ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Θεωρία Μεγάλων Αποκλίσεων & Εφαρμογές

ΕΥΑΓΓΕΛΟΣ ΑΝΤΥΠΑΣ/ ge15095
Επ :ΒΑΣΙΛΕΙΟΣ ΠΑΠΑΝΙΚΟΛΑΟΥ
08/10/2021, Αθήνα

Σκοπός Διπλωματικής Εργασίας

Σε αυτήν τη διπλωματική εργασία ο σκοπός μας ήταν να κάνουμε εισαγωγή στη Θεωρία των Μεγάλων Αποκλίσεων εξετάζοντας τις προεκτάσεις αυτών στην περιοχή των Μαρκοβιανών Αλυσίδων τα θέματα που καλύφθηκαν :

- Η Αρχή των Μεγάλων Αποκλίσεων (LDP)
- Το Θεώρημα Cramer
- Το Θεώρημα Gartner-Ellis
- Το Θεώρημα Sanov
- Εφαρμογή στις Μαρκοβιανές Αλυσίδες

Διάγραμμα Ροής

LDP → Θεώρημα Cramer → Θεώρημα Gartner-Ellis → Markov Chains

1 Εισαγωγή

Τι είναι οι Μεγάλες Αποκλίσεις ;

Το αντικείμενο της Θεωρίας των Μεγάλων Αποκλίσεων είναι η μελέτη **σπανίων ενδεχομένων**. Με τον όρο σπάνιο ενδεχόμενο εννοούμε ένα ενδεχόμενο το οποίο *αποκλίνει* από την τυπική συμπεριφορά στο θεωρούμενο χώρο πιθανότητας, δηλαδή είναι κατά μια έννοια μακριά από τη μέση τιμή.

1 Εισαγωγή

Πως δουλεύουμε στην πράξη;

Το βασικό μας εργαλείο είναι η Αρχή των Μεγάλων Αποκλίσεων(LDP). Η LDP μας προσφέρει έναν πολύ βολικό χαρακτηρισμό της οριακής συμπεριφοράς μιας ακολουθίας μέτρων πιθανότητας $\{\mu_n\}$ πάνω σε ένα μετρήσιμο χώρο (X,$\mathcal{B}$) ως προς μια **συνάρτηση ταχύτητας.** Ο χαρακτηρισμός αυτός γίνεται μέσω ασυμπτωτικών εκθετικών φραγμάτων για τις τιμές που η $\{\mu_n\}$ αποδίδει στα μετρήσιμα σύνολα του X.

<u>Ορισμός :</u> Μια απεικόνιση $I: X \rightarrow [0, \infty]$ θα λέγεται συνάρτηση ταχύτητας εάν και μόνο εάν είναι κάτω ημισυνεχής, δηλαδή για κάθε α>0 το σύνολο $\Psi_I(\alpha) = \{x : I(x) \leq \alpha\}$ είναι κλειστό υποσύνολο του Χ. Επιπλέον εάν το παραπάνω σύνολο είναι συμπαγές θα λέμε ότι η I είναι καλή συνάρτηση ταχύτητας.

2 Η Αρχή των Μεγάλων Αποκλίσεων

<u>Ορισμός:</u> Έστω X ένας πολωνικός χώρος και $\mathcal{B}$ η σ-άλγεβρα Borel του X και $\{\mu_n\}$ ακολουθία μέτρων πιθανότητας ορισμένη πάνω στο (X,$\mathcal{B}$) τότε θα λέμε ότι η $\{\mu_n\}$ ικανοποιεί την Αρχή των Μεγάλων Αποκλίσεων με συνάρτηση ταχύτητας I εάν για κάθε μετρήσιμο σύνολο B ισχύει:

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

όπου η ακολουθία $\alpha_n \to 0, \;\; \kappa\alpha\theta\acute{\omega}\varsigma\; n \to \infty$.

2 Η Αρχή των Μεγάλων Αποκλίσεων

Σχόλιο: Με μια πρώτη ματιά η LDP φαίνεται μάλλον τρομακτική, ωστόσο εάν καταφέρουμε να την αποκωδικοποιήσουμε θα δούμε ότι ο ορισμός που δώσαμε είναι όχι μόνον απολύτως λογικός αλλά και πολύ χρήσιμος.

