ΘΕΩΡΙΑ ΜΕΓΑΛΩΝ ΑΠΟΚΛΙΣΕΩΝ

Θερινό Σχολειό Μαθηματικών, Ηράκλειο 2005

Εισαγωγή

Η θεωρία των μεγάλων αποκλίσεων συνίσταται στην εκτίμηση της πιθανότητας σπάνιων ενδεχομένων-ενδεχομένων δηλαδή που *αποκλίνουν* από την τυπική συμπεριφορά στο θεωρούμενο χώρο πιθανότητας. Ας πάρουμε για παράδειγμα μια ακολουθία ανεξάρτητων ισόνομων τυχαίων μεταβλητών $\{X_i\}_{i\in\mathbb{N}}$ με μηδενική μέση τιμή. Από τον (ασθενή) νόμο των μεγάλων αριθμών, το ενδεχόμενο ο μέσος όρος των $\{X_1,\ldots,X_n\}$ να βρίσκεται έξω από μια περιοχή του 0 γίνεται σπάνιο καθώς το n γίνεται μεγάλο: για κάθε $\delta>0$,

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

καθώς $n\to\infty$. Τι μπορούμε να πούμε όμως για την ταχύτητα αυτής της σύγκλισης;

Αν υποθέσουμε ότι η κοινή κατανομή των X_i είναι η κανονική $\mathcal{N}(0,\sigma^2)$, μπορούμε να προχωρήσουμε την εκτίμησή μας περισσότερο. Η κατανομή μ_n του μέσου όρου είναι η κανονική $\mathcal{N}(0,\frac{\sigma^2}{n})$, επομένως:

$$\begin{aligned}\mathbb{P}\left(\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right|\geq\delta\}\right) &= 2\int_{\delta}^{\infty}\sqrt{\frac{n}{2\pi\sigma^2}}e^{-\frac{nx^2}{2\sigma^2}}\,dx\\ &= \frac{2}{\sqrt{2\pi}}\int_{\frac{\delta\sqrt{n}}{\sigma}}^{\infty}e^{-x^2/2}\,dx.\end{aligned}$$

Χρησιμοποιώντας τις ανισότητες:

$$\frac{x}{1+x^2}e^{-x^2/2}\leq\int_x^{\infty}e^{-y^2/2}\,dy\leq\frac{1}{x}e^{-x^2/2}\qquad\forall x>0,\tag{1}$$

συμπεραίνουμε πως:

$$\frac{1}{n}\log\mathbb{P}\left(\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right|\geq\delta\}\right)=\frac{1}{n}\log\mu_n\left((-\delta,\delta)^c\right)\longrightarrow-\frac{\delta^2}{2\sigma^2}.$$

Τέτοιου είδους αποτελέσματα και οι συνέπειές τους είναι το αντικείμενο της θεωρίας των μεγάλων αποκλίσεων. Φυσικά, το να μελετά κανείς την πιθανότητα σπάνιων ενδεχομένων θέτει τα πράγματα σ' ένα πολύ γενικό πλαίσιο. Έτσι, από τη μια μεν η θεωρία των μεγάλων αποκλίσεων βρίσκει εφαρμογή σε περιοχές τόσο διαφορετικές όσο η Στατιστική, η Θεωρία Πληροφορίας, οι Τηλεπικοινωνίες και η Στατιστική Μηχανική, από την άλλη όμως συχνά χρειάζεται να καταφύγει κανείς σε ειδικές για το εκάστοτε πρόβλημα μεθόδους. Στις διαλέξεις αυτές θα προσπαθήσουμε μέσα από σχετικά απλά παραδείγματα να αναδείξουμε κάποιες από τις κεντρικές ιδέες και τεχνικές της θεωρίας.

1 Το θεώρημα του Cramér

Το πρώτο μας παράδειγμα αφορά στην κατανομή του μέσου όρου ανεξάρτητων ισόνομων τυχαίων μεταβλητών $\{X_i\}_{i\in\mathbb{N}}$, όχι κατ' ανάγκη κανονικών. Θα υποθέσουμε ότι η κοινή τους κατανομή μ έχει την ιδιότητα:

$$M(\lambda):=\int e^{\lambda x}\mu(dx)<+\infty,\qquad\forall\lambda\in\mathbb{R}.\tag{2}$$

Έστω τώρα $\alpha \geq \bar{x}$, όπου $\bar{x} = \int x d\mu(x)$ είναι η μέση τιμή της μ. Από την ανισότητα του Chebyshev:

$$\begin{aligned} \mu_n([\alpha, +\infty)) &= \mathbb{P}\left(\{\frac{1}{n}\sum_{i=1}^{n} X_i \geq \alpha\}\right) \leq e^{-\lambda n\alpha}\mathbb{E}\left(\exp(\lambda\sum_{i=1}^{n} X_i)\right) \\ &= e^{-\lambda n\alpha} M^n(\lambda), \qquad \forall \lambda \geq 0. \end{aligned}$$

Επομένως,

$$\frac{1}{n}\log \mu_n([\alpha,\infty)) \leq -\sup_{\lambda\geq 0}(\lambda\alpha - \log M(\lambda)) = -\sup_{\lambda}(\lambda\alpha - \log M(\lambda)). \tag{3}$$

(Η τελευταία ισότητα προκύπτει από την ανισότητα του Jensen. Μπορείτε να δείτε πώς;)

Πόσο καλή είναι άραγε η παραπάνω εκτίμηση; Την απάντηση σ' αυτό το ερώτημα δίνει το λήμμα που ακολουθεί. Πριν το διατυπώσουμε όμως ας δώσουμε ένα όνομα στο δεξί σκέλος της (3). Ορίζουμε

$$I(x) = \sup_{\lambda}(\lambda x - \log M(\lambda)). \tag{4}$$

Λήμμα 1 Για κάθε $\delta > 0$ και $x \geq \bar{x}$ έχουμε:

$$\liminf \frac{1}{n}\log \mu_n([x, x+\delta)) = \liminf \frac{1}{n}\log \mathbb{P}\left(\{x \leq \frac{1}{n}\sum_{i=1}^{n} X_i < x+\delta\}\right) \geq -I(x). \tag{5}$$

Απόδειξη: Θα διακρίνουμε δύο περιπτώσεις, ανάλογα με το αν το supremum στην (4) λαμβάνεται σε πεπερασμένο λ ή όχι. Ας υποθέσουμε πρώτα ότι υπάρχει πεπερασμένο λ^* τέτοιο ώστε

$$I(x) = \sup_{\lambda}(\lambda x - \log M(\lambda)) = \lambda^* x - \log M(\lambda^*).$$

Από την υπόθεση (2) η συνάρτηση $\log M(\lambda)$ είναι παραγωγίσιμη και

$$x = \frac{M'(\lambda^*)}{M(\lambda^*)} = \int y \frac{e^{\lambda^* y}}{M(\lambda^*)} d\mu(y). \tag{6}$$

Ορίζουμε τώρα ένα νέο μέτρο πιθανότητας μ^* ως εξής:

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

Για κάθε ε με $0 < \varepsilon < \delta$ έχουμε:

$$\begin{aligned} \mathbb{P}\left(\{x \leq \frac{1}{n}\sum_{i=1}^{n} X_i < x+\delta\}\right) &\geq \int \cdots \int_{nx \leq \sum_{i=1}^{n} x_i \leq nx+n\varepsilon} d\mu(x_1)\cdots d\mu(x_n) \\ &\geq \int \cdots \int_{nx \leq \sum_{i=1}^{n} x_i \leq nx+n\varepsilon} e^{\lambda^* \sum_{i=1}^{n} x_i - n\lambda^* x - n|\lambda^*|\varepsilon} d\mu(x_1)\cdots d\mu(x_n) \\ &= e^{-n\lambda^* x - n|\lambda^*|\varepsilon}\, M^n(\lambda^*)\, \mathbb{P}^*(\{0 \leq \frac{1}{n}\sum_{i=1}^{n} X_i - x \leq \varepsilon\}). \end{aligned}$$

