Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Σχολή Θετικών Επιστημών

Τμήμα Μαθηματικών

ΠΜΣ – ΜΔΕ «Εφαρμοσμένα Μαθηματικά»

Μεταπτυχιακή Εργασία:

‘‘Θεωρία Μεγάλων Αποκλίσεων και Εφαρμογές’’

Μαθιουδάκης Ιωάννης

ΠΕΡΙΛΗΨΗ

Η θεωρία των μεγάλων αποκλίσεων ασχολείται με τις πιθανότητες των σπάνιων γεγονότων (ή διακυμάνσεων) που είναι εκθετικά μικρές ως συνάρτηση κάποιας παραμέτρου, π.χ., ο αριθμός των τυχαίων συστατικών ενός συστήματος, ο χρόνος κατά τον οποίο ένα στοχαστικό σύστημα παρατηρείται, το πλάτος του θορύβου, ο οποίος διαταράσσει ένα δυναμικό σύστημα ή την θερμοκρασία μιας χημικής αντίδρασης. Η θεωρία έχει εφαρμογές σε πολλούς διαφορετικούς επιστημονικούς τομείς, που κυμαίνονται από την θεωρία ουρών στη στατιστική και από τις οικονομικές επιστήμες στις επιστήμες των μηχανικών. Επίσης, χρησιμοποιείται όλο και περισσότερο στη στατιστική φυσική για τη μελέτη ισορροπημένων και μη-ισορροπημένων συστημάτων. Σε αυτό το πλαίσιο, βαθιές αναλογίες μπορούν να γίνουν μεταξύ γνωστών εννοιών της στατιστικής φυσικής, όπως η εντροπία και η ελεύθερη ενέργεια, και εννοιών της θεωρίας μεγάλων αποκλίσεων, οι οποίες έχουν περισσότερο τεχνικές ονομασίες, όπως η συνάρτηση ρυθμού και η ανηγμένη αθροιστική γεννήτρια συνάρτηση. Το πρώτο κεφάλαιο αυτής της μεταπτυχιακής εργασίας παρουσιάζει τα βασικά στοιχεία της θεωρίας των μεγάλων αποκλίσεων στα μαθηματικά. Εστιάζουμε στις απλές αλλά πολύ σημαντικές ιδέες πίσω από τη συγκεκριμένη θεωρία, αναφέρεται σε μη τεχνικούς όρους, και για την εφαρμογή αυτών των ιδεών σε απλές στοχαστικές διαδικασίες, όπως οι δειγματικοί μέσοι των ανεξάρτητων και όμοια κατανεμημένων τυχαίων μεταβλητών και οι διαδικασίες Markov. Στο δεύτερο κεφάλαιο, το αριθμητικό πρόβλημα της αξιολόγησης πιθανοτήτων μεγάλης απόκλισης αντιμετωπίζεται σε βασικό επίπεδο. Η βασική ιδέα της δειγματοληψίας σημαντικότητας εισάγεται εκεί μαζί με την συγγενική της ιδέα, την εκθετική αλλαγή του μέτρου. Επίσης, υπάρχει μια βασική αναφορά σχετικά με τις εφαρμογές της θεωρίας των μεγάλων αποκλίσεων στη φυσική, τις οικονομικές επιστήμες και την ασφάλιση. Τέλος, στο τρίτο κεφάλαιο, θα επανεξετάσουμε τη θεωρία των μεγάλων αποκλίσεων για στοχαστικές διαφορικές εξισώσεις στο όριο μικρού θορύβου.

ABSTRACT

The theory of large deviations deals with the probabilities of rare events (or fluctuations) that are exponentially small as a function of some parameter, e.g., the number of random components of a system, the time over which a stochastic system is observed, the amplitude of the noise perturbing a dynamical system or the temperature of a chemical reaction. The theory has applications in many different scientific fields, ranging from queuing theory to statistics and from finance to engineering. It is also increasingly used in statistical physics for studying both equilibrium and nonequilibrium systems. In this context, deep analogies can be made between familiar concepts of statistical physics, such as the entropy and the free energy, and concepts of large deviation theory having more technical names, such as the rate function and the scaled cumulant generating function. The first part of this Master thesis introduces the basic elements of large deviation theory in mathematics. The focus there is on the simple but powerful ideas behind large deviation theory, stated in non-technical terms, and on the application of these ideas in simple stochastic processes, such as sums of independent and identically distributed random variables and Markov processes. In the second part, the problem of numerically evaluating large deviation probabilities is treated at a basic level. The fundamental idea of importance sampling is introduced there together with its sister idea, the exponential change of measure. Also, there is a basic reference about the applications of large deviation theory in physics, finance and insurance. Finally, in the third part, we revisit large deviation theory for stochastic differential equations in the small-noise limit.

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1º

Εισαγωγή στη Θεωρία των Μεγάλων Αποκλίσεων

1.1 Εισαγωγή

Η περιοχή των μεγάλων αποκλίσεων είναι ένα σύνολο ασυμπτωτικών αποτελεσμάτων στις πιθανότητες σπάνιων γεγονότων και ένα σύνολο μεθόδων, οι οποίες παράγουν τέτοια αποτελέσματα. Οι μεγάλες αποκλίσεις είναι μία πολύ ενεργή περιοχή των εφαρμοσμένων πιθανοτήτων και σε σχετικά ερωτήματα ακραίων γεγονότων στις οικονομικές και στις ασφαλιστικές εφαρμογές, παίζει ένα αυξανόμενα σημαντικό ρόλο. Για παράδειγμα, πρόσφατες εφαρμογές έντονου ενδιαφέροντος αφορούν πιθανότητες καταστροφής (χρεοκοπίας) στη θεωρία κινδύνου (διαχείριση κινδύνου), αποτίμηση δικαιωμάτων προαίρεσης, καθώς επίσης και στη στατιστική μηχανική.

Η θεωρία των Μεγάλων Αποκλίσεων εμφανίστηκε ιστορικά στα ασφαλιστικά μαθηματικά με το πρόβλημα εκτίμησης της πιθανότητας χρεοκοπίας στο κλασικό μοντέλο Cramér – Lundberg. Το πρόβλημα ακολούθως επεκτάθηκε σε πιο γενικά μοντέλα, όπως για παράδειγμα στις διαδικασίες Lévy. Στα οικονομικά, η θεωρία των Μεγάλων Αποκλίσεων εμφανίζεται σε διάφορες περιπτώσεις. Λαμβάνουν χώρα στη διαχείριση ακραίου κινδύνου για τον υπολογισμό της πιθανότητας μεγάλων απωλειών ενός κεφαλαίου επενδύσεων. Οι μέθοδοι των Μεγάλων Αποκλίσεων χρησιμοποιούνται ευρέως στη προσομοίωση σπάνιων γεγονότων και εμφανίζονται στην προσέγγιση αποτίμησης δικαιωμάτων προαίρεσης.

Επίσης, η συγκεκριμένη θεωρία συνδέεται με την εκθετική μείωση των πιθανοτήτων μεγάλων διακυμάνσεων σε τυχαία συστήματα. Αυτές οι πιθανότητες είναι σημαντικές για πολλά επιστημονικά πεδία όπως η στατιστική, τα οικονομικά και η επιστήμη των μηχανικών, καθώς αποδίδουν αξιόλογες πληροφορίες για τις μεγάλες διακυμάνσεις ενός τυχαίου συστήματος γύρω από τη πιο πιθανή κατάστασή του ή τροχιά του. Στο επίπεδο της στατιστικής μηχανικής σε ισορροπία, η θεωρία των Μεγάλων Αποκλίσεων ορίζει εκτιμήσεις εκθετικής τάξεως των πιθανοτήτων, οι οποίες βελτιώνουν και γενικεύουν τη θεωρία των διακυμάνσεων του Einstein. Είναι μία θεωρία συλλογής τεχνικών υπολογισμού εντροπιών και ελεύθερων ενεργειών και μία ακριβής έκφραση προσεγγίσεων σαγματικών σημείων στη στατιστική μηχανική. Το συμπέρασμα είναι ότι οι πιθανότητες μπορούν να εκφραστούν σε σχέση με τις συναρτήσεις εντροπίας και η θεωρία αυτή περιέχεται στην αρχή της μεγάλης απόκλισης, όπου η εντροπία καλείται ως συνάρτηση ρυθμού.

Ας αναφέρουμε ένα παράδειγμα για να εξηγήσουμε τον στόχο αυτής της εισαγωγής. Έστω **X** μία τυχαία μεταβλητή, η οποία παίρνει πραγματικές τιμές, και θεωρούμε το πρόβλημα υπολογισμού ή εκτίμησης της $P[\boldsymbol{X} > l]$, της πιθανότητας όπου το **X** υπερβαίνει κάποιο επίπεδο l. Στα οικονομικά, μπορούμε να θεωρήσουμε το **X** ως την απώλεια ενός επενδυτικού κεφαλαίου στην αγορά υψηλού κινδύνου, και να ενδιαφερόμαστε για τη πιθανότητα μεγάλης ή πλήρους απώλειας. Για να εκτιμήσουμε το $p = P[\boldsymbol{X} > l]$, μία βασική τεχνική είναι η προσομοίωση Monte Carlo: Δημιουργούμε n ανεξάρτητες τιμές $\mathbf{X_1, \ldots\ldots, X_n}$ του **X**, και χρησιμοποιούμε τον δειγματικό μέσο:

$$\boldsymbol{\bar{S}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \mu\varepsilon\ Y_i = 1_{X_i > l}}.$$

Η σύγκλιση αυτής της εκτίμησης (όταν $n \to \infty$) απορρέει από το νόμο των Μεγάλων Αριθμών, ενώ ο ρυθμός σύγκλισης δίνεται, μέσω του Κεντρικού Οριακού Θεωρήματος, σε σχέση με τη διασπορά $v = p(1-p)$ του Y_i:

$$\boldsymbol{P\left[|\bar{S}_n - p| \geq \frac{a}{\sqrt{n}}\right] \to 2\Phi\left(-\frac{a}{\sqrt{v}}\right)},$$

όπου Φ είναι η συνάρτηση αθροιστικής κατανομής της κανονικής κατανομής. Αυτή η σύγκλιση της εκτιμήτριας $\bar{S}_n$ προσδιορίζεται επακριβώς με το αποτέλεσμα των μεγάλων αποκλίσεων, γνωστό ως Θεώρημα του Cramér, το οποίο αφορά τη προσέγγιση των πιθανοτήτων σπάνιων γεγονότων $P[\bar{S}_n \epsilon A]$, και τυπικά ισχύει ότι

$$\boldsymbol{P[|\bar{S}_n - p \geq a| \cong Ce^{-\gamma n}},$$

για κάποιες σταθερές C και γ.

Όπως αναφέραμε πιο πάνω, ο ρυθμός της σύγκλισης της εκτιμήτριας $\bar{S}_n$ ορίζεται ως εξής:

$$\boldsymbol{Var(\bar{S}_n) = \frac{Var(1_{X>l})}{n} = \frac{p(1-p)}{n}},$$

και το σχετικό σφάλμα είναι

$$\boldsymbol{relative\ error = \frac{\tau\upsilon\pi\iota\kappa\acute{\eta}\ \alpha\pi\acute{o}\kappa\lambda\iota\sigma\eta\ \tau o\upsilon\ \bar{S}_n}{\mu\acute{\varepsilon}\sigma\eta\ \tau\iota\mu\acute{\eta}\ \tau o\upsilon\ \bar{S}_n} = \frac{\sqrt{p(1-p)}}{\sqrt{n}p}}.$$

Γι' αυτό το λόγο, εάν $p = P[\boldsymbol{X} > l]$ είναι μικρό, και ενώ $\sqrt{p - p^2}\big/p \to \infty$ καθώς το p πηγαίνει στο μηδέν, βλέπουμε ότι χρειάζεται ένα μεγάλο δείγμα έτσι ώστε η εκτιμήτρια να αποδώσει ένα λογικό σχετικό σφάλμα. Αυτή είναι μία συνηθισμένη παρατήρηση όταν εκτιμούμε σπάνια γεγονότα. Για να βελτιώσουμε την εκτίμηση της ουράς της πιθανότητας $P[\boldsymbol{X} > l]$, χρησιμοποιούμε τη μέθοδο της δειγματοληψίας σημαντικότητας για να μειώσουμε τη διακύμανση, και έτσι επιταχύνουμε τον υπολογισμό απαιτώντας λιγότερα δείγματα. Αυτό βασίζεται στο γεγονός ότι δίνουμε μεγαλύτερο βάρος σε «σημαντικά» αποτελέσματα. Αφού η θεωρία Μεγάλων Αποκλίσεων ασχολείται με τα σπάνια γεγονότα, μπορούμε να δούμε τη σχέση που έχει με τη παραπάνω μέθοδο δειγματοληψίας.

1.2 Παραδείγματα αποτελεσμάτων της Θεωρίας Μεγάλων Αποκλίσεων

Πριν εμβαθύνουμε στη θεωρία των Μεγάλων Αποκλίσεων, είναι χρήσιμο να δούμε μερικά παραδείγματα, τα οποία περικλείουν τυχαία αθροίσματα για να αποκτήσουμε μία αίσθηση για το τι είναι οι μεγάλες αποκλίσεις και πού εμφανίζονται οι αποκλίσεις αυτές. Τα παραδείγματα θα είναι συνοπτικά, αλλά όχι απλά.

Παράδειγμα 1.2.1 (Τυχαία τμήματα – Random bits).

Θεωρούμε μία ακολουθία $b = (b_1, b_2, \dots, b_n)$ από n ανεξάρτητα τυχαία τμήματα, τα οποία παίρνουν τη τιμή 0 και 1 με ίση πιθανότητα, και

$$R_n = \frac{1}{n}\sum_{i=1}^{n} b_i \qquad (\mathbf{1.1})$$

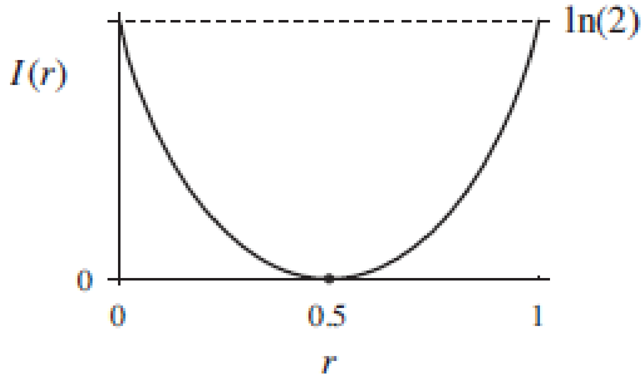
είναι το κλάσμα των τιμών ίσων με 1, οι οποίες περιέχονται στην ακολουθία. Ενδιαφερόμαστε να υπολογίσουμε τη πιθανότητα $P(R_n = r)$ όπου το R_n υποτίθεται είναι μία από τις εύλογες τιμές $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}$. Αφού τα τμήματα είναι ανεξάρτητα και αμερόληπτα, έχουμε $P(b) = 2^{-n}$ για όλα τα $b \in \{0,1\}^n$, οπότε

$$P(R_n = r) = \sum_{b:R_{n(b)=r}} P(b) = \frac{1}{2^n}\frac{n!}{(rn)!\,[(1-r)n]!} \qquad (\mathbf{1.2})$$

Χρησιμοποιώντας τη προσέγγιση Stirling, $n! \approx n^n e^{-n}$, μπορούμε να εξάγουμε από αυτό το αποτέλεσμα μία κυρίαρχη συνεισφορά έχοντας το τύπο

$$P(R_n = r) \approx e^{-nI(r)}, \qquad I(r) = ln2 + rlnr + (1-r)\ln(1-r) \qquad (\mathbf{1.3})$$

για n μεγάλο. Η συνάρτηση *I(r)*, η οποία υπάρχει στο εκθετικό είναι θετική και κυρτή για $r \in [0,1]$, όπως φαίνεται στο σχήμα 1, και έχει ένα μοναδικό ελάχιστο στο μηδέν για *r* = ½.



