ήταν υποεκθετικά κατανεμημένες; Αυτό το ερώτημα δεν είναι ήσσονος σημασίας, καθώς συμπληρώνει με μη-τετριμμένο τρόπο τα κλασικά ευρήματα των συγγραφέων στο [6]. Έτσι, λοιπόν, βάσει του ανάλογου της αρχής του Gibbs για υποεκθετικές κατανομές, θα περιμέναμε στο όριο η συμπεριφορά των τυχαίων μεταβλητών να είναι πολύ διαφορετική απ' αυτήν που παρουσιάσαμε παραπάνω. Όσο φυσιολογικό κι αν φαίνεται αυτό, ωστόσο, είναι απλώς μια υπόθεση, στην οποία δεν έχει δοθεί απάντηση έως τώρα. Με μια νότα αισιοδοξίας, λοιπόν, ολοκληρώνουμε το παρόν πόνημα, αφήνοντας εδώ αυτό το ερώτημα ως έναν γόνιμο σπόρο μελλοντικής έρευνας.

Βιβλιογραφία

[1] Bobrowski A. *Functional Analysis for Probability and Stochastic Processes.* Cambridge University Press, 2005.

[2] Embrechts P. & Klüppelberg C. & Mikosch T. & Wentzell A. *Modelling Extremal Events for Finance and Insurance.* Springer-Verlag, 1997.

[3] Freidlin M. & Wentzell A. *Random Pertrubations of Dynamical Systems.* Springer-Verlag, 2012.

[4] Monaghan Angela. Us wealth inequality - top 0.1% worth as much as the bottom 90%, Nov 2014. Accessed on January 18, 2023. URL: `https://www.theguardian.com/business/2014/nov/13/us-wealth-inequality-top-01-worth-as-much-as-the-bottom-90`.

[5] Villani C. *Optimal transport, old and new.* Springer-Verlag, 2008.

[6] Dembo A. & Deutschel J-D. & Duffie D. Large portfolio losses. *Finance and Stochastics*, 8(1):3–16, 2004.

[7] MacKay D. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, 2003.

[8] Stroock D. *Probability Theory: An Analytic View.* Cambridge University Press, 2011.

[9] Dupuis P. & Richard E.S. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley & Sons, Inc., 1997.

[10] Csiszár I. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

[11] Gálfi V.M. & Lucarini V. & Ragone F. & Wouters J. Applications of large deviation theory in geophysical fluid dynamics and climate science. *La Rivista del Nuovo Cimento*, 44:291–363, 2021.

[12] Armendáriz I & Loulakis M. Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic Processes and Their Applications*, 121(5):1138–1147, 2011.

[13] Costin O. *Asymptotics and Borel Summability.* CRC Press, 2009.

[14] Dembo A. & Zeitouni O. Refinements of the gibbs conditioning principle. *Probab. Theory Relat. Fields*, 104:1–14, 1996.

[15] Dembo A. & Zeitouni O. *Large Deviations Techniques and Applications.* Springer-Verlag, 1998.

[16] Billingsley P. *Convergence of Probability Measures.* Wiley & Sons, Inc., 1999.

[17] Papageorgiou N.S. & Winkert P. *Applied Nonlinear Functional Analysis.* De Gryter, 2018.

[18] Foss S. & Korshunov D. & Zachary S. *An Introduction to Heavy-Tailed and Subexponential Distributions.* Springer-Verlag, 2013.

[19] Varadhan S.R.S. *Large Deviations.* American Mathematical Society, 2016.

[20] Rassoul-Agha F. & Seppalainen T. *A Course on Large Deviations with an Introduction to Gibbs Measures.* American Mathematical Society, 2014.