Θα επιχειρήσουμε να αποδομήσουμε τον ορισμό διατυπώνοντας το θεώρημα Cramer το οποίο είναι ίσως το κεντρικότερο αποτέλεσμα των μεγάλων αποκλίσεων.

3 Το Θεώρημα Cramer στο $\mathbb{R}$

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

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

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

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

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

Το Θεώρημα Cramer στο $\mathbb{R}$

<u>Ισοδύναμα</u>: Εάν μ_n είναι η κατανομή του δειγματικού μέσου $\overline{X}_n$, δηλαδή $\mu_n(\mathrm{B})$=$\mathbb{P}\left(\overline{X}_n \epsilon \mathrm{~B}\right)$, όπου B μετρήσιμο σύνολο, τότε το θεώρημα Cramer μας λέει ότι η $\{\mu_n\}$ ικανοποιεί την LDP με συνάρτηση ταχύτητας την Λ^* και $\alpha_n = \frac{1}{n}$.

- Γιατί είναι λογικό αυτό το αποτέλεσμα ;
- Πως σκεφτήκαμε να ορίσουμε τις Λ^*, Λ ;

Ένα Κινητήριο Παράδειγμα

Το επόμενο παράδειγμα θα πρέπει να μας πείσει ότι πράγματι υπάρχει μια πολύ ισχυρή διαίσθηση πίσω από όσα έχουμε διατυπώσει μέχρι στιγμής.

Παράδειγμα: Έστω ότι εργαζόμαστε σε ένα χώρο πιθανότητας (Ω, F, P) και $X_1, \ldots, X_n \backsim N(0,1)$. Ο Ασθενής Νόμος των Μεγάλων Αριθμών μας υπαγορεύει ότι για κάθε **δ**>0 η πιθανότητα $p_n = P\left(\overline{X}_n > \delta\right) \rightarrow \mathbf{0}$, $\kappa\alpha\theta\acute{\omega}\varsigma\, n \rightarrow \infty$. Το ενδιαφέρον ερώτημα εδώ είναι πόσο μικρές γίνονται αυτές οι πιθανότητες, δηλαδή με τι **ταχύτητα** λαμβάνει χώρα αυτή η σύγκλιση.

Το Θεώρημα Cramer στο $\mathbb{R}$

Από το Κεντρικό Οριακό Θεώρημα γνωρίζουμε ότι $P\left(\sqrt{n}\overline{X}_n > \sqrt{n}\delta\right) = 1 - \Phi(\sqrt{n}\delta)$, όπου Φ είναι η αθροιστική συνάρτηση κατανομής της τυπικής κανονικής κατανομής.

Ορίζουμε τη συνάρτηση πιθανότητας ουράς $Q(x) = P(X > x) =$
$=1 - \Phi(\mathrm{x})$.

Για την Q μπορεί να αποδειχθεί ότι ισχύουν τα εξής φράγματα:

$$\frac{x\varphi(x)}{1+x^2} < Q(x) \leq \frac{\varphi(x)}{x}, \qquad x > 0$$

όπου φ(x) είναι η συνάρτηση πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής.

Επιπλέον για δύο συναρτήσεις f, g θα γράφουμε $f \backsim g$ εάν $\lim\limits_{x \to \infty} \frac{f(x)}{g(x)} = 1$.
Δεν είναι δύσκολο να δούμε τώρα ότι $\frac{Q(x)}{\varphi(x)} \backsim \frac{1}{x}$ το οποίο όμως συνεπάγεται ότι

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

Αυτό σημαίνει ότι με πολύ μικρή πιθανότητα της τάξεως του $\frac{1}{\sqrt{n}} e^{-\frac{n\delta^2}{2}}$ έχουμε $\overline{X}_n > \delta$.