Σχήμα 1: Η συνάρτηση ρυθμού (rate function) *I(r)* για το παράδειγμα 1.2.1.

Η προσέγγιση που έχουμε στη σχέση (3) είναι ένα παράδειγμα προσέγγισης μεγάλης απόκλισης. Ο τύπος της εκθετικής μείωσης αυτής της προσέγγισης, συνδυαζόμενος με την έκφραση της μείωσης ή αλλιώς τη συνάρτηση ρυθμού (*rate function*) *I(r)*, δείχνει ότι οι «μη ισορροπημένες» ακολουθίες των n τμημάτων περιέχουν περισσότερες τιμές 0 παρά 1, ή αντιστρόφως, αλλά είναι δύσκολο να παρατηρηθούν καθώς το n γίνεται μεγαλύτερο επειδή η $P(R_n)$ ελαττώνεται εκθετικά με το n για το οποίο $R_n \neq 1/2$. Μόνο οι «ισορροπημένες» ακολουθίες όπου $R_n \approx 1/2$ έχουν μία μη αμελητέα πιθανότητα να παρατηρηθούν καθώς το n γίνεται μεγάλο.

Παράδειγμα 1.2.2 (Δειγματικός μέσος κανονικά κατανεμημένων τυχαίων μεταβλητών (τ. μ.) – Gaussian sample mean).

Η τυχαία μεταβλητή R_n, η οποία ορίστηκε στο προηγούμενο παράδειγμα ως το άθροισμα των n τυχαίων μεταβλητών διαιρεμένο με το n, ονομάζεται στα μαθηματικά δειγματικός μέσος. Στο συγκεκριμένο παράδειγμα, θεωρούμε έναν παρόμοιο δειγματικό μέσο, ο οποίος ορίζεται ως

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad (\mathbf{1.4})$$

και υποθέτουμε ότι οι τυχαίες μεταβλητές X_i είναι ανεξάρτητες και όμοια κατανεμημένες σύμφωνα με τη συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής

$$p(X_i = x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/(2\sigma^2)}. \quad \mathbf{(1.5)}$$

Οι παράμετροι μ και σ^2 αναπαριστούν, ως συνήθως, τη μέση τιμή και τη διακύμανση, αντίστοιχα, των X_i.

Η πυκνότητα πιθανότητας του S_n μπορεί να γραφεί ως το ολοκλήρωμα

$$p(S_n = s) = \int_{\{x\in\mathbb{R}^n : S_n(x)=s\}} p(x)dx, \quad \mathbf{(1.6)}$$

όπου $x = (x_1, x_2, \dots, x_n)$ είναι το διάνυσμα των τυχαίων μεταβλητών, και

$$p(x) = p(x_1, x_2, \dots, x_n) = p(x_1)p(x_2)\cdots p(x_n) \quad \mathbf{(1.7)}$$

η πυκνότητα γινομένου τους. Η λύση αυτού του ολοκληρώματος είναι

$$p(S_n = s) = \sqrt{\frac{n}{2\pi\sigma^2}} e^{-n(s-\mu)^2/(2\sigma^2)}, \quad \mathbf{(1.8)}$$

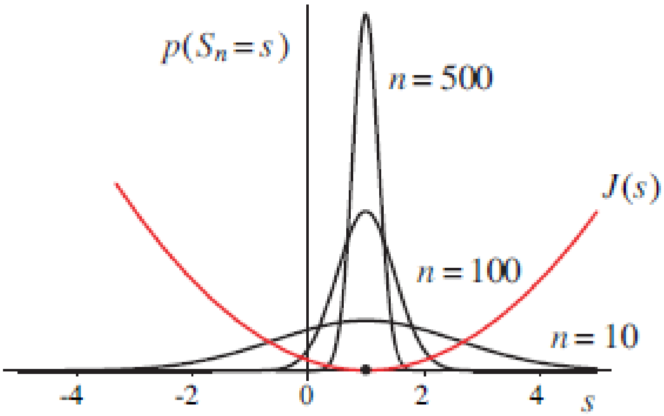
αφού ένα άθροισμα κανονικά κατανεμημένων τυχαίων μεταβλητών ακολουθεί κανονική κατανομή. Μία προσέγγιση μεγάλης απόκλισης έπεται από αυτό το αποτέλεσμα μηδενίζοντας τον όρο $\sqrt{n}$, ο οποίος είναι υποκυρίαρχος σε σχέση με την εκθετική μείωση, γι' αυτό έχουμε

$$p(S_n = s) \approx e^{-nJ(s)}, \qquad J(s) = \frac{(s-\mu)^2}{2\sigma^2}, \quad s \in \mathbb{R}. \quad \mathbf{(1.9)}$$

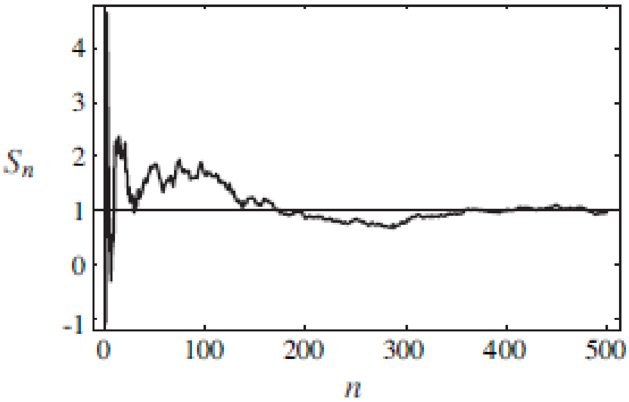
Η συνάρτηση ρυθμού *J(s)*, την οποία υπολογίσαμε στο συγκεκριμένο παράδειγμα είναι όμοια με τη συνάρτηση *I(r)*, την οποία υπολογίσαμε στο πρώτο παράδειγμα. Και οι δύο είναι κυρτές συναρτήσεις και κατέχουν ένα μοναδικό ελάχιστο (σχήμα 2). Όπως και στη περίπτωση της *I(r)*, το ελάχιστο της *J(s)* έχει επίσης την επίδραση, ότι καθώς το n μεγαλώνει, το $p(S_n = s)$ συγκεντρώνεται όλο και περισσότερο γύρω από τη μέση τιμή μ επειδή η μέση τιμή είναι το μόνο σημείο όπου *J(s)* = 0, και για το οποίο το $p(S_n = s)$ δεν μειώνεται εκθετικά. Στα μαθηματικά, αυτή η συγκεντρωτική ιδιότητα εκφράζεται από το παρακάτω όριο:

$$\lim_{n\to\infty} P(S_n \in [\mu-\delta, \mu+\delta]) = 1 \quad \mathbf{(1.10)}$$

όπου δ είναι ένας οποιοσδήποτε θετικός αριθμός. Οποτεδήποτε αυτό το όριο υπάρχει, θεωρούμε ότι ο S_n συγκλίνει στη μέση τιμή (σχήμα 3) και ο δειγματικός μέσος υπακούει στον Νόμο των Μεγάλων Αριθμών. Γενικά, τα αθροίσματα ανεξάρτητων και όμοια κατανεμημένων τυχαίων μεταβλητών, τα οποία περιέχουν διαφορετικών συναρτήσεων πυκνότητας πιθανότητας προσθετέους, έχουν διαφορετικές συναρτήσεις ρυθμού.



Σχήμα 2: Δειγματικός μέσος κανονικά κατανεμημένων τ. μ. με μ = σ =1 και πυκνότητα πιθανότητας $p(S_n = s)$ για αυξανόμενες τιμές του n αντιστοιχώντας τη συνάρτηση ρυθμού *J(s)*.



Σχήμα 3: Ο δειγματικός μέσος S_n συγκλίνει στη μέση τιμή του.

Παράδειγμα 1.2.3 (Εντροπία μη αλληλεπιδρουσών ιδιοστροφορμών)

Θεωρούμε n ιδιοστροφορμές (spins) $\sigma_1, \sigma_2, \dots, \sigma_n$ οι οποίες παίρνουν τιμές στο σύνολο $\{-1,1\}$. Είναι γνωστό ότι ο αριθμός $\Omega(m)$ των διαμορφώσεων των ιδιοστροφορμών $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ έχοντας κάποια μαγνήτιση ανά ιδιοστροφορμή [5]

$$\frac{1}{n}\sum_{i=1}^{n}\sigma_i \qquad (\mathbf{1.11})$$

ίση με *m* δίνεται από έναν δυωνυμικό τύπο

$$\Omega(m) = \frac{n!}{[(1-m)n/2]!\,[(1+m)n/2]!}. \qquad (\mathbf{1.12})$$

Η ομοιότητα αυτού του αποτελέσματος με αυτό που είδαμε στο παράδειγμα 1.2.1 είναι προφανής. Μπορούμε να χρησιμοποιήσουμε τη προσέγγιση Stirling για να έχουμε μία προσέγγιση μεγάλης απόκλισης για το *Ω(m)*, το οποίο γράφουμε ως

$$\Omega(m) \approx e^{ns(m)}, \qquad s(m) = -\frac{1-m}{2}ln\frac{1-m}{2} - \frac{1+m}{2}ln\frac{1+m}{2}, \quad m \in [-1,1]. \qquad (\mathbf{1.13})$$

Η συνάρτηση *s(m)* είναι η εντροπία, η οποία σχετίζεται με τη μέση μαγνήτιση.

Μπορούμε να μετρήσουμε τον αριθμό *Ω(l)* των διαμορφώσεων των ιδιοστροφορμών, συμπεριλαμβανομένου ενός σχετικού αριθμού l_+ των +1 ιδιοστροφορμών και ενός σχετικού αριθμού l_- των -1 ιδιοστροφορμών. Αυτοί οι δύο σχετικοί αριθμοί ή συχνότητες είναι συνιστώσες του διδιάστατου εμπειρικού διανύσματος $l = (l_+, l_-)$, για το οποίο βρίσκουμε

$$\Omega(l) \approx e^{n\tilde{s}(l)}, \qquad \tilde{s}(l) = -l_+ lnl_+ - l_- lnl_- \qquad (\mathbf{1.14})$$

για n μεγάλο. Η συνάρτηση $\tilde{s}(l)$, η οποία παίζει το ρόλο της συνάρτησης ρυθμού, ονομάζεται εντροπία, η οποία συνδέεται με το εμπειρικό διάνυσμα.

1.3 Η Θεωρία των Μεγάλων Αποκλίσεων (Large Deviation Theory)

Το πλέον βασικό στοιχείο της θεωρίας των μεγάλων αποκλίσεων είναι η εκθετική προσέγγιση. Αυτή η προσέγγιση εμφανίζεται τόσο συχνά σε προβλήματα περιλαμβάνοντας πολλές τυχαίες μεταβλητές, συγκεκριμένα αυτές που μελετώνται στη στατιστική μηχανική, γι' αυτό της δίνουμε ένα όνομα: *Η αρχή της μεγάλης απόκλισης*.

1.3.1 Η αρχή της μεγάλης απόκλισης (The large deviation principle)

Μία βασική προσέγγιση του τύπου $P_n \approx e^{-nI}$, όπου P_n είναι μία πιθανότητα, n μία παράμετρος, η οποία υποτίθεται ότι είναι μεγάλη, και I μία θετική σταθερά, η οποία αναφέρεται ως μία *αρχή μεγάλης απόκλισης*. Έστω A_n μία τυχαία μεταβλητή, έχοντας δείκτη τον ακέραιο n, και έστω $P(A_n \in B)$ η πιθανότητα όπου το A_n παίρνει μία τιμή από το σύνολο B. Λέμε ότι $P(A_n \in B)$ ικανοποιεί μία *αρχή μεγάλης απόκλισης* με ρυθμό I_B εάν το όριο

$$\lim_{n\to\infty} -\frac{1}{n}\ln P(A_n \in B) = I_B \qquad (\mathbf{1.15})$$

υπάρχει.

Η ιδέα πίσω από αυτό το όριο πρέπει να είναι ξεκάθαρη. Το τί εννοούμε όταν γράφουμε $P(A_n \in B) \approx e^{-nI_B}$ είναι ότι η κυρίαρχη συμπεριφορά του $P(A_n \in B)$ είναι η εκθετική μείωση στο n. Χρησιμοποιώντας το *ο* σύμβολο του Landau, αυτό σημαίνει ότι

$$-\ln P(A_n \in B) = nI_B + o(n), \qquad (\mathbf{1.16})$$

όπου I_B είναι μία θετική σταθερά. Για να εξάγουμε τη σταθερά αυτή, διαιρούμε και τις δύο πλευρές της παραπάνω έκφρασης με το *n* για να πάρουμε

$$-\frac{1}{n}\ln P(A_n \in B) = I_B + o(1), \qquad (\mathbf{1.17})$$

και περνάμε στο όριο $n \to \infty$, για να απαλλαγούμε από τη συνεισφορά *o(1)*. Το τελικό αποτέλεσμα αυτών των βημάτων είναι το όριο των μεγάλων αποκλίσεων. Γι' αυτό το λόγο, εάν $P(A_n \in B)$ έχει μία κυρίαρχη εκθετική συμπεριφορά στο *n*, τότε αυτό το όριο μπορεί να υπάρχει με $I_B \neq 0$. Εάν το όριο δεν υπάρχει, τότε είτε $P(A_n \in B)$ είναι τόσο σπάνιο για να έχει ένα όριο ή $P(A_n \in B)$ μειώνεται με το *n* γρηγορότερα από το e^{-na} με $\alpha > 0$. Σε αυτή τη περίπτωση, λέμε ότι $P(A_n \in B)$ μειώνεται υπερ – εκθετικά (*super – exponentially*) και θέτουμε $I = \infty$. Το όριο μεγάλων αποκλίσεων μπορεί να είναι και μηδέν για οποιοδήποτε σύνολο B εάν $P(A_n \in B)$ είναι υπό – εκθετικό (*sub – exponentially)* στο *n*, δηλαδή εάν $P(A_n \in B)$ μειώνεται με το *n* πιο αργά από το e^{-na}, $\alpha > 0$.