Η σχέση (6) δηλώνει ότι το x είναι η μέση τιμή της κατανομής μ^*. Αν λοιπόν η κοινή κατανομή των $\{X_i\}$ είναι η μ^*, ο μέσος όρος των $\{X_1, \ldots, X_n\}$ συγκεντρώνεται γύρω από το x. Αυτή η συμπεριφορά που πριν ήταν άτυπη, με την αλλαγή του μέτρου γίνεται η τυπική συμπεριφορά. Αυτό είναι μια συνηθισμένη τεχνική στην απόδειξη κάτω φραγμάτων για την πιθανότητα σπάνιων ενδεχομένων στην θεωρία των μεγάλων αποκλίσεων. Από το κεντρικό οριακό θεώρημα:

$$\mathbb{P}^*(\{0 \le \frac{1}{n}\sum_{i=1}^{n} X_i - x \le \varepsilon\}) \longrightarrow \frac{1}{2}, \qquad n \to \infty.$$

Επομένως,

$$\liminf \frac{1}{n} \log \mu_n([x, x+\delta)) \ge -\lambda^* x + \log M(\lambda^*) - |\lambda^*|\varepsilon = -I(x) - |\lambda^*|\varepsilon.$$

Παίρνοντας $\varepsilon \to 0$ λαμβάνουμε την (5).

Δεν είναι όμως πάντα δυνατόν να βρούμε τέτοιο λ^*. Αν λ.χ. $\mu(x, +\infty) = 0$ (και μόνο τότε όπως μπορείτε εύκολα να διαπιστώσετε):

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

και άρα

$$I(x) = \sup_{\lambda}(\lambda x - \log \int e^{\lambda y}\ d\mu(y)) = -\lim_{\lambda \to \infty} \log \int_{-\infty}^{x} e^{\lambda(y-x)}\ d\mu(y) = -\log \mu(\{x\}).$$

Και σ' αυτήν την περίπτωση όμως έχουμε

$$\mu_n([x, x+\delta)) = \mathbb{P}(X_1 = X_2 = \cdots = X_n = x) = \mu^n(\{x\}),$$

επομένως

$$\frac{1}{n} \log \mu_n([x, x+\delta)) = \log \mu(\{x\}) = -I(x).$$

□

Συνδυάζοντας το λήμμα που μόλις δείξαμε με το άνω φράγμα (3) έχουμε για κάθε $\alpha \ge \bar{x}$:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \ge \alpha\right) = -I(\alpha). \tag{7}$$

Με τον ίδιο τρόπο μπορούμε να δούμε ότι για κάθε $\beta \le \bar{x}$:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \le \beta\right) = -I(\beta). \tag{8}$$

Ας συνοψίσουμε σ' αυτό το σημείο μερικές ιδιότητες της I:

1. Η I είναι κάτω ημισυνεχής, δηλ. $x_n \to x \Rightarrow I(x) \le \liminf I(x_n)$.
2. Η I είναι κυρτή.
3. $I(x) \ge 0$.
4. $I(\bar{x}) = 0$.
5. Η $I(x)$ είναι φθίνουσα στο $(-\infty, \bar{x})$ και αύξουσα στο $(\bar{x}, +\infty)$.

Παρατήρηση: Αν και κάναμε την υπόθεση (2) για να απλουστεύσουμε την απόδειξη, τα αποτελέσματα (7), (8), αλλά και όλες οι προηγούμενες ιδιότητες της I ισχύουν και χωρίς αυτήν. Επιπλέον είναι εύκολο να δει κανείς ότι αν η ανισότητα στην (2) ικανοποιείται για όλα τα λ σε μια περιοχή του μηδενός, τότε $I(x) > 0$, $\forall x \neq \bar{x}$. (Για την ακρίβεια αν υπάρχει $\varepsilon > 0$ ώστε $M(\lambda) < \infty$, $\forall \lambda \in [0, \varepsilon)$ τότε $I(x) > 0$ για κάθε $x > \bar{x}$ και αντίστοιχα για $\lambda \in (-\varepsilon, 0]$.) Αν η μ δεν έχει ούτε αυτή την ιδιότητα, αυτό που μπορεί να συμβεί είναι η I να είναι μηδέν σε μια περιοχή του $\bar{x}$ (ακόμη και σε όλο το $\mathbb{R}$). Σ' αυτήν την περίπτωση αν και οι (7), (8) ισχύουν δεν παρέχουν καμμιά πληροφορία, αφού στην περίπτωση αυτή η ταχύτητα σύγκλισης είναι υποεκθετική. Τέλος αν $M(\varepsilon) < \infty$ και $M(-\varepsilon) < \infty$ για κάποιο $\varepsilon > 0$ έχουμε

$$\lim_{|x| \to \infty} I(x) = +\infty.$$

Πράγματι, για κάθε $N > 0$ έχουμε

$$\{x : I(x) \leq N\} \subset \left[-\frac{N + \log M(-\varepsilon)}{\varepsilon}, \frac{N + \log M(\varepsilon)}{\varepsilon} \right].$$

Ασκήσεις

1. Αν η μ είναι Bernoulli(p), τότε $I(x) = x \log(\frac{x}{p}) + (1-x) \log(\frac{1-x}{1-p})$ αν $x \in [0, 1]$, και $I(x) = \infty$ αν $|x| > 1$.

2. Αν η μ είναι εκθετική(γ), τότε $I(x) = \gamma x - \log(\gamma x) - 1$ για $x > 0$, και $I(x) = \infty$ για $x \leq 0$.

3. Αν η μ είναι Poisson(θ), τότε $I(x) = \theta - x - x \log(x/\theta)$ για $x \geq 0$, και $I(x) = \infty$ για $x < 0$.

4. Αν η μ είναι $\mathcal{N}(0, \sigma^2)$, τότε $I(x) = x^2/2\sigma^2$.

5. Σ' ένα παιχνίδι η πιθανότητα νίκης μας είναι $p < 1/2$. Παίζουμε n φορές αυτό το παιχνίδι με τα ενδεχόμενα νίκης σε κάθε παιχνίδι ανεξάρτητα μεταξύ τους. Αν S_n είναι το πλήθος των νικών μας δείξτε ότι:
$$\mathbb{P}(S_n \geq \frac{n}{2}) \leq \left(2\sqrt{p(1-p)} \right)^n.$$

6. Αν $\alpha \in \mathbb{R}$, υπολογίστε το όριο:
$$\lim_{n \to \infty} \frac{1}{n} \log \left[\binom{2n}{n} + \binom{2n}{n+1} \alpha + \ldots + \binom{2n}{2n} \alpha^n \right].$$

Εφαρμογή: Η θεωρία των μεγάλων αποκλίσεων ξεκίνησε από προβλήματα Στατιστικής. Ας δούμε λοιπόν σ' αυτό το σημείο μια εφαρμογή του αποτελέσματος που αποδείξαμε στον ελέγχο υποθέσεων. Ας υποθέσουμε ότι παρατηρούμε μια ακολουθία από ανεξάρτητες ισόνομες τυχαίες μεταβλητές $X_1, X_2, \ldots$, των οποίων η κοινή κατανομή είναι είτε μ είτε ν ($\mu \neq \nu$). Δεν γνωρίζουμε όμως ποιά από τις δύο και θέλουμε να αποφασίσουμε γιαυτό. Αν υπάρχουν ενδεχόμενα με θετική πιθανότητα ως προς το ένα μέτρο αλλά μηδενική ως προς το άλλο το πρόβλημα είναι εύκολο (γιατί;) Ας υποθέσουμε λοιπόν επιπλέον ότι τα μ και ν είναι ισοδύναμα ($\mu \sim \nu$), δηλ.:

$$\mu(A) = 0 \Leftrightarrow \nu(A) = 0.$$

Από ένα κλασσικό θεώρημα της θεωρίας μέτρου (Radon-Nikodym) υπάρχει μια συνάρτηση $f > 0$, με $\int f \, d\mu = 1$, ώστε $\nu(A) = \int_A f \, d\mu$. Αν τα μ, ν είναι διακριτές κατανομές αυτό σημαίνει ότι $f(x) = \nu(\{x\})/\mu(\{x\})$ αν $\mu(x) \neq 0$ και 1 διαφορετικά. Αν πάλι η μ έχει πυκνότητα $g(x)$ και η ν έχει πυκνότητα $h(x)$, τότε $f(x) = h(x)/g(x)$ αν $g(x) \neq 0$ και 1 διαφορετικά. Η συνάρτηση f ονομάζεται στη Στατιστική "πιθανοφάνεια". Η απόφασή μας δεδομένων των $X_1, \ldots, X_n$ είναι μια συνάρτηση $H_n(X_1, \ldots, X_n)$ με τιμή μ είτε ν και η απόδοσή της ελέγχεται από τις πιθανότητες λάθους:

$$a_n = \mu^n(H_n = \nu), \qquad b_n = \nu^n(H_n = \mu).$$

Αν δεν επιβάλλουμε περιορισμούς στο a_n μπορούμε να κάνουμε το $b_n = 0$ διαλέγοντας $H_n = \nu$ (οπότε $a_n = 1$.) Συνήθως λοιπόν διαλέγουμε ένα επίπεδο σφάλματος a_n και προσπαθούμε να ελαχιστοποιήσουμε το b_n. Το κριτήριο που επιτυγχάνει αυτή τη βελτιστοποίηση είναι το κριτήριο Neyman-Pearson που περιγράφουμε παρακάτω. Έστω

$$S_n = \frac{1}{n}\sum_{j=1}^{n} \log f(X_j)$$

και ένα "κατώφλι" γ_n. Το κριτήριο Neyman-Pearson ορίζει $H_n(X_1, \ldots, X_n) = \mu$ αν $S_n \leq \gamma_n$, διαφορετικά $H_n(X_1, \ldots, X_n) = \nu$. Ορίζουμε τώρα

$$\bar{x}_0 = \int d\mu \log f(x) < \log\left(\int d\mu f(x)\right) = \log \int 1 \; d\nu = 0$$

Η ανισότητα Jensen που χρησιμοποιήσαμε παραπάνω είναι αυστηρή γιατί $\mu(f \neq 1) > 0$. Ομοίως:

$$\bar{x}_1 = \int d\nu \log f(x) > -\log\left(\int d\nu 1/f(x)\right) = -\log \int 1 \; d\mu = 0.$$

Λήμμα 2 Ο έλεγχος Neyman-Pearson με σταθερό κατώφλι $\gamma \in (\bar{x}_0, \bar{x}_1)$ ικανοποιεί τα εξής:

$$\limsup_{n\to\infty} \frac{1}{n}\log a_n \leq -I_\mu(\gamma) < 0, \qquad \lim_{n\to\infty} \frac{1}{n}\log b_n = -I_\nu(\gamma) = \gamma - I_\mu(\gamma) < 0.$$

Απόδειξη: Η απόδειξη είναι άμεση συνέπεια των (3), (8) και της παρατήρησης ότι (από την ανισότητα Hölder)

$$M_\mu(\lambda) = \int d\mu[e^{\lambda \log f}] = \int d\mu[f^\lambda] \leq 1, \qquad \forall \lambda \in [0,1]$$

και ομοίως

$$M_\nu(\lambda) = \int d\nu[e^{\lambda \log f}] = \int d\nu[(\frac{1}{f})^{-\lambda}] \leq 1, \qquad \forall \lambda \in [-1,0].$$

Συνεπώς $I_\mu(\gamma) > 0$ εφόσον $\gamma > \bar{x}_0$ και $I_\nu(\gamma) > 0$ εφόσον $\gamma < \bar{x}_1$.

Λήμμα 3 (Stein) Έστω b_n^ε το infimum των b_n ανάμεσα σ' όλους τους ελέγχους για τους οποίους $a_n < \varepsilon < 1$. Τότε:

$$\lim_{n\to\infty} \frac{1}{n}\log b_n^\varepsilon = \bar{x}_0.$$

Απόδειξη: Θεωρούμε πρώτα έναν έλεγχο με σταθερό $\gamma > \bar{x}_0$. Από το προηγούμενο λήμμα $a_n < \varepsilon$ για κατάλληλα μεγάλο n, επομένως

$$\limsup_{n\to\infty} \frac{1}{n}\log b_n^\varepsilon \leq \lim_{n\to\infty} \frac{1}{n}\log b_n = \gamma - I_\mu(\gamma) \leq \gamma.$$

Εφόσον η παραπάνω εκτίμηση ισχύει για οποιοδήποτε $\gamma > \bar{x}_0$ έχουμε

$$\limsup_{n\to\infty} \frac{1}{n}\log b_n^\varepsilon \leq \bar{x}_0.$$

Για την αντίστροφη ανισότητα αρκεί να υποθέσουμε ότι $\bar{x}_0 > -\infty$. Επίσης μπορούμε να υποθέσουμε ότι τα κατώφλια γ_n ικανοποιούν $\liminf \gamma_n \geq \bar{x}_0$, διαφορετικά από το νόμο των μεγάλων αριθμών θα είχαμε $\limsup a_n = 1$. Έχουμε τώρα για κάθε $\delta > 0$ και κατάλληλα μεγάλο n:

$$\begin{aligned}
\frac{1}{n}\log b_n &= \frac{1}{n}\log \mathbb{P}^\nu(S_n \leq \gamma_n) = \frac{1}{n}\log \mathbb{E}^\mu(e^{nS_n}\chi(S_n \leq \gamma_n)) \\
&\geq \frac{1}{n}\log \mathbb{E}^\mu(e^{nS_n}\chi(\bar{x}_0 - \delta \leq S_n \leq \gamma_n)) \\
&\geq \bar{x}_0 - \delta + \frac{1}{n}\log \mathbb{P}^\mu(S_n \in [\bar{x}_0 - \delta, \gamma_n]).
\end{aligned}$$

Από τον ασθενή νόμο των μεγάλων αριθμών και το γεγονός ότι $a_n < \varepsilon$ όμως έχουμε

$$\liminf_{n\to\infty} \mathbb{P}^{\mu}(S_n \in [\bar{x}_0 - \delta, \gamma_n]) > 1 - \varepsilon,$$

οπότε

$$\liminf \frac{1}{n} \log b_n \geq \bar{x}_0 - \delta.$$

Από την βελτιστότητα των ελέγχων Neyman-Pearson και παίρνοντας $\delta \to 0$ λαμβάνουμε το ζητούμενο. $\square$

Τα αποτελέσματα στη θεωρία των μεγάλων αποκλίσεων διατυπώνονται συνήθως μ΄ ένα συγκεκριμένο τρόπο που περιγράφουμε παρακάτω. Ας πούμε λοιπόν πως έχουμε μια οικογένεια (Borel) μέτρων πιθανότητας $\{\mu_\varepsilon\}$ σ΄ έναν (πλήρη διαχωρίσιμο μετρικό) χώρο $\mathbb{X}$, που "συγκεντρώνονται" στο σημείο $x_0 \in \mathbb{X}$ (συμβολίζουμε $\mu_\varepsilon \Rightarrow \delta_{x_0}$), με την έννοια του ότι αν $x_0 \notin \bar{A}$ τότε $\mu_\varepsilon(A) \to 0$ καθώς $\varepsilon \to 0$. Θα μας ενδιέφερε να μελετήσουμε την ταχύτητα σύγκλισης, ιδιαίτερα δε να εξετάσουμε αν είναι εκθετική και να υπολογίσουμε εκφράσεις όπως:

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A), \qquad \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A).$$

Ορισμός: Θα λέμε ότι η $\{\mu_\varepsilon\}$ ακολουθεί την *αρχή της μεγάλης απόκλισης* (LDP) με *συνάρτηση ταχύτητας* I αν:
α) Η $I : \mathbb{X} \mapsto [0, \infty]$ είναι μια κάτω ημισυνεχής συνάρτηση.
β) Για κάθε ανοικτό σύνολο $G \subset \mathbb{X}$ έχουμε:

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(G) \geq - \inf_{x\in G} I(x). \tag{9}$$

γ) Για κάθε κλειστό σύνολο $C \subset \mathbb{X}$ έχουμε:

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(C) \leq - \inf_{x\in C} I(x). \tag{10}$$

Είναι εύκολο να δούμε ότι οι (9),(10) μπορούν να αντικατασταθούν από την:

$$- \inf_{x\in \mathring{A}} I(x) \leq \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \leq \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \leq - \inf_{x\in \bar{A}} I(x),$$

για κάθε Borel σύνολο A.