3 Το Θεώρημα Cramer στο $\mathbb{R}$

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

3 Το Θεώρημα Cramer στο $\mathbb{R}$

Για να κατανοήσουμε λίγο καλύτερα τι θα πει ότι μπορούμε να αγνοήσουμε τον πολυωνυμικό όρο αρκεί να δούμε την παραπάνω έκφραση υπό λίγο διαφορετική σκοπιά, δηλαδή αντί να γράφουμε :

$$p_n \sim (\sigma\tau\alpha\vartheta.)\frac{1}{\sqrt{n}}e^{-\frac{n\delta^2}{2}}$$

Εξετάζουμε τον όρο :

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

Το Θεώρημα Cramer στο $\mathbb{R}$

Άρα είδαμε ότι

$$\frac{1}{n}\log \mathbb{P}(\bar{X}_n > \delta) \to -\frac{\delta^2}{2}, \quad \kappa\alpha\vartheta\acute{\omega}\varsigma \quad n \to \infty.$$

Αυτή είναι μια κλασική εκτίμηση στο πλαίσιο της Θεωρίας των Μεγάλων Αποκλίσεων, ουσιαστικά η ποσότητα $\frac{\delta^2}{2}$ είναι η *ταχύτητα* που μας ενδιέφερε.

Γενικεύοντας το Παράδειγμα

Το παράδειγμα που μελετήσαμε στις προηγούμενες διαφάνειες ήταν εστιασμένο στην τυπική κανονική κατανομή, το πρόβλημα που μας απασχολεί τώρα είναι τι συμβαίνει στην περίπτωση που οι ανεξάρτητες X_i ακολουθούν μια αυθαίρετη κατανομή μ για την οποία η μέση τιμή είναι πεπερασμένη.

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

3 Το Θεώρημα Cramer στο $\mathbb{R}$

Αν υποθέσουμε ότι η κατανομή μ έχει αρκούντως ελαφριά ουρά, δηλαδή η πιθανότητα ουράς μειώνεται εκθετικά τότε είναι λογικό να αναμένουμε ότι θα μπορούμε να βρούμε μια σχέση της μορφής :

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

όπου A, b σταθερές και c>0 εξαρτάται από το δ και την κατανομή μ.

3 Το Θεώρημα Cramer στο $\mathbb{R}$

Ερώτημα : Πως υπολογίζουμε το c=c(δ , μ) για γενική κατανομή μ;

Σε αυτό ακριβώς το ερώτημα μας απαντάει το θεώρημα Cramer.
Η προσέγγιση που ακολουθούμε είναι η εξής :

Υποθέτουμε ότι η ροπογεννήτρια συνάρτηση είναι πεπερασμένη, δηλαδή:

$$M_{X_1}(\theta) = \mathbb{E}(e^{\theta X_1}) = \int_{\mathbb{R}} e^{\theta x} d\mu(x) < \infty, \quad \forall \theta > 0$$

3 Το Θεώρημα Cramer στο $\mathbb{R}$

Σε αυτήν την περίπτωση με χρήση της εκθετικής ανισότητας Markov θα έχουμε:

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

εάν ορίσουμε τώρα τη λογαριθμική ροπογεννήτρια :

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

Το Θεώρημα Cramer στο $\mathbb{R}$

Η ανισότητα μας παίρνει τη μορφή :

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

Επομένως εάν καταφέρουμε να βρούμε ένα $\theta \ge 0$ τέτοιο ώστε

$$-n(\theta a - \Lambda(\theta)) < 0$$

θα έχουμε ένα άνω φράγμα το οποίο συγκλίνει στο 0 εκθετικά γρήγορα. Πιο συγκεκριμένα, το πιο *σφιχτό* τέτοιο άνω φράγμα προκύπτει ως λύση του παρακάτω προβλήματος βελτιστοποίησης:

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

3 Το Θεώρημα Cramer στο $\mathbb{R}$

Επιπλέον εφόσον α>m ισχύει ότι : $\sup_{\theta\geq 0}(\theta a - \Lambda(\theta)) = \sup_{\theta\in\mathbb{R}}(\theta a - \Lambda(\theta))$

άρα πλέον έχουμε κάνει πολύ σημαντική πρόοδο. Μετατρέψαμε το γενικό πρόβλημα που είχαμε στα χέρια μας σε ένα πρόβλημα βελτιστοποίησης μιας ποσότητας η οποία εξαρτάται αποκλειστικά από την κατανομή μ. Σε αυτό το πλαίσιο οδηγούμαστε με πολύ φυσικό τρόπο στον επόμενο ορισμό:

Καλούμε μετασχηματισμό Fenchel-Legendre της Λ τη συνάρτηση:

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

Το Θεώρημα Cramer στο $\mathbb{R}$

Επομένως ανακεφαλαιώνοντας, μέχρι στιγμής έχουμε δείξει ότι:

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

δηλαδή, ισοδύναμα :

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

Χρησιμοποιώντας ένα επιχείρημα αλλαγής μέτρου θα μπορούσαμε να πάρουμε κι ένα κάτω φράγμα παρόμοιας μορφής :

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

Σχόλιο: Επομένως βλέπουμε ότι το θεώρημα Cramer κωδικοποιεί με πολύ φυσικό τρόπο τη διαδικασία που ακολουθήσαμε στο προηγούμενο παράδειγμα . Οι ορισμοί των Λ^*, Λ αλλά και της LDP δεν είναι αυθαίρετοι αλλά ανακύπτουν φυσικά μέσα από τη μελέτη τέτοιου είδους προβλημάτων.

Επιπλέον όταν η μέση τιμή m υπάρχει μπορούμε να σκεφτόμαστε τη Λ^* ως ένα μέτρο απόστασης από την m, όσο πιο μακριά βρισκόμαστε από τη m τόσο γρηγορότερα έχουμε σύγκλιση στο 0.

3 Το Θεώρημα Cramer στο $\mathbb{R}$

<u>Επαλήθευση:</u> Κάνουμε τους υπολογισμούς του θεωρήματος Cramer για την κανονική κατανομή .

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

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

και παραγωγίζοντας ως προς θ *θα πάρουμε,*

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

συνεπώς

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

Άρα για $x > m$, *έχουμε*

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

πράγματι έχουμε συμφωνία.

Ένα Παράδειγμα από την Πραγματική Ζωή

Ας υποθέσουμε ότι μας ανήκει μια ασφαλιστική εταιρία με n πελάτες, η ασφαλιστική αποζημίωση που μπορεί να κληθούμε να πληρώσουμε για κάθε πελάτη είναι μια τυχαία μεταβλητή X_i , άρα έχουμε $X_1, \dots, X_n \backsim \mu$ ανεξάρτητες και η μ έχει πεπερασμένη μέση τιμή m. Συνεπώς εάν η εταιρία μας χρεώνει κάποιο ασφάλιστρο α>m η πιθανότητα οικονομικής απώλειας είναι :

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

3α Το Θεώρημα Cramer στο $\mathbb{R}^d$

Η γενίκευση του θεωρήματος Cramer για τυχαία διανύσματα γίνεται με φυσιολογικό τρόπο , ορίζουμε τις ποσότητες :

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

Τότε:

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

όπου $D_\Lambda = \{\theta \in \mathbb{R}^d | \Lambda(\theta) < \infty\}$.

4 Το Θεώρημα Gartner-Ellis

Εστιάζουμε την προσοχή μας τώρα στο γεγονός ότι για να ισχύει το θεώρημα Cramer θα πρέπει οι $\mathrm{X}_1, \ldots, \mathrm{X}_n$ να είναι ανεξάρτητες. Τι γίνεται στην περίπτωση που χάνουμε την ανεξαρτησία ;

Μπορούμε και πάλι να πούμε κάτι για της συμπεριφορά της $\{\mu_n\}$;

Η απάντηση είναι θετική και είναι το αντικείμενο του θεωρήματος Gartner-Ellis.

4 Το Θεώρημα Gartner-Ellis

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

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

Κάνουμε την εξής υπόθεση :

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

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

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

Πριν διατυπώσουμε το θεώρημα χρειαζόμαστε μερικούς τεχνικούς ορισμούς :

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

$$\langle \theta, y \rangle - \Lambda^*(y) > \langle \theta, x \rangle - \Lambda^*(x)$$

Το θ στην παραπάνω σχέση καλείται εκτεθειμένο υπερεπίπεδο (exposing hyperplane).