Το όριο, το οποίο ορίζει την αρχή των μεγάλων αποκλίσεων, όπως τα περισσότερα όρια που εμφανίζονται εδώ, πρέπει να κατανοηθούν σε ένα πρακτικό και όχι σε αυστηρά μαθηματικό επίπεδο. Ομοίως, ο ορισμός μας για την αρχή των μεγάλων αποκλίσεων δεν πρέπει να θεωρηθεί αυστηρά μαθηματικός ορισμός. Υπάρχουν πολλές μαθηματικές λεπτομέρειες που πρέπει να ληφθούν υπόψη σε σχέση με τις πιθανότητες και τα όρια. Μπορούμε όμως να αναφέρουμε δύο βασικές λεπτομέρειες:

- Το όριο που εμπεριέχεται στον ορισμό της αρχής της μεγάλης απόκλισης μπορεί και να μην υπάρχει. Σε αυτή τη περίπτωση, κάποιος μπορεί να βρει ένα άνω φράγμα και ένα κάτω φράγμα στο $P(A_n \in B)$ τα οποία είναι εκθετικά και τα δύο στο n:
$$e^{-nI_B^-} \leq P(A_n \in B) \leq e^{-nI_B^+} \qquad (\mathbf{1.18})$$
Τα δύο φράγματα δίνουν ένα ακριβές νόημα στον ισχυρισμό ότι $P(A_n \in B)$ είναι εκθετικά μειούμενο με το n, και δίνει αφορμή για δύο αρχές μεγάλης απόκλισης: η μία ορίζεται σε σχέση με ένα «κατώτερο όριο» το οποίο παραγάγει τον I_B^-, και μία ορίζεται με ένα «ανώτερο όριο» το οποίο παραγάγει τον I_B^+. Υποθέτουμε ότι πάντα ισχύει $I_B^- = I_B^+$. Γι' αυτό το λόγο, ο ορισμός μας για την αρχή της μεγάλης απόκλισης περιέχει ένα μόνο όριο.

- Οι διακριτές τυχαίες μεταβλητές συνήθως αντιμετωπίζονται ως συνεχείς στο όριο $n \to \infty$ (όριο συνεχούς – continuum limit). Σε όλα τα προηγούμενα παραπάνω παραδείγματα, οι προσεγγίσεις των μεγάλων αποκλίσεων που παρήχθησαν είναι συνεχείς προσεγγίσεις περιέχοντας συνεχείς συναρτήσεις ρυθμού (continuous rate functions).

Η αντικατάσταση των διακριτών τυχαίων μεταβλητών από συνεχείς τ. μ. δικαιολογούνται από μαθηματικής απόψεως υπό την έννοια της ασθενούς σύγκλισης (weak convergence). Έστω A_n μία διακριτή τ. μ. με κατανομή πιθανότητας $P(A_n = a)$ που ορίζεται σε ένα υποσύνολο τιμών $\alpha \in \mathbb{R}$, και έστω $\tilde{A}_n$ μία συνεχής τ. μ. με πυκνότητα πιθανότητας $p(\tilde{A}_n)$ ορισμένη στο $\mathbb{R}$. Το να πούμε ότι το A_n συγκλίνει ασθενώς στο $\tilde{A}_n$ σημαίνει, ουσιαστικά, ότι οποιοδήποτε άθροισμα το οποίο περιέχει το A_n μπορεί να προσεγγιστεί, για n μεγάλο, από ολοκληρώματα που περιέχουν το $\tilde{A}_n$, για παράδειγμα,
$$\sum_{\alpha} f(a)P(A_n = a) \approx^{n\to\infty} \int f(a)p(\tilde{A}_n = a)da, \qquad (\mathbf{1.19})$$
όπου f είναι οποιαδήποτε συνεχής και φραγμένη συνάρτηση ορισμένη στον $\mathbb{R}$. Αυτό το είδος προσέγγισης είναι συνηθισμένο στις φυσικές επιστήμες, και προτείνεται ο ακόλουθος κανόνας αντικατάστασης:
$$P(A_n = a) \to p\big(\tilde{A}_n = a\big)da \qquad (\mathbf{1.20})$$

ως ένα τυπικό τέχνασμα για να παίρνουμε το συνεχές όριο του A_n.

Επίσης, μπορούμε να γράφουμε ότι

$$P(A_n \in [a, a+da]) \approx e^{-nI(a)}da, \qquad (\mathbf{1.21})$$

εννοώντας ότι το A_n, είτε διακριτό είτε συνεχές, ικανοποιεί μία αρχή μεγάλης απόκλισης. Αυτή η επιλογή συμβολισμού είναι βολική αλλά αυθαίρετη: Θα μπορούσαμε να αποφύγουμε την έκφραση με το απειροελάχιστο στοιχείο da, αλλά εμείς θα το χρησιμοποιήσουμε τελικά, παρόλο που το στοιχείο αυτό δεν είναι εκθετικό με το n.

$$\langle f(A_n)\rangle = \int f(a)P(A_n \in [a, a+da]) \asymp \int f(a)e^{-nI(a)}da \qquad (\mathbf{1.22})$$

Υπάρχουν δύο σημεία που πρέπει να διευκρινιστούν για να προχωρήσουμε βαθύτερα στη θεωρία των μεγάλων αποκλίσεων. Πρώτα απ' όλα, θα χρησιμοποιούμε την συμπαγή έκφραση $P(A_n \in da)$ εννοώντας ότι $P(A_n \in [a, a+da])$. Ύστερα, θα χρησιμοποιήσουμε το σύμβολο " $\asymp$ " αντί του " $=$ " οποτεδήποτε αναπτύσσονται αρχές μεγάλης αποκλίσεως. Γι' αυτό γράφουμε

$$P(A_n \in da) \asymp e^{-nI(a)}da \qquad (\mathbf{1.23})$$

εννοώντας ότι το A_n ικανοποιεί μία αρχή μεγάλης απόκλισης, με συνάρτηση ρυθμού $I(a)$. Το σύμβολο " $\asymp$ " συνηθίζει να τονίζει ότι, καθώς $n \to \infty$, το κυρίαρχο μέρος του $P(A_n \in da)$ μειώνεται εκθετικά όπως το $e^{-nI(a)}$. Επίσης, το σύμβολο αυτό χρησιμοποιείται για να εκφράσει μία σχέση ισοδυναμίας και ισότητας σε λογαριθμική κλίμακα: Έτσι ερμηνεύουμε ότι $a_n \asymp b_n$ σημαίνει ότι

$$\lim_{n\to\infty} \frac{1}{n} \ln a_n = \lim_{n\to\infty} \frac{1}{n} \ln b_n . \qquad (\mathbf{1.24})$$

1.3.2 Ασθενής σύγκλιση

Επικεντρωνόμαστε στην περίπτωση όπου αντιμετωπίζουμε την ακολουθία συναρτήσεων f_n σαν μια ακολουθία τυχαίων μεταβλητών X_n.

Θεωρούμε μία ακολουθία από μέτρα μ_n σε ένα μετρήσιμο χώρο $(\mathbb{X}, \mathcal{F})$ και ένα μέτρο μ στο ίδιο χώρο.

Ορισμός 1: Θα λέμε ότι η ακολουθία μ_n συγκλίνει ασθενώς στο μέτρο μ και θα συμβολίζουμε $\mu_n \to \mu$, αν για κάθε πραγματική συνεχή και φραγμένη συνάρτηση f ισχύει $\int_{\mathbb{X}} f d\mu_n \to \int_{\mathbb{X}} f d\mu$.

<u>Θεώρημα 1</u>: Έστω μία ακολουθία μέτρων μ_n τέτοια ώστε

$$\lim_{m\to\infty} sup_n \mu_n([-m,m]^c) = 0 \qquad (\mathbf{1.25})$$

Τότε υπάρχει υπακολουθία μ_{n_k} που συγκλίνει ασθενώς σε κάποιο μέτρο μ.

Η ιδιότητα (**1.25**) ονομάζεται tightness και είναι το ανάλογο της συμπάγειας για τα μέτρα.

Αν τα μέτρα μ_n είναι μέτρα πιθανότητας τότε η σύγκλιση αυτή ονομάζεται σύγκλιση σε κατανομή. Αν X_n είναι μια ακολουθία τυχαίων μεταβλητών τότε μπορούμε να καταλάβουμε τα μ_n σαν τα επαγόμενα μέτρα πιθανότητας από την ακολουθία τυχαίων μεταβλητών X_n. Η ασθενής σύγκλιση λοιπόν μπορεί να ερμηνευθεί ως $\mathbb{E}[f(X_n)] \to \mathbb{E}[(f(X))]$ για κάθε συνεχή και φραγμένη συνάρτηση f. Θα συμβολίζουμε $X_n \xrightarrow{\mathrm{D}} X$. Η έννοια αυτή της σύγκλισης είναι πολύ χρήσιμη στη στατιστική.

Είναι πολύ σημαντικό να καταλάβουμε ότι στην σύγκλιση σε κατανομή δεν είναι απαραίτητο όλες οι τυχαίες μεταβλητές να ορίζονται στο ίδιο χώρο πιθανότητας. Αυτό συμβαίνει πολλές φορές και στις εφαρμογές, π.χ. σε επαναλαμβανόμενα πειράματα.

Αν όλες οι τυχαίες μεταβλητές ορίζονται στον ίδιο χώρο πιθανότητας μπορούμε να συγκρίνουμε την σύγκλιση σε κατανομή με τις άλλες συγκλίσεις.

<u>Πρόταση 1</u>: Αν $X_n \xrightarrow{\mu} X$ τότε $X_n \xrightarrow{\mathrm{D}} X$

Θεωρούμε $(\mathbb{X}, \mathcal{F}, \mu)$ ένα χώρο με μέτρο. Η περίπτωση όπου $\mathbb{X} = \Omega$, ένα σύνολο ενδεχομένων και $\mu = P$ ένα μέτρο πιθανότητας είναι ειδική περίπτωση.

Επίσης $f_n : \mathbb{X} \to \mathbb{R}$ θα είναι μια ακολουθία συναρτήσεων. Στην ειδική περίπτωση όπου $\mathbb{X} = \Omega$ ένα σύνολο ενδεχομένων, η ακολουθία συναρτήσεων θα συμβολίζεται με $X_n : \Omega \to \mathbb{R}$ και θα ονομάζεται ακολουθία τυχαίων μεταβλητών.

<u>Ορισμός 2</u>: Έστω $(\Omega, \mathcal{F}, \mu)$ ένας χώρος πιθανότητας. Αν $X : \Omega \to \mathbb{X}$ μια τυχαία μεταβλητή θα ονομάζουμε επαγόμενο μέτρο $\mu_X : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ το μέτρο που ορίζεται από την σχέση $\mu_X(A) = \mu(X^{-1}(A))$ για κάθε $A \in \mathcal{B}(\mathbb{R})$

<u>Ορισμός 3</u>: Θα λέμε ότι η ακολουθία τυχαίων μεταβλητών X_n συγκλίνει σε κατανομή στην τυχαία μεταβλητή X ($X_n \xrightarrow{D} X$) αν τα επαγόμενα μέτρα $\mu_{X_n} \to \mu_X$, δηλαδή για κάθε συνεχή και φραγμένη συνάρτηση f ισχύει $\mathbb{E}[f(X_n)] \to \mathbb{E}[(f(X))]$.

Σύγκλιση σ. β. $\Longrightarrow$ Σύγκλιση σε πιθανότητα $\Longrightarrow$ Σύγκλιση σε κατανομή

$\Uparrow$

Σύγκλιση L^p

Το παραπάνω διάγραμμα μας εξασφαλίζει ότι αν μια ακολουθία $X_n \xrightarrow{\sigma\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha} X$ τότε $X_n \xrightarrow{\mathrm{D}} X$.

Το αντίστροφο δεν είναι απαραίτητα αληθές. Μια ειδική περίπτωση όπου ισχύει είναι όταν $X = c$ (σταθερά) σχεδόν βέβαια.

<u>Πρόταση 2</u>: Η ακολουθία τυχαίων μεταβλητών X_n έχει υπακολουθία X_{n_k} τέτοια ώστε $X_{n_k} \xrightarrow{D} X$ για κάποια τυχαία μεταβλητή X, αν η ακολουθία επαγόμενων μέτρων μ_{X_n}είναι tight δηλαδή αν $\lim_m sup_n \mu_{X_n}([-m, m]^c) = 0$.

Για να ελέγξουμε στην πράξη αν $X_n \xrightarrow{\mathrm{D}} X$ δεν χρειάζεται να ελέγξουμε ότι $\mathbb{E}[f(X_n)] \to \mathbb{E}[(f(X))]$ για κάθε φραγμένη και συνεχή συνάρτηση f, αλλά μόνο για τις φραγμένες και συνεχείς κατά Lipschitz συναρτήσεις f.

Κάνοντας χρήση της παραπάνω παρατήρησης έχουμε το πολύ χρήσιμο σε εφαρμογές στατιστικής θεώρημα Slutsky.

<u>Θεώρημα 2</u>: Αν $X_n, Y_n : \Omega \to \mathbb{R}^n$ δύο ακολουθίες τυχαίων μεταβλητών τέτοιες ώστε $X_n \xrightarrow{\mathrm{D}} X$ και $\|X_n - Y_n\| \xrightarrow{\sigma\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha} 0$ τότε $Y_n \xrightarrow{D} X$

1.3.3 Υπολογίζοντας συναρτήσεις ρυθμού (rate functions)

Η θεωρία των μεγάλων αποκλίσεων μπορεί να περιγραφεί ως μία συλλογή από μεθόδους, οι οποίες αναπτύχθηκαν και συλλέχτηκαν όλες μαζί, ώστε να λύσουν δύο προβλήματα:

- Να αποδείξουν ότι η αρχή μεγάλης απόκλισης υπάρχει για μία δεδομένη τυχαία μεταβλητή
- Να εξάγουν την έκφραση της αντίστοιχης συνάρτησης ρυθμού.

Και τα δύο αυτά προβλήματα μπορούν να επιλυθούν, όπως κάναμε και στα προηγούμενα παραδείγματα, υπολογίζοντας απευθείας την κατανομή πιθανότητας μίας τυχαίας μεταβλητής, και εξάγοντας από αυτή τη κατανομή μία προσέγγιση μεγάλης απόκλισης χρησιμοποιώντας τη προσέγγιση Stirling ή μία οποιαδήποτε άλλη ασυμπτωτική μέθοδο. Γενικά, είναι δύσκολο ή και μερικές φορές αδύνατο να εξάγουμε αρχές μεγάλης απόκλισης μέσω αυτού του άμεσου υπολογισμού. Συνδυαστικές μέθοδοι βασισμένες στη προσέγγιση Stirling δεν μπορούν να χρησιμοποιηθούν, για παράδειγμα, για συνεχείς τυχαίες μεταβλητές, και γίνονται αρκετά σύνθετες όταν σε σχέση με αθροίσματα διακριτών τυχαίων μεταβλητών, οι οποίες δεν είναι όμοια κατανεμημένες και ανεξάρτητες. Σε αυτές τις περιπτώσεις, μία πιο γενική μέθοδος υπολογισμού εμφανίστηκε από ένα θεμελιώδες αποτέλεσμα της θεωρίας μεγάλων αποκλίσεων γνωστό ως θεώρημα Gärtner – Ellis (Gärtner – Ellis Theorem).