Άσκηση: Έστω ότι η $\{\mu_\varepsilon\}$ ακολουθεί την LDP με συνάρτηση ταχύτητας I. Δείξτε ότι $\inf I(x) = 0$. Δείξτε τώρα ότι αν τα $\mu_\varepsilon \Rightarrow \delta_{x_0}$ τότε $I(x_0) = 0$.

Το θεώρημα του Cramér περιγράφει τις μεγάλες αποκλίσεις από το νόμο των μεγάλων αριθμών για την κατανομή μ_n του μέσου όρου μιας ακολουθίας ανεξάρτητων και ισόνομων τυχαίων μεταβλητών.

Θεώρημα 1 (Cramér): Η κατανομή μ_n του $n^{-1} \sum_{i=1}^{n} X_i$ ακολουθεί την LDP με συνάρτηση ταχύτητας:

$$I(x) = \sup_{\lambda} (\lambda x - \log M(\lambda)).$$

Απόδειξη: Η απόδειξη έχει ουσιαστικά δοθεί με τις (7), (8) και το λήμμα 1. Έστω F ένα κλειστό υποσύνολο του $\mathbb{R}$. Αν $\bar{x} \in F$ η (10) ικανοποιείται αφού $I(\bar{x}) = 0$. Αν $\bar{x} \notin F$, και (β, α) η συνεκτική συνιστώσα του F^c που περιέχει το $\bar{x}$ τότε η (10) προκύπτει είτε άμεσα από τις (7), (8), όταν το (β, α) είναι άπειρο είτε ως ακολούθως:

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) &\leq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n((-\infty, \beta] \cup [\alpha, +\infty)) \\
&= -\min\{I(\beta), I(\alpha)\} \\
&= -\inf_{x\in F} I(x),
\end{aligned}$$

από την ιδιότητα 5 της συνάρτησης ταχύτητας και το γεγονός ότι $\beta, \alpha \in F$.
Αν τώρα U είναι ένα ανοικτό υποσύνολο του $\mathbb{R}$ και $x \in U$ θεωρούμε $\delta > 0$ ώστε $(x-\delta, x+\delta) \subset U$ και από το λήμμα 1:

$$\liminf \frac{1}{n} \log \mu_n(U) \geq \liminf \frac{1}{n} \log \mu_n(x-\delta, x+\delta) \geq -I(x).$$

Μεγιστοποιώντας για $x \in U$ το δεξί σκέλος και λαμβάνουμε την (9).
□

Παρατήρηση: Το θεώρημα του Cramér γενικεύεται στον $\mathbb{R}^d$. Ορίζουμε την $M : \mathbb{R}^d \to [0, \infty]$ ως εξής:

$$M(\lambda) = \int e^{\lambda\cdot x}\, d\mu(x).$$

Αν $M(\lambda) < \infty$ για κάθε λ με $|\lambda| < \varepsilon$ η συνάρτηση ταχύτητας I που ελέγχει τις μεγάλες αποκλίσεις της κατανομής του μέσου όρου ανεξάρτητων τυχαίων διανυσμάτων από το νόμο των μεγάλων αριθμών δίνεται από την:

$$I(x) = \sup_{\lambda\in\mathbb{R}^d} (\lambda\cdot x - \log M(\lambda)).$$

Ακόμη κι αν η M δεν ικανοποιεί ούτε αυτή τη συνθήκη πάλι μπορεί να αποδειχθεί ότι για κάθε ανοικτό κυρτό $A \subset \mathbb{R}^d$ ισχύει:

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) = -\inf_{x\in A} I(x).$$

2 Μεγάλες αποκλίσεις για εξαρτημένες τυχαίες μεταβλητές

Μια προσεκτική ματιά στην απόδειξη του θεωρήματος του Cramér θα σας πείσει ότι η ανεξαρτησία των $X_1, \ldots, X_n$ δεν χρησιμοποιείται και τόσο αποφασιστικά. Μας βοήθησε να υπολογίσουμε την έκφραση

$$M_n(n\lambda) = \mathbb{E}\left[e^{\lambda\cdot\sum_{i=1}^n X_i}\right] = \int e^{n\lambda\cdot x}\, d\mu_n(x).$$

Τι γίνεται όμως αν μπορούμε να υπολογίσουμε το όριο:

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log M_n(n\lambda)$$

με κάποιον άλλο τρόπο; Επίσης χρησιμοποιήσαμε την ανεξαρτησία στην απόδειξη του κάτω φράγματος όπου επικαλεστήκαμε το κεντρικό οριακό θεώρημα για να αποδείξουμε ότι η $\mathbb{P}^*$ πιθανότητα ενός ενδεχομένου είναι φραγμένη μακριά από το μηδέν. Θα δούμε παρακάτω ότι μπορούμε να χρησιμοποιήσουμε το άνω φράγμα των μεγάλων αποκλίσεων για να δείξουμε ότι η πιθανότητα του συμπληρώματός του τείνει στο μηδέν. Σαν εφαρμογή θα μελετήσουμε μεγάλες αποκλίσεις μαρκοβιανών αλυσίδων.

Θεώρημα 2 (**Gärtner-Ellis**) Έστω ότι το όριο

$$\Lambda(\lambda) = \lim_{\varepsilon\to 0} \varepsilon \Lambda_\varepsilon(\frac{\lambda}{\varepsilon}) = \lim_{\varepsilon\to 0} \varepsilon \log \int e^{\frac{\lambda\cdot x}{\varepsilon}}\ d\mu_\varepsilon(x) \tag{11}$$

υπάρχει για κάθε $\lambda \in \mathbb{R}^d$ και είναι παραγωγίσιμη συνάρτηση του λ. Τότε η μ_ε ακολουθεί την LDP με συνάρτηση ταχύτητας

$$I(x) = \sup_\lambda(\lambda \cdot x - \Lambda(\lambda)).$$

Παρατήρηση: Η I είναι κάτω ημισυνεχής σαν supremum συνεχών συναρτήσεων. Είναι επίσης κυρτή σαν supremum γραμμικών συναρτήσεων και μη αρνητική αφού $\Lambda(0) = 0$. Μια και δουλεύουμε στον $\mathbb{R}^d$ όμως θα χρειαστεί να διαφοροποιήσουμε την απόδειξη του άνω φράγματος (10) αφού δεν μπορούμε να χρησιμοποιήσουμε το επιχείρημα με τη μονοτονικότητα της I. Επιπλέον θα υποθέσουμε ότι για κάθε $x \in \mathbb{R}^d$ υπάρχει $\lambda^* \in \mathbb{R}^d$ ώστε $I(x) = \lambda^* \cdot x - \Lambda(\lambda^*)$ (και θα περιγράψουμε σε μια άσκηση που ακολουθεί πώς μπορούμε να χειριστούμε τη γενική περίπτωση.) Η απόδειξη του θεωρήματος θα βασιστεί σε μια σειρά από λήμματα.