4 Το Θεώρημα Gartner-Ellis

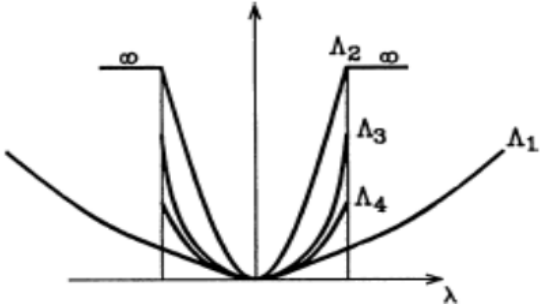
Ορισμός *Θα λέμε ότι μια κυρτή συνάρτηση* $\Lambda : \mathbb{R}^d \longrightarrow (-\infty, \infty]$ *είναι ουσιωδώς λεία εάν:*

(α) Το D_Λ^o *είναι μη κενό.*

(β) Η Λ *είναι διαφορίσιμη σε ολόκληρο το* D_Λ^o.

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

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



4 Το Θεώρημα Gartner-Ellis

Είμαστε τώρα σε θέση να διατυπώσουμε το θεώρημα :

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

(α) Για κάθε κλειστό σύνολο F,

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

(β) Για κάθε ανοιχτό σύνολο G,

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

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

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

<u>Σχόλιο:</u> Το σημείο που είναι άξιο παρατήρησης εδώ είναι ότι συγκριτικά με το θεώρημα Cramer η διατύπωση του θεωρήματος Gartner-Ellis είναι τεχνικά πιο περίπλοκη, αυτό είναι το κόστος που πληρώνουμε για τη χαλάρωση της απαίτησης της ανεξαρτησίας.

Μεγάλες Αποκλίσεις για Μαρκοβιανές Αλυσίδες

Οπλισμένοι με τη γενικότητα που μας προσφέρει το θεώρημα Gartner-Ellis μπορούμε να επεκτείνουμε τα αποτελέσματα μας στην περίπτωση που οι $X_1, \dots, X_n, \dots$ παρουσιάζουν μαρκοβιανή εξάρτηση.

Προαπαιτούμενα:

Για τη συνέχεια θεωρούμε ότι οι μαρκοβιανές αλυσίδες που μελετάμε λαμβάνουν τιμές σε ένα πεπερασμένο σύνολο Σ, το οποίο μπορούμε χωρίς βλάβη της γενικότητας να θεωρήσουμε ότι είναι Σ={1,…,r}.

Μεγάλες Αποκλίσεις για Μαρκοβιανές Αλυσίδες

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

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

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

Μεγάλες Αποκλίσεις για Μαρκοβιανές Αλυσίδες

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

Μεγάλες Αποκλίσεις για Μαρκοβιανές Αλυσίδες

Το επόμενο θεώρημα μας δίνει μια σημαντική ιδιότητα των ανάγωγων πινάκων:

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

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

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

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

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

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

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

LDP για Αθροιστικά Συναρτησιακά Μ.Α.

Το αντικείμενο αυτής της ενότητας είναι η μελέτη συναρτησοειδών της μορφής

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

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

Μεγάλες Αποκλίσεις για Μαρκοβιανές Αλυσίδες

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

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

Επειδή οι ποσότητες $e^{\langle\theta, f(j)\rangle}$ είναι πάντοτε θετικές, εάν ο $\pi(i,j)$ είναι ανάγωγος τότε και ο $\pi_\theta(i,j)$ θα είναι ανάγωγος. Επιπλέον για κάθε $\theta \in \mathbb{R}^d$ θα συμβολίζουμε με $\rho(\mathbf{\Pi}_\theta)$ την ιδιοτιμή Perron-Frobenius του πίνακα $\mathbf{\Pi}_\theta$.

Μεγάλες Αποκλίσεις για Μαρκοβιανές Αλυσίδες

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

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

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

$$- \inf_{z \in \Gamma^o} I(z) \leq \liminf_{n \to \infty} \frac{1}{n} \log P_\sigma^\pi (Z_n \in \Gamma) \leq$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \log P_\sigma^\pi (Z_n \in \Gamma) \leq - \inf_{z \in \bar{\Gamma}} I(z).$$

Συμπέρασμα: Άρα βλέπουμε ότι η ιδιοτιμή Perron-Frobenius παίζει το ρόλο της Λ(θ).

ΤΕΛΟΣ ΠΑΡΟΥΣΙΑΣΗΣ

ΕΥΧΑΡΙΣΤΩ ΠΟΛΥ