Θεώρημα Gärtner – Ellis (Gärtner – Ellis Theorem)

Θεωρούμε μία πραγματική τυχαία μεταβλητή A_n παραμετροποιημένη από έναν θετικό ακέραιο αριθμό n, και ορίζεται η *ανηγμένη αθροιστική γεννήτρια συνάρτηση* της A_n από το όριο

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}[e^{nkA_n}], \qquad (\mathbf{1.26})$$

όπου $k \in \mathbb{R}$ και

$$\mathbb{E}[e^{nkA_n}] = \int_{\mathbb{R}} e^{nka} P(A_n \in da). \qquad (\mathbf{1.27})$$

Το θεώρημα Gärtner – Ellis δηλώνει ότι, εάν το $\lambda(k)$ υπάρχει και είναι διαφορίσιμο για όλα τα $k \in \mathbb{R}$, τότε η A_n ικανοποιεί μία αρχή μεγάλης απόκλισης, για παράδειγμα,

$$P(A_n \in da) \asymp e^{-nI(a)} da, \qquad (\mathbf{1.28})$$

με μία συνάρτηση ρυθμού $I(\alpha)$, η οποία δίνεται από

$$I(\alpha) = sup_{k\in\mathbb{R}}\{ka - \lambda(k)\}. \qquad (\mathbf{1.29})$$

Ο μετασχηματισμός που ορίζεται με το supremum είναι μία επέκταση του μετασχηματισμού Legendre, αναφερόμενος ως *μετασχηματισμός Legendre – Fenchel (Legendre – Fenchel transform).* Δηλαδή το θεώρημα Gärtner – Ellis δηλώνει ότι όταν η ανηγμένη αθροιστική γεννήτρια συνάρτηση $\lambda(k)$ της A_n είναι διαφορίσιμη, τότε η A_n υπακούει σε μία αρχή μεγάλης απόκλισης με μία συνάρτηση ρυθμού $I(\alpha)$, η οποία δίνεται από το μετασχηματισμό Legendre – Fenchel της $\lambda(k)$.

Εισάγοντας τη προσέγγιση της εξίσωσης (1.28) στην εξίσωση (1.27), η οποία ορίζει την αναμενόμενη τιμή, έχουμε

$$\mathbb{E}[e^{nkA_n}] \asymp \int_{\mathbb{R}} e^{n[ka - I(a)]} da. \qquad (\mathbf{1.30})$$

Στη συνέχεια, προσεγγίζουμε το ολοκλήρωμα από τη μέγιστη ολοκληρωτέα ποσότητα, η οποία βρίσκεται τοποθετώντας το μέγιστο του $ka - I(a)$. Αυτή η προσέγγιση, η οποία είναι γνωστή ως προσέγγιση σαγματικού σημείου ή προσέγγιση Laplace, είναι μία φυσική προσέγγιση που θεωρούμε εδώ επειδή το σφάλμα που συνδέεται με αυτή είναι της ίδιας τάξεως με το σφάλμα που σχετίζεται με τη προσέγγιση μεγάλης απόκλισης. Γι' αυτό, θεωρώντας ότι το μέγιστο του $ka - I(a)$ υπάρχει και είναι μοναδικό, γράφουμε

$$\mathbb{E}[e^{nkA_n}] \asymp \exp(n sup_{a\in\mathbb{R}}\{ka - I(a)\}) \qquad \mathbf{(1.31)}$$

άρα

$$\lambda(k) = \lim_{n\to\infty}\frac{1}{n}\ln \mathbb{E}[e^{nkA_n}] = sup_{a\in\mathbb{R}}\{ka - I(a)\}. \qquad \mathbf{(1.32)}$$

Για να βρούμε το $I(a)$ σε σχέση με το $\lambda(k)$, χρησιμοποιούμε τη περίπτωση όπου οι μετασχηματισμοί Legendre – Fenchel μπορούν να αντιστραφούν όταν το $\lambda(k)$ είναι παντού διαφορίσιμη. Σε αυτή τη περίπτωση, ο μετασχηματισμός Legendre – Fenchel είναι αυτό-αντίστροφος (*self – inverse*), άρα

$$I(a) = sup_{k\in\mathbb{R}}\{ka - \lambda(k)\}, \qquad \mathbf{(1.33)}$$

το οποίο είναι το αποτέλεσμα της εξίσωσης (1.29).

Έχουμε δύο σημαντικά σημεία για τη θεωρία των μεγάλων αποκλίσεων. Το πρώτο είναι ότι οι μετασχηματισμοί Legendre – Fenchel εμφανίζονται μέσα στη θεωρία ως μία φυσική συνέπεια της προσέγγισης Laplace. Το δεύτερο σημείο είναι ότι το θεώρημα Gärtner – Ellis είναι μία συνέπεια της αρχής μεγάλης απόκλισης συνδυαζόμενη με τη προσέγγιση Laplace.

1.3.4 Θεώρημα Cramér (Cramér's Theorem)

Η εφαρμογή του θεωρήματος Gärtner – Ellis σ' έναν δειγματικό μέσο

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \mathbf{(1.34)}$$

των ανεξάρτητων και όμοια κατανεμημένων τυχαίων μεταβλητών μας δίνει ένα κλασικό αποτέλεσμα της θεωρίας πιθανοτήτων, γνωστό και ως θεώρημα Cramér (*Cramér's theorem*). Σε αυτή τη περίπτωση, η ανηγμένη αθροιστική γεννήτρια συνάρτηση έχει την απλή μορφή

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\left[e^{k \sum_{i=1}^{n} X_i}\right] = \lim_{n\to\infty} \frac{1}{n} \ln \prod_{i=1}^{n} \mathbb{E}[e^{kX_i}] = \ln \mathbb{E}[e^{kX}], \qquad (\mathbf{1.35})$$

όπου X είναι οποιοσδήποτε από τους προσθετέους X_i. Ως αποτέλεσμα, έχουμε μία αρχή μεγάλης απόκλισης για τον S_n απλά υπολογίζοντας την *αθροιστική γεννήτρια συνάρτηση* $\ln \mathbb{E}[e^{kX}]$ ενός απλού προσθετέου, και παίρνοντας τον μετασχηματισμό Legendre – Fenchel του αποτελέσματος. Επίσης, η συνθήκη διαφορισιμότητας του θεωρήματος Gärtner – Ellis δεν χρειάζεται να εξεταστεί για όμοια κατανεμημένους δειγματικούς μέσους, επειδή η *γεννήτρια συνάρτηση* ή ο μετασχηματισμός Laplace $\mathbb{E}[e^{kX}]$ μίας τυχαίας μεταβλητής X είναι πάντα πραγματική και αναλυτική όταν υπάρχει για όλα τα $k \in \mathbb{R}$.

Θεώρημα Cramér (Cramér's Theorem)

Για κάθε $x \geq \mathbb{E}X_1$, τότε έχουμε

$$\lim_{n\to\infty} \frac{1}{n} \ln P\left[\frac{S_n}{n} \geq x\right] = -I(x) = -\, inf_{y\geq x} I(y). \qquad (\mathbf{1.36})$$

<u>Απόδειξη</u>: 1) Άνω φράγμα. Το κύριο βήμα για το άνω φράγμα της εξίσωσης (1.36) βασίζεται στην ανισότητα Chebychev συνδυαζόμενο με την ανεξαρτησία και την όμοια κατανομή στο X_i:

$$P\left[\frac{S_n}{n} \geq x\right] = \mathbb{E}\left[1_{\frac{S_n}{n}\geq x}\right] \leq \mathbb{E}\left[e^{k(S_n - nx)}\right] = \exp(n\lambda(k) - knx), \quad \forall k \geq 0.$$

Παίρνοντας το infimum στο $k \geq 0$, και αφού $I(x) = sup_{a\geq 0}[kx - \lambda(k)]$ για $x \geq \mathbb{E}X_1$, έχουμε

$$P\left[\frac{S_n}{n} \geq x\right] \leq \exp\big(-nI(x)\big).$$

Αυτό είναι το άνω φράγμα της εξίσωσης (35).

2) Κάτω φράγμα. Αφού $P\left[\frac{S_n}{n} \geq x\right] \geq P\left[\frac{S_n}{n} \in [x, x+\varepsilon)\right]$, για όλα τα $\varepsilon > 0$, αρκεί να δείξουμε ότι

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} inf \frac{1}{n} \ln P[\frac{S_n}{n} \in [x, x+\varepsilon)] \geq -I(x). \qquad (\mathbf{1.37})$$

Υποθέτουμε ότι το μ είναι φραγμένο έτσι ώστε το λ να είναι πεπερασμένο παντού. Υποθέτουμε επίσης ότι υπάρχει μία λύση $k = k(x) > 0$ στην εξίσωση σαγματικού σημείου $\lambda'(k) = x$, για παράδειγμα πετυχαίνοντας το supremum στο $I(x) = k(x)x - \lambda(k(x))$. Το σημαντικό σημείο είναι να εισάγουμε τη νέα κατανομή πιθανότητας μ_k και P_k το αντίστοιχο μέτρο πιθανότητας (με αντίστοιχη κατανομή πιθανότητας μ και μέτρο πιθανότητας P) στον $(\Omega, \mathcal{F})$ με λόγο πιθανοφάνειας (ή η παράγωγος Radon – Nikodym):

$$\frac{dP_k}{dP} = \prod_{i=1}^{n} \frac{d\mu_k}{d\mu}(X_i) = \exp\big(kS_n - n\lambda(k)\big).$$

Τότε, έχουμε για όλα τα $\varepsilon > 0$:

$$P\left[\frac{S_n}{n} \in [x, x+\varepsilon)\right] = \mathbb{E}_k\left[\exp\big(-kS_n + k\lambda(k)\big)\, 1_{\frac{S_n}{n}\in[x,x+\varepsilon)}\right]$$

$$= \; e^{-n(kx-\lambda(k))}\mathbb{E}_k\left[\exp\left(-nk\left(\frac{S_n}{n} - x\right)\right) 1_{\frac{S_n}{n}\in[x,x+\varepsilon)}\right]$$

$$\geq \; e^{-n(kx-\lambda(k))}e^{-n|k|\varepsilon}P_k\left[\frac{S_n}{n} \in [x, x+\varepsilon)\right],$$

και άρα

$$\frac{1}{n}\ln P[\tfrac{S_n}{n} \in [x, x+\varepsilon)] \;\geq\; -[kx - \lambda(k)] - |k|\varepsilon + \frac{1}{n}\ln P_k[\tfrac{S_n}{n} \in [x, x+\varepsilon)]. \qquad (\mathbf{1.38})$$

Τώρα, αφού $\lambda'(k) = x$, έχουμε $\mathbb{E}_k[X_1] = x$, και από το νόμο των Μεγάλων Αριθμών και το Κεντρικό Οριακό Θεώρημα: $lim_n P_k\left[\frac{S_n}{n} \in [x, x+\varepsilon)\right] = \frac{1}{2}$ (> 0). Επίσης έχουμε $I(x) = kx - \lambda(k)$. Γι’ αυτό το λόγο, πηγαίνοντας το $n \to \infty$, τότε και $\varepsilon \to 0$ στην εξίσωση (1.38), παίρνουμε την (1.37).

Τώρα, εάν το supremum στο $I(x)$ δεν επιτευχθεί, μπορούμε να βρούμε μία ακολουθία $(k_i)_{i\to\infty}$, όπως $k_i x - \lambda(k_i) \to I(x)$. Αφού $\mathbb{E}[e^{k_i(X_1-x)}1_{X_1<x}] \to 0$, τότε παίρνουμε

$$\mathbb{E}\big[e^{k_i(X_1-x)}1_{X_1\geq x}\big] \to e^{-I(x)},$$

καθώς το k πηγαίνει στο άπειρο. Αυτό είναι πιθανό μόνο εάν $P[X_1 > x] = 0$ και $P[X_1 = x] = e^{-I(x)}$. Από την υπόθεση της ανεξαρτησίας και της ταυτόσημης κατανομής της τ. μ. X_i, αυτό συνεπάγεται $P\left[{}^{S_n}\!/_n \geq x\right] \geq (P[X_1 \geq x])^n = e^{-nI(x)}$, το οποίο αποδεικνύει την εξίσωση (1.37).

Υποθέτουμε τώρα ότι το $\boldsymbol{\mu}$ είναι μη φραγμένο, και φτιάχνουμε $\boldsymbol{M}$ μεγάλο αρκετά έτσι ώστε $\mu([-M, M]) > 1 - \varepsilon$. Από τη προηγούμενη απόδειξη, Το κάτω φράγμα (1.37) ισχύει με το νόμο του

${S_n}/{n}$ όρου στο $\{|X_i| \leq M, i = 1, \dots, n\}$, και με μία αθροιστική γεννήτρια συνάρτηση ίση με την αθροιστική γεννήτρια συνάρτηση του γενόμενου υπό όρους νόμου του X_1 δεδομένου ότι με $|X_i| \leq M$:

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} inf \frac{1}{n} \ln P[\frac{S_n}{n} \in [x, x+\varepsilon) \Big| |X_i| \leq M, i = 1, \dots, n] \geq -\tilde{I}_M(x) := -sup_{k\in\mathbb{R}}[kx - \tilde{\lambda}_M(k)], \quad (\mathbf{1.39})$$

με $\lambda_M(k) = \ln \mathbb{E}[e^{kX_1} | |X_1| \leq M] = \lambda_M(k) - \ln \mu([-M, M]),\ \ \lambda_M(k) = \ln \mathbb{E}\big[e^{kX_1} 1_{|X_1|\leq M}\big]$.

Τώρα, γράφοντας το τύπο του Bayes ότι $P\left[\frac{S_n}{n} \in [x, x+\varepsilon)\right] = P\left[\frac{S_n}{n} \in [x, x+\varepsilon) \Big| |X_i| \leq M, i = 1, \dots, n\right] \cdot (\mu([-M, M]))^n$, παίρνουμε με τη (1.39)

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} inf \frac{1}{n} \ln P\left[\frac{S_n}{n} \in [x, x+\varepsilon)\right]$$

$$\geq \lim_{\varepsilon\to 0} \lim_{n\to\infty} inf \frac{1}{n} \ln P[\frac{S_n}{n} \in [x, x+\varepsilon) \Big| |X_i| \leq M, i = 1, \dots, n] + \ln \mu([-M, M])]$$

$$\geq -I_M(x) := -sup_{k\in\mathbb{R}}[kx - \lambda_M(k)].$$

Το απαιτούμενο αποτέλεσμα αποκτιέται στέλνοντας το M στο άπειρο.

Παράδειγμα 1.3.4.1 (Δειγματικός μέσος κανονικά κατανεμημένων τ. μ. - Gaussian sample mean)

Θεωρούμε ξανά τον δειγματικό μέσο S_n των n κανονικά κατανεμημένων (IID) τυχαίων μεταβλητών όπως στο παράδειγμα 1.2.2. Για τη πυκνότητα πιθανότητας της εξίσωσης (1.5), $\lambda(k)$ εύκολα υπολογίζεται ότι είναι

$$\lambda(k) = \ln \mathbb{E}[e^{kX}] = \mu k + \frac{1}{2}\sigma^2 k^2,\ k \in \mathbb{R}. \quad (\mathbf{1.40})$$

Όπως είναι αναμενόμενο, το $\lambda(k)$ είναι παντού διαφορίσιμο, άρα $P(S_n \in ds) \asymp e^{-nI(s)}ds$ με

$$I(s) = sup_k\{ks - \lambda(k)\}. \quad (\mathbf{1.41})$$

Αυτό το αποτέλεσμα επικαλείται το θεώρημα Cramér. Το supremum, το οποίο ορίζεται από τον μετασχηματισμό Legendre – Fenchel, και δίνεται αμέσως με συνηθισμένο υπολογισμό. Το αποτέλεσμα είναι

$$I(s) = k(s)s - \lambda\big(k(s)\big) = \frac{(s-\mu)^2}{2\sigma^2},\ s \in \mathbb{R}, \qquad \mathbf{(1.42)}$$

όπου $k(s)$ είναι το μοναδικό μέγιστο σημείο του $ks - \lambda(k)$ ικανοποιώντας το $\lambda'(k) = s$. Αυτό επιβεβαιώνει ακριβώς το αποτέλεσμα του παραδείγματος 1.2.2 γνωρίζοντας ότι $P(S_n \in ds) = p(S_n = s)ds$.