Λήμμα 4 Αν η $\{\mu_\varepsilon\}$ ικανοποιεί τις συνθήκες του θεωρήματος 2 τότε για κάθε συμπαγές σύνολο K

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K) \leq -\inf_{x\in K} I(x) \tag{12}$$

Απόδειξη: Έστω $x \in K$, $\lambda \in \mathbb{R}^d$ ώστε $I(x) = \lambda \cdot x - \Lambda(\lambda)$. Δοθέντος $\delta > 0$ θεωρούμε μια σφαίρα U_x με κέντρο το x και ακτίνα ρ ώστε $|\lambda|\rho \leq \delta$.

$$\mu_\varepsilon(U_x) \leq e^{\frac{-\lambda\cdot x + |\lambda|\rho}{\varepsilon}} \int_{U_x} e^{\frac{\lambda\cdot y}{\varepsilon}}\ d\mu_\varepsilon(y).$$

Επομένως,

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(U_x) \leq -\lambda \cdot x + |\lambda|\rho + \Lambda(\lambda) = -I(x) + \delta.$$

Η ένωση όλων των U_x για $x \in K$ είναι μια ανοικτή κάλυψη του K. Διαλέγουμε λοιπόν $x_1, \ldots, x_N$ ώστε $K \subset \bigcup_{i=1}^N U_{x_i}$, και έχουμε:

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K) &\leq \limsup_{\varepsilon\to 0}(N\varepsilon + \max_{1\leq i\leq N} \varepsilon \log \mu_\varepsilon(U_{x_i})) \\ &\leq \delta - \min_{1\leq i\leq N} I(x_i) \leq \delta - \inf_{x\in K} I(x).\end{aligned}$$

Εφόσον αυτό ισχύει για κάθε $\delta > 0$ ο ισχυρισμός του λήμματος είναι αληθής.
□

Ορισμός: Μια οικογένεια μέτρων πιθανότητας $\{\mu_\varepsilon\}$ ονομάζεται *εκθετικά ισοσυμπαγής* (exponentially tight) αν για κάθε $L > 0$ υπάρχει ένα συμπαγές σύνολο K_L ώστε:

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K_L^c) < -L.$$

Λήμμα 5 Αν

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(C) \leq -\inf_{x\in C} I(x) \tag{13}$$

για κάθε συμπαγές σύνολο C και η $\{\mu_\varepsilon\}$ είναι εκθετικά ισοσυμπαγής τότε η (13) ικανοποιείται για κάθε κλειστό σύνολο C.

Απόδειξη: Για κάθε $L > 0$ θεωρούμε συμπαγές K_L όπως στον προηγούμενο ορισμό. Το C είναι κλειστό άρα το $C \cap K_L$ είναι συμπαγές. Έτσι,

$$\begin{aligned}
\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(C) &\leq \max\{\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(C \cap K_L), \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K_L^c)\} \\
&\leq \max\{-\inf_{x\in C\cap K_L} I(x), -L\} \\
&\leq \max\{-\inf_{x\in C} I(x), -L\}.
\end{aligned}$$

Αφήνουμε τώρα το $L \to \infty$ για να ολοκληρώσουμε την απόδειξη.
□

Λήμμα 6 Αν $\Lambda(\lambda) < \infty$ για $\lambda = \pm e_1, \ldots, \pm e_d$ τότε η $\{\mu_\varepsilon\}$ είναι εκθετικά ισοσυμπαγής.

Απόδειξη: Αν $K_L = \bigcap_{i=1}^d \{-L \leq x_i \leq L\}$ είναι εύκολο να δούμε ότι:

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K_L^c) < -L + \max_{1\leq i\leq d} \Lambda(\pm e_i).$$

□

Από τα τρία προηγούμενα λήμματα έχουμε ότι το άνω φράγμα (10) ικανοποιείται για κάθε κλειστό σύνολο C. Αποδεικνύουμε και το κάτω φράγμα στο επόμενο λήμμα:

Λήμμα 7 Αν η $\{\mu_\varepsilon\}$ ικανοποιεί τις συνθήκες του θεωρήματος 2 τότε για κάθε ανοικτό σύνολο G έχουμε:

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(G) \geq -\inf_{x\in G} I(x)$$

Απόδειξη: Όπως και για το θεώρημα του Cramér αρκεί να δείξουμε ότι για κάθε σφαίρα $B_\delta(x)$ με κέντρο το x και ακτίνα $\delta > 0$

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(B_\delta(x)) \geq -I(x).$$

Έστω λ^* ώστε $I(x) = \lambda^* \cdot x - \Lambda(\lambda^*)$. Ορίζουμε τα νέα μέτρα πιθανότητας μ_ε^* ως εξής:

$$d\mu_\varepsilon^*(y) = \exp\left(\frac{\lambda^* \cdot y}{\varepsilon} - \Lambda_\varepsilon(\frac{\lambda^*}{\varepsilon})\right) d\mu_\varepsilon(y)$$

Όπως και στο προηγούμενο κεφάλαιο έπειτα από μερικούς υπολογισμούς έχουμε ότι για κάθε $0 < \rho < \delta$:

$$\begin{aligned}
\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(B_\delta(x)) &\geq \liminf_{\varepsilon\to 0} \left(\varepsilon \Lambda_\varepsilon(\frac{\lambda^*}{\varepsilon}) - \lambda^* \cdot x - |\lambda^*|\rho + \varepsilon \log \mu_\varepsilon^*(B_\rho(x))\right) &(14)\\
&= -I(x) - |\lambda^*|\rho + \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon^*(B_\rho(x)). &(15)
\end{aligned}$$

Τα μ_ε^* όμως ικανοποιούν τις συνθήκες του θεωρήματος 2 όπως μπορείτε εύκολα να διαπιστώσετε και

$$\Lambda^*(\lambda) = \Lambda(\lambda^* + \lambda) - \Lambda(\lambda^*),$$

οπότε και

$$I^*(y) = I(y) - \lambda^* \cdot y + \Lambda(\lambda^*) \geq \Lambda(\lambda^*) - \Lambda(\lambda^* + \delta z) + \delta z \cdot y.$$

Αν λοιπόν $I^*(y) = 0$ τότε διαιρώντας με δ και παίρνοντας $\delta \to 0$ έχουμε ότι για κάθε $z \in \mathbb{R}^d$:

$$z \cdot (y - \nabla\Lambda(\lambda^*)) \leq 0.$$

Τούτο μπορεί μόνο να συμβεί αν $y = \nabla\Lambda(\lambda^*) = x$. Χρησιμοποιώντας το άνω φράγμα του θεωρήματος 2 για τα μ_ε^* έχουμε ότι $\mu_\varepsilon^*(B_\rho^c(x)) \to 0$ και συνεπώς $\varepsilon \log \mu_\varepsilon^*(B_\rho(x)) \to 0$. Παίρνοντας τώρα $\rho \to 0$ στην

(15) ολοκληρώνουμε την απόδειξη του λήμματος και του θεωρήματος 2.
□

Άσκηση: Θα δούμε εδώ πώς μπορούμε να παρακάμψουμε την υπόθεση ότι το supremum στον ορισμό της I επιτυγχάνεται σε πεπερασμένο λ^*. Στη γενική περίπτωση θεωρήστε τα μέτρα πιθανότητας $\tilde{\mu}_\varepsilon = \mu_\varepsilon * \nu_{\varepsilon\delta}$, όπου $\nu_{\varepsilon\delta} = \mathcal{N}(0, \varepsilon\delta I_d)$. (Αν $\varepsilon = n^{-1}$ και μ_n είναι η κατανομή του μέσου όρου των τυχαίων μεταβλητών $X_1, \ldots, X_n$ αυτό ισοδυναμεί με το να θεωρήσουμε ανεξάρτητες (από τις X_i και μεταξύ τους) τυχαίες μεταβλητές Y_i με κατανομή $\mathcal{N}(0, \delta)$ και να μελετήσουμε το μέσο όρο των $\tilde{X}_i = X_i + Y_i$. Η τεχνική αυτή ονομάζεται ομαλοποίηση.)

$$\int e^{\lambda \cdot x}\ d\tilde{\mu}_\varepsilon = \int \int e^{\lambda \cdot (x+y)}\ d\mu_\varepsilon(x)\ d\nu_{\varepsilon\delta}(y)$$

συνεπώς $\tilde{\Lambda}(\lambda) = \Lambda(\lambda) + \delta|\lambda|^2/2$. Το supremum στον ορισμό της $\tilde{I}$ λαμβάνεται σε πεπερασμένο σημείο και άρα η απόδειξή μας είναι εντάξει. Δείτε τώρα ότι από τα (9) και (10) για τα $\tilde{\mu}_\varepsilon$ μπορούμε να συνάγουμε τις ίδιες ανισότητες για τα μ_ε παίρνοντας $\delta \to 0$.