1.3.5 Ιδιότητες των λ και I

Τώρα παράθετουμε και αποδεικνύουμε έναν αριθμό ιδιοτήτων των ανηγμένων αθροιστικών γεννητριών συναρτήσεων (**scaled cumulant generating functions - s.c.g.f**) και των συναρτήσεων ρυθμού (**rate functions**) στη περίπτωση που οι τελευταίες αποκτούνται μέσω του θεωρήματος Gärtner – Ellis. Οι ιδιότητες που θα δούμε ισχύουν για οποιαδήποτε αυθαίρετη τυχαία μεταβλητή A_n.

1. *Ιδιότητες για λ στο k=0*

Αφού τα μέτρα πιθανότητας κανονικοποιούνται, $\lambda(0) = 0$. Επιπλέον,

$$\lambda'(0) = \lim_{n\to\infty} \frac{\mathbb{E}[A_n e^{nkA_n}]}{\mathbb{E}[e^{nkA_n}]}\bigg|_{k=0} = \lim_{n\to\infty} \mathbb{E}[A_n], \qquad \mathbf{(1.43)}$$

υπό τον όρο ότι το $\lambda'(0)$ υπάρχει. Για (IID) δειγματικούς μέσους, αυτό ανάγεται σε $\lambda'(0) = \mathbb{E}[X] = \mu$ (Σχήμα 4). Ομοίως,

$$\lambda''(0) = \lim_{n\to\infty} n(\mathbb{E}[A_n^2] - \mathbb{E}[A_n]^2) = \lim_{n\to\infty} nvar(A_n), \qquad \mathbf{(1.44)}$$

το οποίο ανάγεται στο $\lambda''(0) = var(X) = \sigma^2$ για (IID) δειγματικούς μέσους.

2. *Κυρτότητα (Convexity) του λ*

Η συνάρτηση $\lambda(k)$ είναι πάντα κυρτή (convex function). Αυτό προκύπτει ως γενική συνέπεια της ανισότητας Hölder (Hölder’s inequality):

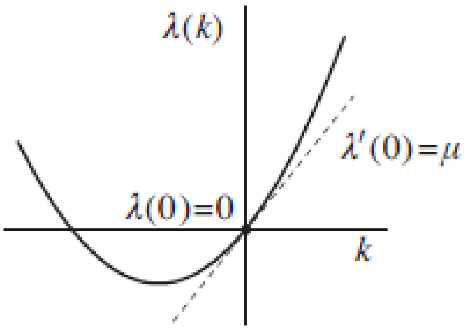
$$\textstyle\sum_i |y_i z_i| \le \left(\sum_i |y_i|^{\frac{1}{p}}\right)^p \left(\sum_i |z_i|^{\frac{1}{q}}\right)^q, \qquad \mathbf{(1.45)}$$

όπου $0 \le p, q \le 1, p + q = 1$. Εφαρμόζοντας τη συγκεκριμένη ανισότητα στο $\lambda(k)$ έχουμε

$$\alpha \ln \mathbb{E}[e^{nk_1A_n}] + (1-\alpha)\ln \mathbb{E}[e^{nk_2A_n}] \geq \ln \mathbb{E}\left[e^{n[\alpha k_1 + (1-\alpha)k_2]A_n}\right] \qquad \mathbf{(1.46)}$$

για $\alpha \in [0,1]$. Γι' αυτό το λόγο,

$$\alpha\lambda(k_1) + (1-\alpha)\lambda(k_2) \geq \lambda(\alpha k_1 + (1-\alpha)k_2) \qquad \mathbf{(1.47)}$$



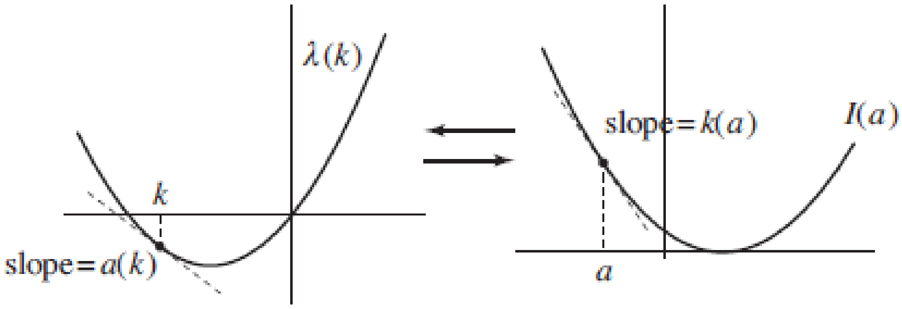
Σχήμα 4: Ιδιότητες του λ(k) στο k=0

3. *Μετασχηματισμός Legendre*

Έχουμε δει όταν υπολογίζουμε τις συναρτήσεις ρυθμού των δειγματικών μέσων κανονικά και εκθετικά κατανεμημένων τυχαίων μεταβλητών ότι ο μετασχηματισμός Legendre – Fenchel εμπλεκόμενος με το θεώρημα Gärtner – Ellis γίνεται

$$I(a) = k(a)a - \lambda\big(k(a)\big), \qquad \mathbf{(1.48)}$$

όπου $k(a)$ είναι η μοναδική ρίζα της εξίσωσης $\lambda'(k) = a$. Η εξίσωση αυτή ορίζει το μετασχηματισμό Legendre (*Legendre transform*) της $\lambda(k)$. Μία σημαντική ιδιότητα του συγκεκριμένου μετασχηματισμού ισχύει όταν η $\lambda(k)$ είναι διαφορίσιμη και αυστηρά κυρτή (*strictly convex*), η οποία είναι κυρτή χωρίς γραμμικά μέρη. Σε αυτή τη περίπτωση, το $\lambda'(k)$ αυξάνεται μονότονα, έτσι ώστε η συνάρτηση $k(a)$, ικανοποιώντας τη σχέση $\lambda'(k(\alpha)) = a$ μπορεί να αντιστρέφεται, έτσι ώστε να έχουμε μία συνάρτηση $\alpha(k)$ ικανοποιώντας τη σχέση $\lambda'(k) = a(k)$. Από την εξίσωση, η οποία ορίζει το μετασχηματισμό Legendre, έχουμε $I'\big(a(k)\big) = k$ και $I'(a) = k(a)$. Σε αυτή τη περίπτωση μόνο, οι κλίσεις της λ είναι μία προς μία σχετιζόμενες με τις κλίσεις της *I*. Αυτή η ιδιότητα ονομάζεται δυαδικότητα (duality) του μετασχηματισμού Legendre (σχήμα 5).



Σχήμα 5: Η κλίση της λ στο k είναι το σημείο όπου η κλίση της I είναι k.

4. *Θεώρημα Varadhan* (*Varadhan's Theorem*)

Στο θεώρημα Gärtner – Ellis, δείξαμε ότι εάν η A_n ικανοποιεί μία αρχή μεγάλης απόκλισης με συνάρτηση ρυθμού $I(a)$, τότε το $\lambda(k)$ είναι ο μετασχηματισμός Legendre – Fenchel της $I(a)$:

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[e^{nkA_n}] = sup_a \{ka - I(a)\}. \qquad (\mathbf{1.49})$$

Αντικαθιστώντας το γινόμενο kA_n από μία αυθαίρετη συνεχή συνάρτηση f της A_n έχουμε το πιο γενικό αποτέλεσμα

$$\lambda(f) = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\left[e^{nf(A_n)}\right] = sup_a\{f(a) - I(a)\}, \qquad (\mathbf{1.50})$$

το οποίο είναι γνωστό ως θεώρημα Varadhan. Η συνάρτηση $\lambda(f)$ είναι ένα συναρτησιακό της f.

Είναι σημαντικό να παρατηρήσουμε ότι το συγκεκριμένο θεώρημα είναι μία συνέπεια της προσέγγισης Laplace μόνο όταν η A_n είναι μία πραγματική τυχαία μεταβλητή. Για άλλους τύπους τυχαίων μεταβλητών, όπως οι τυχαίες συναρτήσεις, το συγκεκριμένο θεώρημα εφαρμόζεται και εκτείνει τη προσέγγιση Laplace σε αυτές τις τυχαίες μεταβλητές. Επίσης, το θεώρημα ισχύει όταν το $f(a) - I(a)$ έχει περισσότερα από ένα μέγιστα, όταν το ολοκλήρωμα που ορίζει την αναμενόμενη τιμή $\mathbb{E}\left[e^{nf(A_n)}\right]$ έχει περισσότερα του ενός σαγματικά σημεία.

5. *Θετικότητα των συναρτήσεων ρυθμού* (*Positivity of rate functions*)

Οι συναρτήσεις ρυθμού είναι πάντα θετικές. Αυτό ακολουθεί σημειώνοντας ότι $\lambda(0) = 0$ και το $\lambda(k)$ μπορεί να εκφραστεί πάντα ως ο μετασχηματισμός Legendre – Fenchel της $I(a)$. Γι' αυτό το λόγο,

$$\lambda(0) = sup_a\{-I(a)\} = -inf_a I(a) = 0, \qquad (\mathbf{1.51})$$

Μία αρνητική συνάρτηση ρυθμού συνεπάγεται ότι η $P(A_n \in da)$ αποκλίνει καθώς $n \to \infty$.

6. *Κυρτότητα των συναρτήσεων ρυθμού* (*Convexity of rate functions*)

Οι συναρτήσεις ρυθμού, οι οποίες λαμβάνονται από το θεώρημα Gärtner – Ellis, είναι αναγκαία κυρτές, δηλαδή είναι κυρτές και δεν έχουνε γραμμικά μέρη. Θεωρούμε ότι η $\lambda(k)$ είναι διαφορίσιμη και δεν έχει γραμμικά μέρη. Σε αυτή τη περίπτωση, ο μετασχηματισμός Legendre – Fenchel μετατρέπεται σ' έναν μετασχηματισμό Legendre, και η εξίσωση, η οποία ορίζεται από το μετασχηματισμό Legendre συνεπάγεται

$$I''(\alpha) = k'(a) = \frac{1}{\lambda''(k)}. \qquad (\mathbf{1.52})$$

Αφού η $\lambda(k)$ είναι κυρτή και χωρίς γραμμικά μέρη (($\lambda''(k) > 0$), η $I(\alpha)$ πρέπει να είναι κυρτή χωρίς γραμμικά μέρη ($I''(\alpha) > 0$). Αυτό δείχνει ότι η καμπυλότητα της $I(\alpha)$ είναι αντίστροφη στη καμπυλότητα της $\lambda(k)$. Στη περίπτωση των IID δειγματικών μέσων, συγκεκριμένα,

$$I''(\alpha = \mu) = \frac{1}{\lambda''(0)} = \frac{1}{\sigma^2}. \qquad (\mathbf{1.53})$$

Ένα παρόμοιο αποτέλεσμα ισχύει για μη IID τυχαίες μεταβλητές αντικαθιστώντας το σ^2 με το γενικό αποτέλεσμα της εξίσωσης (1.44).

7. *Νόμος των Μεγάλων Αριθμών* (*Law of Large Numbers*)

Εάν η $I(\alpha)$ παρουσιάζει ολικό ελάχιστο, τότε

$$\alpha^* = \lambda'(0) = \lim_{n\to\infty} \mathbb{E}[A_n] \qquad (\mathbf{1.54})$$

από την εξίσωση (1.43). Εάν η $I(\alpha)$ είναι διαφορίσιμη στο α^*, έχουμε $I'(\alpha^*) = k(a^*) = 0$. Για να αποδείξουμε αυτή την ιδιότητα, απλά εφαρμόζουμε την ιδιότητα της δυικότητας (δυϊσμού) για τον μετασχηματισμό Legendre - Fenchel:

$$I(\alpha^*) = k(a^*)a^* - \lambda\big(k(a^*)\big) = 0 \cdot a^* - 0 = 0. \qquad (\mathbf{1.55})$$

Το ολικό ελάχιστο της $I(\alpha)$ έχει μία ειδική ιδιότητα την οποία παρατηρήσαμε ήδη: Αντιστοιχεί, εάν είναι μοναδική, στη μοναδική τιμή για την οποία η $P(A_n \in da)$ δεν έχει εκθετική μείωση, και η οποία συγκεντρώνεται σε ένα σημείο συνεχεία καθώς $n \to \infty$ (σχήμα 6). Εξ' αιτίας αυτού του φαινομένου, έχουμε

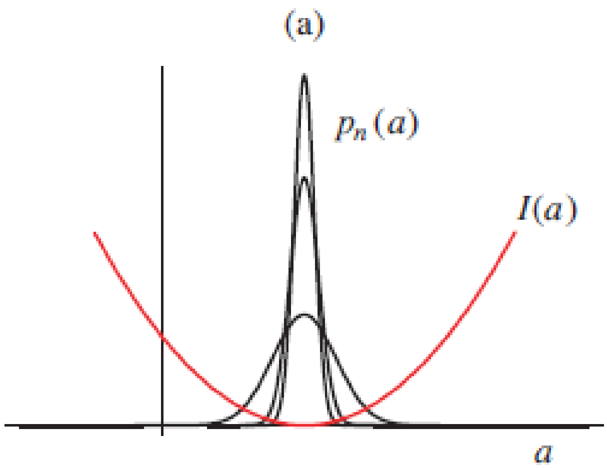
$$\lim_{n\to\infty} P(A_n \in da^*) = \lim_{n\to\infty} P(A_n \in [a^*, a^* + da]) = 1, \qquad (\mathbf{1.56})$$

όπως έχουμε ήδη αναφέρει στην εξίσωση (1.10), και καλούμε το α^* τη *πιο πιθανή ή τυπική* τιμή της A_n. Η ύπαρξη αυτής της τυπικής τιμής είναι μία έκφραση του Νόμου των Μεγάλων Αριθμών, ο οποίος δηλώνει στην ασθενή μορφή του ότι $A_n \to a^*$ με πιθανότητα 1. Μία σημαντική παρατήρηση εδώ είναι ότι η θεωρία της μεγάλης απόκλισης επεκτείνει το Νόμο των Μεγάλων Αριθμών εξασφαλίζοντας πληροφορίες για το πόσο γρήγορα συγκλίνει η A_n στη μέση τιμή της. Για να είμαστε περισσότερο ακριβείς, έστω B ένα οποιοδήποτε σύνολο από τιμές της A_n. Τότε

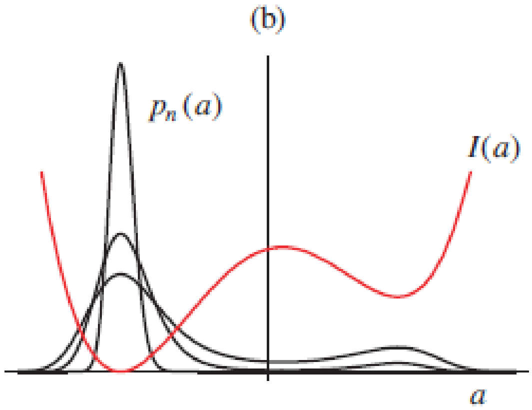
$$P(A_n \in B) = \int_B P(A_n \in da) \asymp \int_B e^{-nI(a)} da \asymp e^{-n \cdot inf_{a \in B} I(a)} \qquad (\mathbf{1.57})$$

εφαρμόζοντας προσέγγιση Laplace. Γι' αυτό το λόγο, $P(A_n \in B) \to 0$ εκθετικά γρήγορα με το n εάν $\alpha^* \notin B$, το οποίο σημαίνει ότι $P(A_n \in B) \to 1$ εκθετικά γρήγορα με το n εάν $\alpha^* \in B$.

Γενικά, η ύπαρξη ενός Νόμου των Μεγάλων Αριθμών για μία τυχαία μεταβλητή A_n είναι ένα καλό σημάδι ότι μία αρχή μεγάλης απόκλισης ισχύει για τη συγκεκριμένη τ. μ. Μπορεί να δοθεί έμφαση στο γεγονός ότι η $I(\alpha)$ μπορεί να έχει περισσότερα του ενός ολικά ελάχιστα, στη περίπτωση που ο Νόμος των Μεγάλων Αριθμών μπορεί να μην ισχύει. Οι συναρτήσεις ρυθμού μπορεί να έχουν τοπικά ελάχιστα επιπροσθέτως με το ένα ολικό. Τα ολικά ελάχιστα δίνουν τυπικές τιμές της τ. μ. στη περίπτωση ενός μόνου ελαχίστου, αν και τα τοπικά ελάχιστα δίνουν 'σχεδόν ασταθείς' τιμές της τυχαίας μεταβλητές για τις οποίες $P(A_n \in da)$ είναι τοπικά αλλά όχι ολικά μέγιστη (σχήμα 7).



Σχήμα 6: Παράδειγμα πυκνότητας πιθανότητας $p_n(a)$ και αντίστοιχης συνάρτησης ρυθμού $I(\alpha)$



Σχήμα7: Παράδειγμα πυκνότητας πιθανότητας $p_n(\alpha)$ χαρακτηριζόμενη από μία μη κυρτή συνάρτηση ρυθμού $I(\alpha)$ έχοντας ένα τοπικό ελάχιστο

8. *Κεντρικό Οριακό Θεώρημα* (*Central Limit Theorem*)

Το Κ.Ο.Θ. εμφανίζεται στη θεωρία των μεγάλων αποκλίσεων όταν μία κυρτή συνάρτηση ρυθμού $I(\alpha)$ κατέχει ένα μόνο ολικό ελάχιστο και μηδέν στο α^* και δύο φορές διαφορίσιμη στο συγκεκριμένο σημείο. Προσεγγίζοντας τη $I(\alpha)$ με το πρώτο τετραγωνικό όρο,

$$I(\alpha) \approx \frac{1}{2} I''(a^*)(a - a^*)^2, \qquad (\mathbf{1.58})$$

τότε φυσικά οδηγεί στη προσέγγιση Gauss (Gaussian approximation)

$$P(A_n \in da) \approx e^{-\frac{nI''(a^*)(a-a^*)^2}{2}} da, \qquad (\mathbf{1.59})$$

το οποίο μπορεί να θεωρηθεί ως μία ασθενής μορφή του Κ.Ο.Θ. Η προσέγγιση Gauss μπορεί να αποδειχθεί ότι είναι ακριβής για τιμές της A_n γύρω από το α^* της τάξεως $O(n^{-\frac{1}{2}})$ ή, ισοδύναμα, για τιμές του nA_n γύρω από το α^* της τάξεως $O(n^{\frac{1}{2}})$. Αυτό εξηγεί το νόημα του ονόματος ‘μεγάλες αποκλίσεις’.

1.3.6 Αρχή της συστολής (Contraction principle)[1]

Έχουμε δει μέχρι τώρα δύο βασικά αποτελέσματα της θεωρίας μεγάλης απόκλισης. Το πρώτο είναι το θεώρημα Gärtner – Ellis, το οποίο μπορεί να χρησιμοποιηθεί για να αποδείξουμε ότι η αρχή της μεγάλης απόκλισης υπάρχει και να υπολογίσουμε την αντίστοιχη συνάρτηση ρυθμού από τη ύπαρξη της $\lambda(k)$. Το δεύτερο αποτέλεσμα είναι το θεώρημα Varadhan, το οποίο μπορεί να χρησιμοποιηθεί

[1] Η συγκεκριμένη αρχή της συστολής στη θεωρία των μεγάλων αποκλίσεων δεν πρέπει να συγχέεται με την αρχή της συστολής ή αλλιώς το θεώρημα σταθερού σημείου του Banach.

για να υπολογίσουμε τη $\lambda(k)$ από την ύπαρξη μίας συνάρτησης ρυθμού. Το τελευταίο αποτέλεσμα το οποίο εισάγουμε είναι ένα χρήσιμο εργαλείο υπολογισμού, το οποίο ονομάζεται αρχή της συστολής, η οποία μπορεί να χρησιμοποιηθεί για τον υπολογισμό μίας συνάρτησης ρυθμού από την ύπαρξη μίας άλλης συνάρτησης ρυθμού.

Το πρόβλημα που υπάρχει με τη συγκεκριμένη αρχή είναι το εξής: Έχουμε μία τ. μ. A_n ικανοποιώντας μία αρχή μεγάλης απόκλισης με συνάρτηση ρυθμού $I_A(a)$, και θέλουμε να βρούμε τη συνάρτηση ρυθμού μίας άλλης τ. μ. B_n όπου $B_n = h(A_n)$, όπου h είναι μία συνεχής συνάρτηση. Ονομάζουμε h συστολή της A_n. Για να υπολογίσουμε τη συνάρτηση ρυθμού της B_n από αυτήν της A_n, χρησιμοποιούμε απλά την αρχή μεγάλης απόκλισης για την A_n και τη προσέγγιση Laplace στο επίπεδο του

$$P(B_n \in db) = \int_{\{a:h(a)=b\}} P(A_n \in da) \qquad \mathbf{(1.60)}$$

για να αποκτήσουμε

$$P(B_n \in db) \asymp \exp\left(-n \cdot inf_{a:h(a)=b} I_A(a)\right) da. \qquad \mathbf{(1.61)}$$

Αυτό δείχνει ότι εάν μία αρχή μεγάλης απόκλισης ισχύει για τη A_n με συνάρτηση ρυθμού $I_A(a)$, τότε μία αρχή μεγάλης απόκλισης ισχύει επίσης κα για τη B_n,

$$P(B_n \in db) \asymp e^{-nI_B(b)} db, \qquad \mathbf{(1.62)}$$

με μία συνάρτηση ρυθμού η οποία δίνεται από

$$I_B(b) = inf_{a:h(a)=b} I_A(a) \qquad \mathbf{(1.63)}$$

Αυτή η γενική αναγωγή από τη μία συνάρτηση ρυθμού στην άλλη είναι αυτό που ονομάζουμε αρχή συστολής. Εάν h είναι μία ένα προς ένα συνάρτηση με την αντίστροφη h^{-1}, τότε $I_B(b) = I_A(h^{-1}(b))$.

Η ερμηνεία της αρχής συστολής θα πρέπει να είναι ξεκάθαρη. Ενώ πιθανότητες στη θεωρία μεγάλης απόκλισης μετρούνται σε εκθετική κλίμακα, η πιθανότητα για οποιαδήποτε μεγάλη διακύμανση θα πρέπει να προσεγγιστεί, ακολουθώντας προσέγγιση Laplace, από τη πιθανότητα του πιο πιθανού γεγονότος, το οποίο οδηγεί σε αυτή τη διακύμανση.

Κεφάλαιο 2ο

Αποτελέσματα στις Μεγάλες Αποκλίσεις – Προσομοίωση και Εφαρμογές

2.1 *Βασικά εργαλεία και αποτελέσματα στις μεγάλες αποκλίσεις*

2.1.1 Συνάρτηση Laplace και εκθετική αλλαγή των μέτρων

Εάν X είναι μία τυχαία μεταβλητή, η οποία παίρνει πραγματικές τιμές, στον $(\Omega, \mathcal{F})$ με κατανομή πιθανότητας $\mu(dx)$, η αθροιστική γεννήτρια συνάρτηση της μ είναι ο λογάριθμος της εξίσωσης Laplace της X, για παράδειγμα:

$$\lambda(k) = \ln \mathbb{E}\,[e^{kX}] = \ln \int e^{kx} \mu(dx) \quad \in \quad (-\infty, \infty], \quad k \in \mathbb{R}.$$

Παρατηρούμε ότι $\lambda(0) = 0$, και η λ είναι κυρτή από την ανισότητα Hölder. Δηλώνουμε ότι $D(\lambda) = \{k \in \mathbb{R} : \lambda(k) < \infty\}$, και για κάθε $k \in D(\lambda)$, ορίζουμε μία νέα κατανομή πιθανότητας μ_k στον $\mathbb{R}$ ως:

$$\mu_k(dx) = \exp\big(kx - \lambda(k)\big)\, \mu(dx) \tag{2.1}$$

Υποθέτουμε ότι $X_1, \dots, X_n, \dots$, είναι μία ακολουθία ανεξάρτητων και όμοια κατανεμημένων τυχαίων μεταβλητών με κατανομή μ και θεωρούμε το νέο μέτρο πιθανότητας $\mathbb{P}_k$ στο $(\Omega, \mathcal{F})$ με λόγο πιθανοφάνειας:

$$\frac{d\mathbb{P}_k}{d\mathbb{P}}(X_1, \dots, X_n) = \prod_{i=1}^{n} \frac{d\mu_k}{d\mu}(X_i) = \exp\left(k \sum_{i=1}^{n} X_i - n\lambda(k)\right). \tag{2.2}$$

Δηλώνοντας ότι $\mathbb{E}_k$ είναι η αντίστοιχη μέση τιμή υπό το μέτρο πιθανότητας $\mathbb{P}_k$, ο τύπος (2.2) μας δίνει για όλα τα $n \in \mathbb{N}^*$,

$$\mathbb{E}[f(X_1, \dots, X_n] = \mathbb{E}_k[f(X_1, \dots, X_n) exp(-k \sum_{i=1}^{n} X_i + n\lambda(k))], \tag{2.3}$$

για όλες τις συναρτήσεις Borel f για τις οποίες η μέση τιμή είναι πεπερασμένη. Εξάλλου, οι τυχαίες μεταβλητές $X_1, \dots, X_n$, $n \in \mathbb{N}^*$, είναι ανεξάρτητες και όμοια κατανεμημένες με κατανομή πιθανότητας μ_k υπό το μέτρο πιθανότητας. Πραγματικά, η σχέση (2.3) εκτείνεται από έναν σταθερό αριθμό βημάτων n σε έναν τυχαίο αριθμό βημάτων. Ακριβέστερα, ένα τ είναι ένας χρόνος στάσης στο σύνολο $\mathbb{N}$ για $X_1, \dots, X_n, \ldots$, το γεγονός $\{\tau < n\}$ είναι μετρήσιμο σε σχέση με την άλγεβρα που δημιουργείται από $\{X_1, \dots, X_n\}$ για όλα τα n, τότε

$$\mathbb{E}[f(X_1, \dots, X_\tau) 1_{\tau<\infty}] = \mathbb{E}_k[f(X_1, \dots, X_\tau) exp(-k \sum_{i=1}^{\tau} X_i + \tau\lambda(k)) 1_{\tau<\infty}], \tag{2.4}$$

για όλες τις συναρτήσεις Borel f για τις οποίες η μέση τιμή είναι πεπερασμένη.

Η αθροιστική γεννήτρια συνάρτηση $\lambda(k)$ δείχνει κάποιες χρήσιμες πληροφορίες στη κατανομή πιθανότητας μ_k. Για παράδειγμα, το $\lambda'(k)$ είναι η μέση τιμή της μ_k. Όντως, για κάθε k στο εσωτερικό του $D(\lambda)$:

$$\lambda'(k) = \frac{\mathbb{E}[Xe^{kX}]}{\mathbb{E}[e^{kX}]} = \mathbb{E}[Xexp(kX - \lambda(k))] = \mathbb{E}_k[X]. \quad \textbf{(2.5)}$$

Ένας παρόμοιος υπολογισμός δείχνει ότι το $\lambda''(k)$ είναι η διακύμανση του μ_k. Παρατηρούμε ιδίως ότι εάν η τιμή 0 υπάρχει στο εσωτερικό του $D(\lambda)$, τότε $\lambda'(0) = \mathbb{E}[X]$ και $\lambda''(0) = Var(X)$.

Παράδειγμα 2.1.1.1 (Κατανομή Poisson)

Έστω μ η κατανομή Poisson με σταθερά λ. Θέλουμε να υπολογίσουμε την αθροιστική γεννήτρια συνάρτηση (c.g.f.) μέσω της ροπο-γεννήτριας συνάρτησης μιας τυχαίας μεταβλητής X που ακολουθεί τη συγκεκριμένη κατανομή.

$$M(k) = \mathbb{E}(\exp(kX)) = \sum_{x=0}^{\infty} e^{kx} e^{-\lambda} \frac{\lambda^x}{x!} = \exp(-\lambda) \sum \frac{[e^k \lambda]^x}{x!} = \exp(-\lambda) \exp[e^k \lambda]$$
$$= \exp\,[-\lambda + e^k \lambda]$$

Η αθροιστική γεννήτρια συνάρτηση δίνεται από τη σχέση

$$\lambda(k) = \ln\big(M(k)\big) = -\,\lambda + e^k \lambda = \lambda(e^k - 1).$$

Ένας άμεσος, απλός αλγεβρικός υπολογισμός δείχνει ότι μ_k είναι η κατανομή Poisson με σταθερά λe^k. Γι' αυτό το λόγο, η συνέπεια της αλλαγής του μέτρου πιθανότητας $\mathbb{P}_k$ είναι να πολλαπλασιάσουμε τη σταθερά με τον παράγοντα e^k. Επίσης, η αντίστοιχη συνάρτηση ρυθμού είναι

$$I(\alpha) = \alpha \ln\left(\frac{\alpha}{\lambda}\right) + \lambda - \alpha \text{ } \gamma\iota\alpha \text{ } \alpha \geq 0 \text{ } \kappa\alpha\iota \text{ } \infty \text{ } \sigma\varepsilon \text{ } \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\acute{\eta} \text{ } \pi\varepsilon\rho\acute{\iota}\pi\tau\omega\sigma\eta \quad \textbf{(2.6)}$$

Παράδειγμα 2.1.1.2 (Κανονική Κατανομή)

Έστω μ η Κανονική κατανομή $\mathcal{N}(0, \sigma^2)$, της οποίας η αθροιστική γεννήτρια συνάρτηση δίνεται από το τύπο

$$\lambda(k) = \frac{k^2 \sigma^2}{2} \quad \textbf{(2.7)}$$

με αντίστοιχη συνάρτηση ρυθμού

$$I(\alpha) = \frac{\alpha^2}{2\sigma^2}, \alpha \in \mathbb{R} \quad \textbf{(2.8)}$$

Παράδειγμα 2.1.1.3 (Εκθετική Κατανομή)

Έστω μ η εκθετική κατανομή σταθεράς λ. Η αθροιστική γεννήτρια συνάρτησή της δίνεται από το τύπο

$$\lambda(k) = \begin{cases} \ln\left(\frac{\lambda}{\lambda-k}\right), & k < \lambda \\ \infty, & k \geq \lambda \end{cases} \qquad \textbf{(2.9)}$$

Η αντίστοιχη συνάρτηση ρυθμού είναι

$$I(\alpha) = \lambda\alpha - 1 - \ln(\lambda\alpha) \ \text{για} \ \alpha > 0 \ \text{και} \ I(\alpha) = \infty \ \text{σε διαφορετική περίπτωση} \qquad \textbf{(2.10)}$$

Παράδειγμα 2.1.1.4 (Κατανομή Bernoulli)

Έστω μ η κατανομή Bernoulli με παράμετρο p. Η αθροιστική γεννήτρια συνάρτησή της δίνεται από το τύπο

$$\lambda(k) = \ln(1 - p - pe^k). \qquad \textbf{(2.11)}$$

Η συνάρτηση ρυθμού είναι η εξής:

$$I(\alpha) = \alpha \ln\left(\frac{\alpha}{p}\right) + (1 - a)\ln\left(\frac{1-\alpha}{1-p}\right) \text{για} \ \alpha \in [0,1] \text{και} \ \infty \ \text{σε διαφορετική περίπτωση} \qquad \textbf{(2.12)}$$

2.1.2 Ασυμπτωτική συμπεριφορά πιθανοτήτων στη Θεωρία Μεγάλων Αποκλίσεων

Οι πιθανότητες $P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq c)$, για οποιοδήποτε $c \geq 0$, παρουσιάζουν ασυμπτωτικό ρυθμό σύγκλισης, γι' αυτό τον λόγο υπολογίζουμε τη συνάρτηση ρυθμού προσεγγίζοντας το όριο $\lim_{n\to\infty} \frac{1}{n} \ln P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq c\right] = -I(c)$. Η ασυμπτωτική επίλυση του ολοκληρώματος, επειδή δεν είναι δυνατόν να υπολογιστεί σε κλειστή μορφή, γίνεται με την μέθοδο του Laplace για να το προσεγγίσουμε. Το αποτέλεσμα αυτής της μεθόδου πρέπει να συμφωνεί με το πιθανοθεωρητικό αποτέλεσμά του.