Αξίζει να δούμε πώς μεταφράζεται το θεώρημα Gärtner-Ellis στην περίπτωση που μ_n είναι η κατανομή του μέσου όρου τυχαίων μεταβλητών -όχι κατ΄ ανάγκη ανεξάρτητων.

Πόρισμα 1 Αν το όριο

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{n\lambda \cdot x}\ d\mu_n(x) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{\lambda \cdot (X_1 + \ldots + X_n)}\right)$$

υπάρχει και είναι παραγωγίσιμη συνάρτηση του λ τότε η $\{\mu_n\}$ ακολουθεί την LDP με συνάρτηση ταχύτητας:

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

Ειδικώτερα, αν $\inf_{x \in \mathring{A}} I(x) = \inf_{x \in \bar{A}} I(x) = c(A)$ τότε

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{X_1 + X_2 + \ldots + X_n}{n} \in A\right) = -c(A).$$

Άσκηση: Μια ανέλιξη διακριτού χρόνου ονομάζεται *στάσιμη* αν για κάθε $k \in \mathbb{Z}$ και $n \in \mathbb{N}$ η κατανομή του διανύσματος $(X_1, \ldots, X_n)$ είναι ίδια με αυτήν του διανύσματος $(X_{k+1}, \ldots, X_{k+n})$. Μια ανέλιξη ονομάζεται ανέλιξη Gauss αν για κάθε k η κατανομή του $(X_1, \ldots, X_k)$ είναι κανονική. Έστω λοιπόν μια στάσιμη πραγματική ανέλιξη Gauss με $\mathbb{E}(X_i = 0)$ και ακολουθία συσχετίσεων $\rho_i = \mathbb{E}(X_n X_{n+i})$. Η ‘ίσχύς ” της ορίζεται ως:

$$P = \lim_{n\to\infty} \sum_{i=-n}^{n} (1 - \frac{|i|}{n})\rho_i.$$

Δείξτε ότι αν η ισχύς της είναι πεπερασμενη, τότε η κατανομή του μέσου όρου των $\{X_1, \ldots, X_n\}$ ακολουθεί την LDP με συνάρτηση ταχύτητας $I(x) = x^2/2P$.

Εφαρμογή: Μεγάλες αποκλίσεις για προσθετικά συναρτησοειδή μαρκοβιανών αλυσίδων.
Έστω $X_0, X_1, \ldots$ μια μαρκοβιανή αλυσίδα σ΄ ένα πεπερασμένο χώρο V με πίνακα μετάβασης $\Pi = (\pi(x,y))_{(x,y)\in V^2}$ και f μια συνάρτηση $f: V \to \mathbb{R}^d$. Θα υποθέσουμε ότι ο Π είναι ανάγωγος, δηλαδή για κάθε $x, y \in V$ υπάρχουν $y_1, \ldots, y_k \in V$ ώστε $\pi(x, y_1) > 0, \ldots, \pi(y_k, y) > 0$. Αυτό σημαίνει ότι η μαρκοβιανή αλυσίδα μπορεί να πάει από οποιαδήποτε κατάσταση σε οποιαδήποτε άλλη σε πεπερασμένο αριθμό βημάτων. Θα χρησιμοποιήσουμε το θεώρημα Gärtner-Ellis για να δείξουμε ότι ο εμπειρικός μέσος όρος της f:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

ακολουθεί την LDP και θα υπολογίσουμε τη συνάρτηση ταχύτητας. Πράγματι αρκεί να υπολογίσουμε το όριο:

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}^x \left(\exp(\lambda \cdot \sum_{i=1}^{n} f(X_i)) \right),$$

και να δείξουμε ότι είναι πεπερασμένη και παραγωγίσιμη συνάρτηση του λ. Έχουμε λοιπόν:

$$\begin{aligned} \mathbb{E}^x \left(\exp(\lambda \cdot \sum_{i=1}^{n} f(X_i)) \right) &= \sum_{v_1,\ldots,v_n \in V} \exp(\lambda \cdot \sum_{i=1}^{n} f(v_i)) \mathbb{P}^x(X_1 = v_1, \ldots, X_n = v_n) \\ &= \sum_{v_1,\ldots,v_n \in V} \prod_{i=1}^{n} \pi(v_{i-1}, v_i) \exp(\lambda \cdot f(v_i)) \\ &= \sum_{y \in V} (\Pi_\lambda)^n(x, y), \end{aligned}$$

όπου Π_λ είναι ο πίνακας με στοιχεία $\pi_\lambda(x, y) = \pi(x, y)e^{\lambda \cdot f(y)}$, ενώ χρησιμοποιήσαμε τη σύμβαση $v_0 = x$. Φυσικά ο πίνακας Π_λ είναι και αυτός ανάγωγος έτσι από το θεώρημα Perron-Frobenius έχει μια πραγματική απλή ιδιοτιμή $\rho(\lambda) > 0$ ώστε για κάθε άλλη ιδιοτιμή γ του Π_λ έχουμε $|\gamma| < \rho(\lambda)$ ενώ το ιδιοδιάνυσμα u που αντιστοιχεί στην λ έχει αυστηρά θετικές συντεταγμένες. Έχουμε τώρα:

$$\sum_{y \in V} (\Pi_\lambda)^n(x, y) \le \sum_{y \in V} (\Pi_\lambda)^n(x, y) \frac{u(y)}{\inf_{z \in V} u(z)} = \frac{\rho^n(\lambda)u(x)}{\inf_{z \in V} u(z)}.$$

Ομοίως:

$$\sum_{y \in V} (\Pi_\lambda)^n(x, y) \ge \sum_{y \in V} (\Pi_\lambda)^n(x, y) \frac{u(y)}{\sup_{z \in V} u(z)} = \frac{\rho^n(\lambda)u(x)}{\sup_{z \in V} u(z)}.$$

οπότε $\Lambda(\lambda) = \rho(\lambda) < \infty$. Επιπλέον, η Λ είναι παραγωγίσιμη αφού οι συντελεστές του χαρακτηριστικού πολυωνύμου του Π_λ είναι ομαλές συναρτήσεις του λ και η $\rho(\lambda)$ είναι απομονωμένη ρίζα (θεώρημα πεπλεγμένης συνάρτησης). Επομένως από το θεώρημα Gärtner-Ellis έχουμε:

Θεώρημα 3 Η κατανομή του εμπειρικού μέσου της f:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

ακολουθεί την LDP με συνάρτηση ταχύτητας $I(x) = \sup_{\lambda \in \mathbb{R}^d} (\lambda \cdot x - \log \rho(\lambda))$.

Άσκηση: Αν διαλέξουμε $f : V \to \{0, 1\}^{|V|}$ με $f(x) = (\delta_{y_1}(x), \delta_{y_2}(x), \ldots, \delta_{y_{|V|}}(x))$, τότε το S_n είναι είναι ένα μέτρο στον V που ονομάζουμε 'έμπειρική κατανομή της αλυσίδασ":

$$S_n(x) = \frac{\text{Αριθμός επισκέψεων στην κατάσταση } x \text{ μέχρι το χρόνο } n}{n}.$$

Γράφουμε $u >> 0$ αν $u(x) > 0, \forall x \in V$. Τότε η συνάρτηση ταχύτητας δίνεται από τη μεταβολική σχέση:

$$I(\mu) = \sup_{u>>0} \int \log \left(\frac{u(x)}{(u\Pi)(x)} \right) \, d\mu(x).$$

3 Το λήμμα του Varadhan

Στον υπολογισμό ολοκληρωμάτων, πολύ μεγάλες τιμές που λαμβάνονται σ' ένα "μικρό" μέρος του χώρου μας παίζουν συχνά καθοριστικό ρόλο στη συμπεριφορά του ολοκληρώματος. Στο κεφάλαιο αυτό θα δείξουμε άλλη μια εφαρμογή της θεωρίας των μεγάλων αποκλίσεων στη ασυμπτωτική μελέτη ολοκληρωμάτων. Το επόμενο λήμμα (που οφείλεται στον Laplace) δεν είναι δύσκολο να αποδειχθεί:

Λήμμα 8 (Laplace): Αν η f είναι μια συνεχής συνάρτηση ορισμένη στο [0,1] τότε:

$$\lim_{\varepsilon\to 0}\varepsilon\log\int_0^1 e^{\frac{f(x)}{\varepsilon}}\,dx=\sup_{x\in[0,1]}f(x).$$

Το επόμενο λήμμα γενικεύει το προηγούμενο αποτέλεσμα στην περίπτωση που τα ολοκληρώματα δεν υπολογίζονται ως προς το ίδιο μέτρο, αλλά ως προς μια οικογένεια μέτρων που ακολουθεί την (LDP). Μια συνάρτηση ταχύτητας I χαρακτηρίζεται "καλή" αν για κάθε $M>0$ τα σύνολα $\Psi_I(M)\{x:I(x)\leq M\}$ είναι συμπαγή.

Λήμμα 9 (Varadhan): Έστω οικογένεια μέτρων πιθανότητας $\{\mu_\varepsilon\}$ που ακολουθεί την (LDP) με μια "καλή" συνάρτηση ταχύτητας I και Φ μια άνω φραγμένη συνεχής συνάρτηση στον $\mathbb{X}$. Τότε:

$$\lim_{\varepsilon\to 0}\varepsilon\log\int e^{\frac{\Phi(x)}{\varepsilon}}\,d\mu_\varepsilon(x)=\sup_{x\in\mathbb{X}}(\Phi(x)-I(x)).$$

Απόδειξη 1.Κάτω φράγμα: Από τη συνέχεια της Φ για κάθε $x\in\mathbb{X}$ και $\delta>0$ υπάρχει περιοχή U_x του x τέτοια ώστε: $\Phi(y)>\Phi(x)-\delta$, για κάθε $y\in U_x$. Έτσι,

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

Επομένως,

$$\liminf_{\varepsilon\to 0}\varepsilon\log\int e^{\frac{\Phi(y)}{\varepsilon}}\,d\mu_\varepsilon(y)\geq\Phi(x)-\delta-\inf_{U_x}I(y)\geq\Phi(x)-I(x)-\delta.$$

Παίρνοντας $\delta\to 0$ και το supremum για $x\in\mathbb{X}$ στο δεξί σκέλος, έχουμε το κάτω φράγμα.
2. Άνω φράγμα: Για κάθε $x\in\mathbb{X}$ υπάρχει (από τη συνέχεια της Φ και την κάτω ημισυνέχεια της I) περιοχή $U_x\ni x$ ώστε

$$\Phi(y)<\Phi(x)+\delta \text{ και } I(y)>I(x)-\delta,\ \text{για κάθε } y\in\bar{U}_x.$$

Για κάθε $M>0$ το σύνολο $\Psi_I(M)$ είναι συμπαγές και η $\cup_{x\in\Psi_I(M)}U_x$ είναι μια ανοικτή κάλυψή του. Ας θεωρήσουμε λοιπόν μια πεπερασμένη υποκάλυψη $\{U_{x_i}\}_{1\leq i\leq N}$ και το κλειστό σύνολο $F=(\cup_{i=1}^N U_{x_i})^c$.

$$\begin{aligned}\int e^{\frac{\Phi(y)}{\varepsilon}}\,d\mu_\varepsilon(y) &= \sum_{i=1}^N\int_{U_{x_i}}e^{\frac{\Phi(y)}{\varepsilon}}\,d\mu_\varepsilon(y)+\int_F e^{\frac{\Phi(y)}{\varepsilon}}\,d\mu_\varepsilon(y)\\ &\leq \sum_{i=1}^N\mu_\varepsilon(U_{x_i})\exp\left(\frac{\Phi(x_i)+\delta}{\varepsilon}\right)+\mu_\varepsilon(F)\exp\left(\frac{\sup\Phi}{\varepsilon}\right)\end{aligned}$$

Από το άνω φράγμα των ανισοτήτων των μεγάλων αποκλίσεων (10) λοιπόν:

$$\begin{aligned}\limsup_{\varepsilon\to 0}\varepsilon\log\int e^{\frac{\Phi(y)}{\varepsilon}}\,d\mu_\varepsilon(y) &\leq \max\{\Phi(x_i)+\delta-\inf_{\bar{U}_{x_i}}I(y),(\sup\Phi-\inf_{x\in F}I(x))\}\\ &\leq \max\{\Phi(x_i)-I(x_i)+2\delta,(\sup\Phi-M)\}\\ &\leq \max\{\sup_{x\in\mathbb{X}}(\Phi(x)-I(x)),\sup\Phi-M\}+2\delta.\end{aligned}$$

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

Είναι ενδιαφέρον οτι ισχύει και το αντίστροφο του παραπάνω λήμματος:

Λήμμα 10 (Bryc): Έστω οικογένεια μέτρων πιθανότητας $\{\mu_\varepsilon\}$ εκθετικά ισοσυμπαγής, και το όριο

$$\Lambda(f) = \lim_{\varepsilon\to 0} \varepsilon \log \int e^{\frac{f(x)}{\varepsilon}} \, d\mu_\varepsilon$$

υπάρχει για κάθε $f \in C_b(\mathbb{X})$. Τότε η $\{\mu_\varepsilon\}$ ακολουθεί την LDP με συνάρτηση ταχύτητας:

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

Στην απόδειξη του λήμματος 1 είπαμε πως με την αλλαγή μέτρου ώστε το x να είναι η μέση τιμή της νέας κατανομής καταστήσαμε τυπική την προηγουμένως άτυπη συμπεριφορά (σύγκλιση του μέσου όρου στο x.) Το γιατί όμως από όλες τις κατανομές με μέση τιμή x διαλέξαμε τη συγκεκριμένη παραμένει ανεξήγητο. Αυτό θα προσπαθήσουμε να καταλάβουμε στην επόμενη διάλεξη, στην οποία θα μελετήσουμε το θεώρημα του Sanov. Σαν προεργασία όμως προσπαθήστε να αποδείξετε ότι:

$$I(x) = \inf_{\int y d\nu(y)=x} H(\nu|\mu),$$

όπου $H(\nu|\mu)$ είναι η *σχετική εντροπία* της κατανομής ν ως προς την κατανομή μ που ορίζεται ως εξής:

$$H(\nu|\mu) = \begin{cases} \int f \log f \, d\mu & ,\text{αν } d\nu = f d\mu \text{ και } \nu(\mathbb{R}) = 1, \\ \infty & ,\text{διαφορετικά.} \end{cases}$$

(Η $H(\cdot|\mu)$ είναι κυρτή συνάρτηση.) Θα δείτε ότι η καινούργια κατανομή που χρησιμοποιήσαμε είναι το μέτρο στο οποίο επιτυγχάνεται το ελάχιστο στο παραπάνω μεταβολικό πρόβλημα.