Ολοκληρώματα Laplace

Θέλουμε να μελετήσουμε έναν τύπο ολοκληρωμάτων που έχουν τη μορφή

$$I(\lambda) = \int_a^b f(t)e^{-\lambda g(t)}dt, \qquad \lambda \gg 1, \qquad \textbf{(2.13)}$$

όπου g είναι μια γνησίως αύξουσα συνάρτηση στο $[a, b]$ με συνεχή παράγωγο g'. Εδώ, $a < b \leq +\infty$, και γράφοντας $\lambda \gg 1$ εννοούμε "καθώς το $\lambda \to +\infty$". Για την ακρίβεια, αρκεί να εξετάσουμε ολοκληρώματα της μορφής

$$I(\lambda) = \int_0^b f(t) e^{-\lambda t}\, dt, \qquad \lambda \gg 1. \qquad \mathbf{(2.14)}$$

Για να διαπιστώσουμε αυτό, αρκεί να κάνουμε την αλλαγή μεταβλητών $s = g(t) - g(a)$ στη σχέση (2.13), οπότε

$$I(\lambda) = e^{-\lambda g(a)} \int_0^{g(b)-g(a)} \frac{f(t(s))}{g'(t(s))} e^{-\lambda s} ds, \qquad \mathbf{(2.15)}$$

όπου $t = t(s)$ είναι η λύση της εξίσωσης $s = g(t) - g(a)$ ως προς t. Παρατηρούμε τώρα ότι το ολοκλήρωμα στο δεύτερο μέλος είναι της μορφής (2.14). Όταν $b = \infty$, το ολοκλήρωμα στη σχέση (2.14) είναι ο μετασχηματισμός Laplace της $f(t)$.

Η βασική ιδέα για την προσέγγιση του ολοκληρώματος (2.14) είναι να προσδιορίσουμε ποιο υποδιάστημα μας δίνει την κυρίαρχη συνεισφορά στο ολοκλήρωμα. Παρατηρούμε ότι η συνάρτηση $e^{-\lambda t}$ φθίνει πολύ γρήγορα (εκθετικά). Άρα, αν η f δεν αυξάνει πολύ γρήγορα στο άπειρο και συμπεριφέρεται αρκετά καλά κοντά στο $t = 0$, τότε είναι λογικό να υποθέσουμε ότι η κύρια συνεισφορά προέρχεται από μια περιοχή του $t = 0$.

Παράδειγμα 2.1.2.1 (Εκθετική Κατανομή)

Έστω $P(X > c) = \int_c^{+\infty} \lambda e^{-\lambda x} dx$ το ασυμπτωτικό ολοκλήρωμα που θέλουμε να προσεγγίσουμε για να υπολογίσουμε τη πιθανότητα μίας τυχαίας μεταβλητής X (η οποία μπορεί να είναι ο δειγματικός μέσος S_n), η οποία ακολουθεί την εκθετική κατανομή με σταθερά λ. Με τη μέθοδο του Laplace έχουμε

$$P(X > c) = \int_c^{+\infty} \lambda e^{-\lambda x} dx \xRightarrow{s=x-c} P(X > c) = \lambda e^{-\lambda c} \int_0^{+\infty} e^{-\lambda s} ds = e^{-\lambda c}$$

Το πιθανοθεωρητικό αποτέλεσμα είναι

$$P(X > c) = \int_c^{+\infty} \lambda e^{-\lambda x} dx = -\frac{1}{\lambda} \lambda e^{-\lambda x} \Big|_c^{+\infty} = e^{-\lambda c}$$

Άρα βλέπουμε ότι τα δύο αποτελέσματα συμπίπτουν.

Παράδειγμα 2.1.2.2 (Κανονική Κατανομή)

Έστω $X\ \tau.\mu. \sim \mathcal{N}(0,1)$ και θέλουμε να προσεγγίσουμε τη πιθανότητα $P(X > c) = \int_c^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. Με τη μέθοδο του Laplace έχουμε

$$P(X > c) = \int_c^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \xRightarrow{y=\frac{x}{a} \Rightarrow x=ay \Rightarrow dx=ady} P(X > c) = \frac{1}{\sqrt{2\pi}} \int_{\frac{c}{a}}^{+\infty} a e^{-\frac{a^2 y^2}{2}} dy, \acute{o}\pi o\upsilon\ \alpha \gg 1$$

$$P(X>c)=\frac{1}{\sqrt{2}\pi}\int_{\frac{c}{a}}^{+\infty} ae^{-\frac{a^2y^2}{2}}dy \xRightarrow{t=\frac{y^2}{2}\Rightarrow y=\sqrt{2t}\Rightarrow dy=\frac{\sqrt{2}}{2\sqrt{t}}dt} P(X>c)=\frac{a}{2\sqrt{\pi}}\int_{\frac{c^2}{2a^2}}^{+\infty}\frac{1}{\sqrt{t}}e^{-a^2t}dt$$

$$P(X>c)=\frac{a}{2\sqrt{\pi}}\int_{\frac{c^2}{2a^2}}^{+\infty}\frac{1}{\sqrt{t}}e^{-a^2t}dt \xRightarrow{s=t-\frac{c^2}{2a^2}\Rightarrow t=s+\frac{c^2}{2a^2}\Rightarrow dt=ds}$$

$$\Rightarrow P(X>c)=\frac{a}{2\sqrt{\pi}}e^{-\lambda\frac{c^2}{2a^2}}\int_{0}^{+\infty}\frac{1}{\sqrt{s+\frac{c^2}{2a^2}}}e^{-\lambda(s+\frac{c^2}{2a^2})}ds \Rightarrow$$

$$\Rightarrow P(X>c)=\frac{a}{2\sqrt{\pi}}e^{-\lambda\frac{c^2}{2a^2}}\frac{\Gamma\left(\frac{1}{2}\right)}{\lambda^{\frac{1}{2}}} \xRightarrow{\lambda=\alpha^2} P(X>c)=\frac{e^{-\frac{c^2}{2}}}{2}$$

Εάν θέσουμε $c=0\Rightarrow P(X>c)=0.5$. Από τον πίνακα τιμών της κανονικής κατανομής $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{x^2}{2}}dx=1-\frac{1}{\sqrt{2\pi}}\int_{z}^{+\infty}e^{-\frac{x^2}{2}}dx$, έχουμε ότι για $z=0\Rightarrow\Phi(z)=0.5$. Άρα τα δύο αποτελέσματα συμπίπτουν.

Το παραπάνω αποτέλεσμα μπορεί να συνδυαστεί με τη συμπεριφορά του δειγματικού μέσου $S_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ ανεξάρτητων, όμοια κατανεμημένων τυχαίων μεταβλητών. Έστω $X_1,X_2,\dots,X_n$ είναι μία ακολουθία ανεξάρτητων τυχαίων μεταβλητών, οι οποίες ακολουθούν την κανονική κατανομή $\mathcal{N}(0,1)$. Αφού ο δειγματικός μέσος S_n είναι ξανά μία τ. μ.$\sim\mathcal{N}(0,1/n)$, τότε για κάθε $c\geq 0$,

$$P(S_n\geq c)\xrightarrow[n\to\infty]{}0,$$

και, για κάθε διάστημα A,

$$P\left(\sqrt{n}S_n\in A\right)\xrightarrow[n\to\infty]{}\frac{1}{\sqrt{2\pi}}\int_A e^{-\frac{x^2}{2}}dx.$$

Σημειώνουμε τώρα ότι

$$P(S_n\geq c)=1-\frac{1}{\sqrt{2\pi}}\int_{-c\sqrt{n}}^{c\sqrt{n}}e^{-\frac{x^2}{2}}dx\Rightarrow\frac{1}{n}lnP(S_n\geq c)\xrightarrow[n\to\infty]{}-\frac{c^2}{2}.$$

Το παραπάνω αποτέλεσμα είναι ένα παράδειγμα αποτελέσματος της θεωρίας των μεγάλων αποκλίσεων: Η τυπική τιμή του δειγματικού μέσου είναι της τάξεως $1/\sqrt{n}$, αλλά με μικρή πιθανότητα (της τάξεως του $e^{-nc^2/2}$), ο δειγματικός μέσος παίρνει σχετικά μεγάλες τιμές. Γενικά, το συμπέρασμα στο οποίο καταλήγουμε είναι ότι εφόσον το όριο του $\frac{1}{n}lnP(S_n\geq c)$ υπάρχει πάντα, η τιμή του εξαρτάται από τη κατανομή, την οποία ακολουθεί η τ. μ. X_i.

2.2 Προσομοίωση αποτελεσμάτων στις μεγάλες αποκλίσεις

2.2.1 Αριθμητική εκτίμηση πιθανοτήτων μεγάλων αποκλίσεων

Οι συναρτήσεις ρυθμού δεν μπορούν να υπολογιστούν ακριβώς από πολλές στοχαστικές διαδικασίες. Ακριβή και σαφή αποτελέσματα συναρτήσεων ρυθμού μπορούν να βρεθούν μόνο σε λίγες απλές περιπτώσεις. Για τις περισσότερες στοχαστικές διαδικασίες επιστημονικού ενδιαφέροντος, πρέπει να βασιστούμε σε αναλυτικές προσεγγίσεις και αριθμητικές μεθόδους για να υπολογίσουμε τις συναρτήσεις ρυθμού.

2.2.1.1 Άμεση δειγματοληψία (Direct sampling)

Το πρόβλημα το οποίο μας απασχολεί είναι να αποκτήσουμε μία αριθμητική εκτίμηση της συνάρτησης πυκνότητας πιθανότητας (σ.π.π.) $p_{S_n}(s)$ για μία πραγματική τυχαία μεταβλητή S_n ικανοποιώντας μία αρχή μεγάλης απόκλισης, και να βρούμε από αυτή μία εκτίμηση της συνάρτησης ρυθμού $I(s)$. Γενικά, παίρνουμε S_n να είναι μία συνάρτηση από n τ. μ. $X_1, X_2, \dots, X_n$. Για να απλοποιήσουμε το συμβολισμό, θα χρησιμοποιήσουμε τη συντομογραφία $\omega = X_1, X_2, \dots, X_n$.

Αριθμητικά, δεν μπορούμε να αποκτήσουμε $p_{S_n}(s)$ ή $I(s)$ για όλα τα $s \in \mathbb{R}$, αλλά μόνο για έναν πεπερασμένο αριθμό τιμών s, τον οποίο θεωρούμε ότι είναι ίσος με ένα μικρό βήμα Δs. Στο πλαίσιο της αρχής μεγάλης απόκλισης, επιχειρούμε να εκτιμήσουμε την αδροποιημένη (γενικευμένη) σ.π.π

$$p_{S_n}(s) = \frac{P(S_n \in [s, s+\Delta s])}{\Delta s} = \frac{P(S_n \in \Delta_s)}{\Delta s}, \tag{2.16}$$

όπου Δ_s δηλώνει το μικρό διάστημα $[s, s+\Delta s]$ συνδεόμενο με τη τιμή s.

Για να κατασκευάσουμε την εκτίμηση αυτή, ακολουθούμε τη στατιστική δειγματοληψία ή Monte Carlo μέθοδο (statistical sampling or Monte Carlo method), η οποία παρουσιάζεται στα παρακάτω βήματα:

1. Δημιουργούμε ένα δείγμα $\{\omega^{(j)}\}_{j=1}^{M}$ από M «υλοποιήσεις» της ακολουθίας ω για την σ.π.π. του $p(\omega)$.
2. Αποκτούμε από αυτό το δείγμα ένα δείγμα $\{s^{(j)}\}_{j=1}^{M}$ από τιμές για S_n:
$$s^{(j)} = S_n(\omega^{(j)}), \quad j = 1, \dots, L. \tag{2.17}$$
3. Εκτιμούμε $P(S_n \in \Delta_s)$ υπολογίζοντας τον δειγματικό μέσο
$$P_M(\Delta_s) = \frac{1}{L}\sum_{j=1}^{M} \mathbf{1}_{\Delta_s}(s^{(j)}), \tag{2.18}$$
όπου $\mathbf{1}_A(x)$ δηλώνει τη χαρακτηριστική (δείκτρια) συνάρτηση (indicator function) για το σύνολο A, η οποία είναι ίση με 1 εάν $x \in A$ και 0 σε διαφορετική περίπτωση.
4. Αλλάζουμε την εκτιμήτρια (estimator) $P_M(\Delta_s)$ της πιθανότητας $P(S_n \in \Delta_s)$ σε μία εκτιμήτρια $p_M(s)$ της πυκνότητας πιθανότητας $p_{S_n}(s)$:

$$p_M(s) = \frac{P_M(\Delta_s)}{\Delta_s} = \frac{1}{M\Delta_s}\sum_{j=1}^{M} \mathbf{1}_{\Delta_s}(s^{(j)}). \qquad \textbf{(2.19)}$$

Ο λόγος για τον οποίο επιλέξαμε το $p_M(s)$ ως εκτιμήτρια της $p_{S_n}(s)$ είναι ότι αυτή είναι μία αμερόληπτη εκτιμήτρια (unbiased estimator) υπό την έννοια ότι

$$\mathbb{E}[p_M(s)] = p_{S_n}(s) \qquad \textbf{(2.20)}$$

για όλες τις M. Εξάλλου, γνωρίζουμε ότι $p_M(s)$ συγκλίνει στην πιθανότητα να είναι η μέση τιμή της $p_{S_n}(s)$ καθώς $M \to \infty$. Γι' αυτό το λόγο, όσο μεγαλύτερο είναι το δείγμα, τόσο πιο κοντά μπορούμε να πάρουμε μία σωστή εκτίμηση της $p_{S_n}(s)$.

Για να εξάγουμε μία συνάρτηση ρυθμού $I(s)$ από $p_M(s)$, απλά υπολογίζουμε

$$I_{n,M}(s) = -\frac{1}{n} ln p_M(s) \qquad \textbf{(2.21)}$$

και επαναλαμβάνουμε όλη τη διαδικασία για μεγαλύτερες και μεγαλύτερες ακέραιες τιμές του n και M μέχρι η $I_{n,M}(s)$ να συγκλίνει σε ένα συγκεκριμένο επίπεδο ακρίβειας.