4 Το θεώρημα του Sanov και η αρχή της συστολής

Ας ξαναγυρίσουμε στην κατάσταση όπου οι $X_1, X_2, \ldots$ είναι ανεξάρτητες ισόνομες τυχαίες μεταβλητές με κοινή κατανομή μ. Ορίζουμε:

$$\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}.$$

Η ν_n είναι μια τυχαία μεταβλητή με τιμές στο σύνολο $\mathcal{M}_1$ των μέτρων πιθανότητας ορισμένων στον $\mathbb{R}$. Μπορούμε να δούμε τον $\mathcal{M}_1$ σαν υποσύνολο του χώρου $\mathcal{M}$ των πεπερασμένων μέτρων στον $\mathbb{R}$. Αν $C_b(\mathbb{R})$ είναι ο χώρος των συνεχών φραγμένων συναρτήσεων στον $\mathbb{R}$ τότε από ένα γνωστό θεώρημα (Riesz) της Συναρτησιακής Ανάλυσης ο $\mathcal{M}$ είναι ο δυϊκός του $C_b(\mathbb{R})$ και αν εφοδιάσουμε τον $\mathcal{M}$ με την w*-τοπολογία, έχουμε $\mathcal{M}^* = C_b(\mathbb{R})$. Κατ΄ αναλογία με την συζήτησή μας για το θεώρημα Gärtner-Ellis υπολογίζουμε για κάθε $f \in C_b(\mathbb{R})$ την:

$$\Lambda(f) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left(e^{n\langle f,\nu_n\rangle}\right) = \log \int e^{f(x)} \, d\mu(x).$$

Μάλιστα η $\Lambda : \mathcal{M}^* \to \mathbb{R}$ είναι (Gateaux)-παραγωγίσιμη:

$$\delta_g \Lambda(f) = \frac{d}{d\varepsilon}\Lambda(f+\varepsilon g)\bigg|_{\varepsilon=0} = \frac{\int g e^f d\mu}{\int e^f d\mu}.$$

Από μια επέκταση του θεωρήματος Gärtner-Ellis (που δεν θα δείξουμε εδώ) έχουμε ότι η κατανομή των ν_n ακολουθεί την LDP με συνάρτηση ταχύτητας:

$$I(\nu) = \sup_{f \in \mathcal{C}_b(\mathbb{R})} (\langle f, \nu \rangle - \log \int e^f \, d\mu). \tag{16}$$

Λήμμα 11 $\quad I(\nu) = H(\nu|\mu)$.

Απόδειξη: Για να δείξουμε ότι $I(\nu) \leq H(\nu|\mu)$ αρκεί να θεωρήσουμε ν τέτοιο ώστε $H(\nu|\mu) < \infty$, οπότε $d\nu = h d\mu$, με $\int h d\mu = 1$. Από τη στοιχειώδη ανισότητα $xy \leq y \log y + e^{x-1}$ (οι $xlogx$ και e^{x-1} είναι ζευγάρι μετασχηματισμού Legendre) έχουμε για κάθε $f \in C_b(\mathbb{R})$:

$$\langle f, \nu \rangle \leq \int h \log h \, d\mu + \int e^{f-1} \, d\mu,$$

άρα εφαρμόζοντας την ανισότητα για $f - c$, όπου $c \in \mathbb{R}$ παίρνουμε:

$$\langle f, \nu \rangle = \langle f - c, \nu \rangle + c \leq \int h \log h \, d\mu + e^{-(1+c)} \int e^f \, d\mu + c.$$

Ελαχιστοποιώντας το δεξί σκέλος ως προς c τέλος, λαμβάνουμε

$$\langle f, \nu \rangle \leq H(\nu|\mu) + \log \int e^f \, d\mu.$$

Η παραπάνω ισχύει για κάθε $f \in C_b(\mathbb{R})$ άρα $I(\nu) \leq H(\nu|\mu)$.
Αντικαθιστώντας την f με $f + c$ στην (16) εύκολα βλέπουμε ότι $H(\nu|\mu) = \infty$, αν $\nu(\mathbb{R}) \neq 1$. Ομοίως αν $\mu(A) = 0$ αλλά $\nu(A) > 0$, παίρνουμε στην (16) $f = c\chi(A)$ και κατόπιν $c \to \infty$. Τέλος αν $d\nu = h d\mu$ προσεγγίζουμε την $\log(h)$ με συνεχείς και φραγμένες συναρτήσεις.
□
Συνοψίζοντας, μπορούμε να διατυπωσουμε το ακόλουθο θεώρημα.

Θεώρημα 4 (Sanov):

$$-\inf_{\nu \in \mathring{\Gamma}} H(\nu|\mu) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\nu_n \in \Gamma) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\nu_n \in \Gamma) \leq -\inf_{\nu \in \bar{\Gamma}} H(\nu|\mu).$$

Θα κλείσουμε αυτές τις διαλέξεις επιστρέφοντας εκεί απ΄ όπου αρχίσαμε, στο θεώρημα του Cramér. Παρατηρήστε:

$$S_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \langle i, \nu_n \rangle,$$

όπου $i(x) = x$ είναι η ταυτοτική συνάρτηση. Η αρχή της συστολής περιγράφει πώς μετασχηματίζεται η αρχή της μεγάλης απόκλισης από συνεχείς συναρτήσεις:

Θεώρημα 5 Αρχή της συστολής: Αν η οικογένεια μέτρων πιθανότητας $\{\mu_\varepsilon\}$ στο χώρο X ακολουθεί την LDP με συνάρτηση ταχύτητας I και f είναι μια συνεχής συνάρτηση $f : X \to Y$, τότε η οικογένεια $\{\mu_\varepsilon \circ f^{-1}\}$ ακολουθεί την LDP με συνάρτηση ταχύτητας

$$I'(y) = \inf\{I(x) : y = f(x)\}.$$

Η απόδειξη είναι στην ουσία ταυτολογία, δοκιμάστε να την γράψετε. Το θεώρημα του Cramér προκύπτει ευριστικά από το θεώρημα του Sanov χρησιμοποιώντας την αρχή της συστολής με $f : \mathcal{M}_1(\mathbb{R}) \to \mathbb{R}$ ώστε $f(\nu) = \langle i, \nu \rangle$. Βέβαια, η f ΔΕΝ είναι συνεχής συνάρτηση, μπορούμε όμως να αποκόψουμε την ταυτοτική συνάρτηση σε κάποιο πεπερασμένο επίπεδο, και ν΄ αφήσουμε στη συνέχεια αυτό να τείνει στο άπειρο, παίρνοντας το θεώρημα του Cramér με συνάρτηση ταχύτητας:

$$I(x) = \inf_{\nu:\int y d\nu(y)=x} H(\nu|\mu).$$

Αυτός είναι και ο λόγος που επέβαλε τη συγκεκριμένη αλλαγή μέτρου στην απόδειξή μας. Η εντροπία ελέγχει τις μεγάλες αποκλίσεις σ΄ ένα ψηλότερο επίπεδο- αυτό των εμπειρικών μέτρων. Θέλαμε την κατανομή με μέση τιμή x και την ελάχιστη εντροπία ως προς την αρχική κατανομή.

Βιβλιογραφία: Το βιβλίο που συμβουλεύτηκα περισσότερο γράφοντας αυτές τις σημειώσεις είναι το:

Dembo- Zeitouni, Large Deviations Techniques and Applications, Springer.

Από εκεί είναι διαλεγμένες και οι περισσότερες ασκήσεις. Άλλα καλά, αλλά πιο προχωρημένα βιβλία είναι:

Varadhan, Large Deviations and Applications, SIAM και

Deuschel- Stroock, Large Deviations, Academic Press.

Για τους ελέγχους Neyman-Pearson μπορείτε αν θέλετε να διαβάσετε στις σημειώσεις Στατιστικής που μπορείτε να βρείτε στην ιστοσελίδα:

www.math.nyu.edu/faculty/varadhan.

Για τα αποτελέσματα από την Ανάλυση που επικαλεστήκαμε μπορείτε να συμβουλευτείτε το

Folland, Real Analysis, Modern Techniques and Applications.

Τέλος το θεώρημα Perron-Frobenius μπορείτε να το διαβάσετε στο:

Senata, Non negative matrices and Markov Chains,Springer.

Ηράκλειο, Ιούλιος 2005,
Μιχάλης Λουλάκης.