2.2.1.2 Δειγματοληψία σημαντικότητας (Importance sampling)

Ένας βασικός κανόνας στη στατιστική δειγματοληψία είναι ότι ένα γεγονός με πιθανότητα P θα εμφανιστεί στο δείγμα μεγέθους L περίπου LP φορές. Γι' αυτό το λόγο για να πάρουμε τουλάχιστον ένα παράδειγμα του γεγονότος αυτού στο δείγμα, πρέπει να έχουμε $L > 1/P$ ως ένα προσεγγιστικό κάτω φράγμα για το μέγεθος του δείγματος.

Εφαρμόζοντας αυτό το αποτέλεσμα στο $p_{S_n}(s)$, βλέπουμε ότι εάν αυτή η σ.π.π ικανοποιεί μία αρχή μεγάλης απόκλισης τύπου $p_{S_n}(s) \approx e^{-nI(s)}$, τότε χρειαζόμαστε να έχουμε $L > e^{nI(s)}$ για να πάρουμε τουλάχιστον ένα παράδειγμα του γεγονότος $S_n \in \Delta_s$ στο δείγμα μας. Με άλλα λόγια, το δείγμα μας πρέπει να είναι εκθετικά μεγάλο με το n για να δούμε οποιαδήποτε μεγάλη απόκλιση.

Αυτός είναι ένας μερικός περιορισμός του δειγματοληπτικού σχήματος και γι' αυτό το λόγο το ονομάζουμε crude Monte Carlo or naive sampling. Ο τρόπος γύρω από αυτό τον περιορισμό είναι να χρησιμοποιήσουμε δειγματοληψία σημαντικότητας (importance sampling – IS), η οποία λειτουργεί ως εξής:

1. Αντί της δειγματοληψίας των X_i σύμφωνα με την σ.π.π. $p(\omega)$, η δειγματοληψία γίνεται σύμφωνα με τη νέα σ.π.π. $q(\omega)$.
2. Υπολογίζουμε αντί της $p_M(s)$ την εκτιμήτρια

$$q_M(s) = \frac{1}{M\Delta_s}\sum_{j=1}^{M} \mathbf{1}_{\Delta_s}\left(S_{\boldsymbol{n}}\left(\omega^{(j)}\right)\right) R(\omega^{(j)}), \qquad \textbf{(2.22)}$$

όπου

$$R(\omega) = \frac{p(\omega)}{q(\omega)} \qquad \textbf{(2.23)}$$

ονομάζεται λόγος πιθανοφάνειας (likelihood ratio). Μαθηματικά, το R αντιστοιχεί στη παράγωγο Radon – Nikodym (Radon – Nikodym derivative) των μέτρων, τα οποία σχετίζονται με τις p και q.

Η νέα εκτιμήτρια $q_M(s)$ είναι μία αμερόληπτη εκτιμήτρια της $p_M(s)$ επειδή

$$\mathbb{E}_q[q_M(s)] = \mathbb{E}_p[p_M(s)]. \tag{2.24}$$

Παρ' όλα αυτά, υπάρχει ένας λόγος για τον οποίο επιλέγουμε την $q_M(s)$ ως τη νέα εκτιμήτρια: μπορούμε να καταλήξουμε με μία κατάλληλη επιλογή για την σ.π.π. q όπως ότι η $q_M(s)$ έχει μικρότερη διακύμανση από την $p_M(s)$. Αυτός είναι ο στόχος της δειγματοληψίας σημαντικότητας: επιλέγουμε $q(\omega)$ μέχρι να ελαχιστοποιήσουμε τη διακύμανση

$$var_q\big(q_M(s)\big) = \mathbb{E}_q[(q_M(s) - p(s))^2]. \tag{2.25}$$

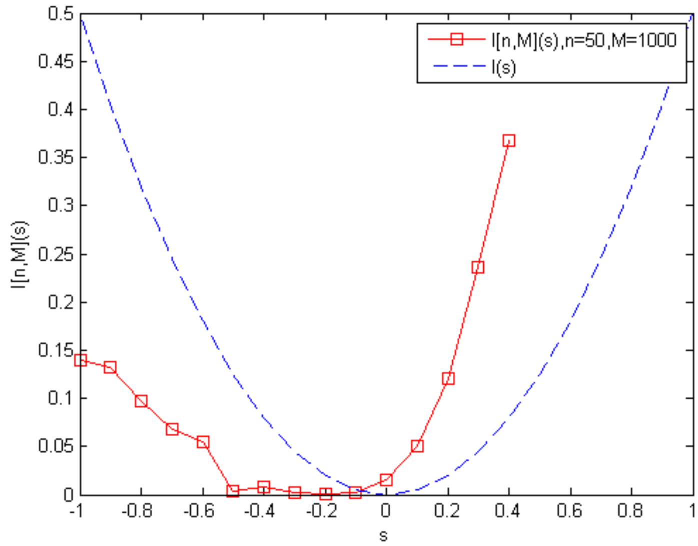
Εάν μπορούμε να επιλέξουμε μία q έτσι ώστε $var_q\big(q_M(s)\big) < var_p\big(p_M(s)\big)$, τότε $q_M(s)$ θα συγκλίνει γρηγορότερα στην $p_{S_n}(s)$ απ' ότι η $p_M(s)$ καθώς αυξάνεται το M.

2.2.2 Προσομοίωση πιθανοτήτων μεγάλων αποκλίσεων

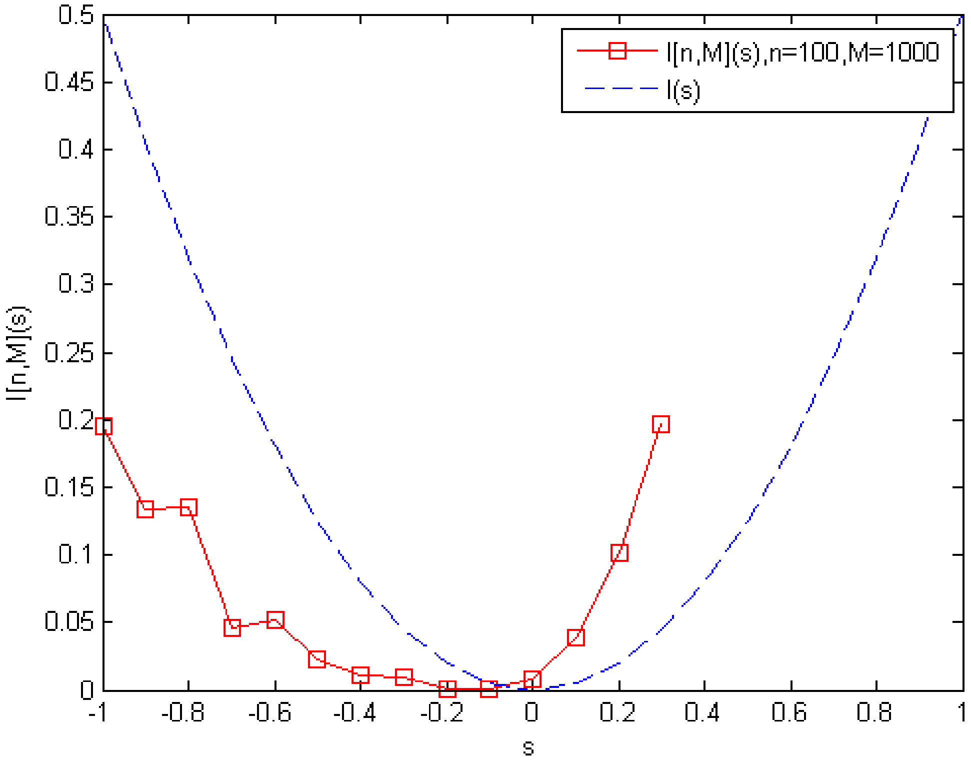
Πραγματοποιήθηκαν αριθμητικά πειράματα εκτίμησης πιθανοτήτων μιας τ. μ. S_n, η οποία είναι ο δειγματικός μέσος για n τ. μ. $X_1, X_2, \ldots, X_n$. Οι τ. μ. $X_1, X_2, \ldots, X_n$ ακολουθούν κάθε φορά και μία διαφορετική κατανομή (έχουμε διαφορετική σ.π.π.). Η Monte Carlo μέθοδος που δημιουργήσαμε χρησιμοποιεί δειγματοληψία σημαντικότητας (importance sampling) και δημιουργήθηκε κώδικας προγράμματος σε Matlab.

2.2.2.1 Κανονική Κατανομή $\mathcal{N}(0, \sigma^2)$

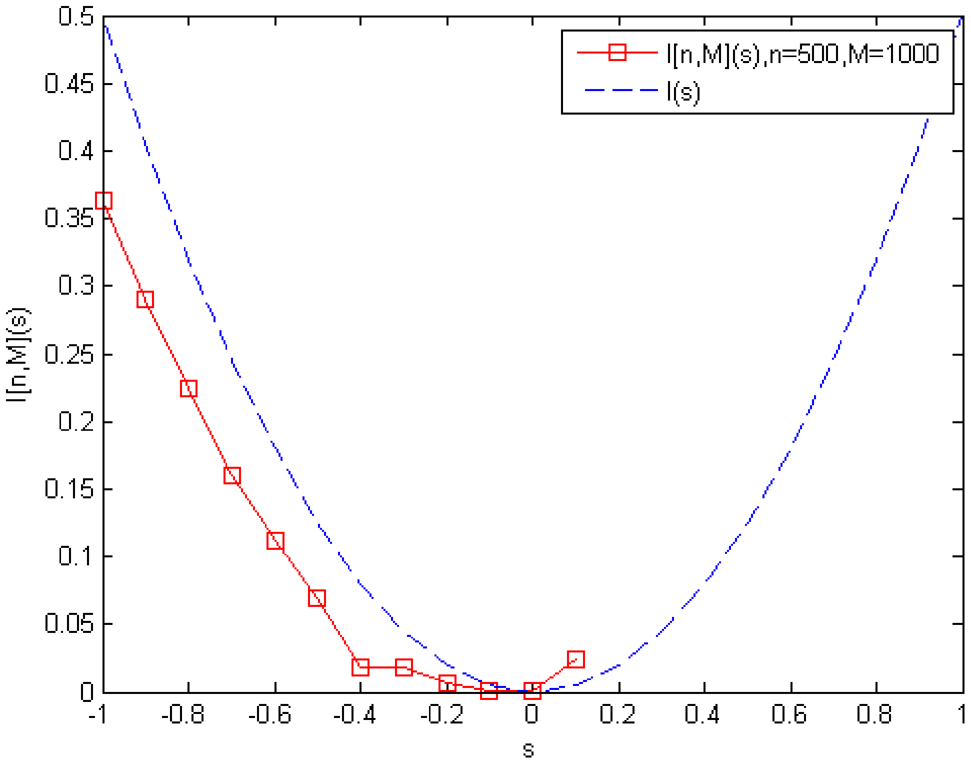
Δειγματοληψία σημαντικότητας (IS) δειγματικού μέσου ανεξάρτητων και όμοια κατανεμημένων τ. μ. κανονικής κατανομής $\mathcal{N}(0,1)$ για n=50 και μέγεθος δείγματος M=1000



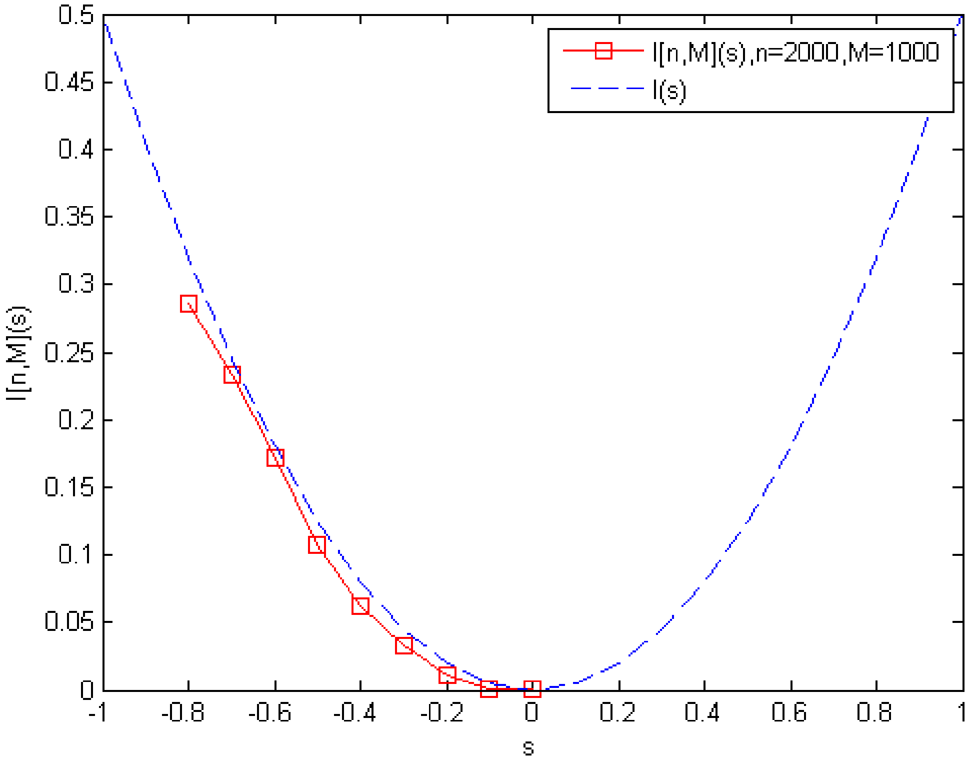
Για n=100 και M=1000:



Για n=500 και M=1000:

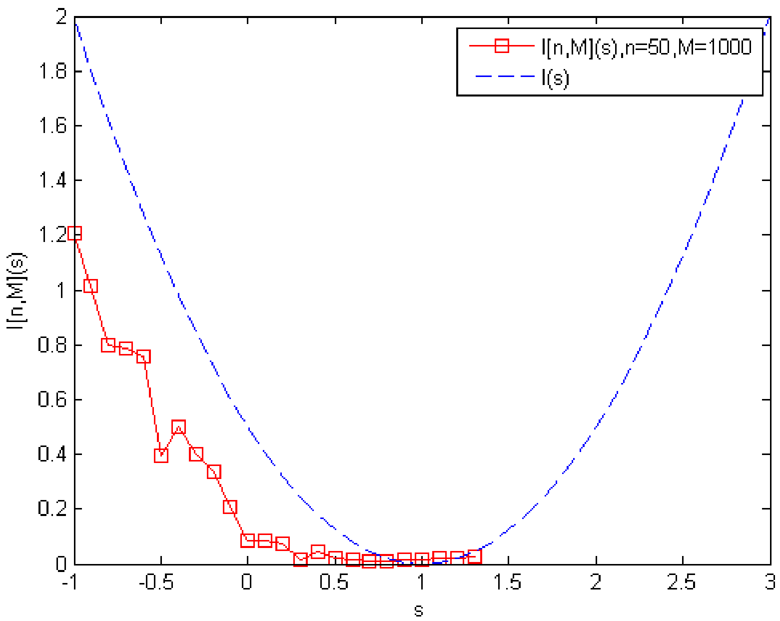


Για n=2000 και M=1000:

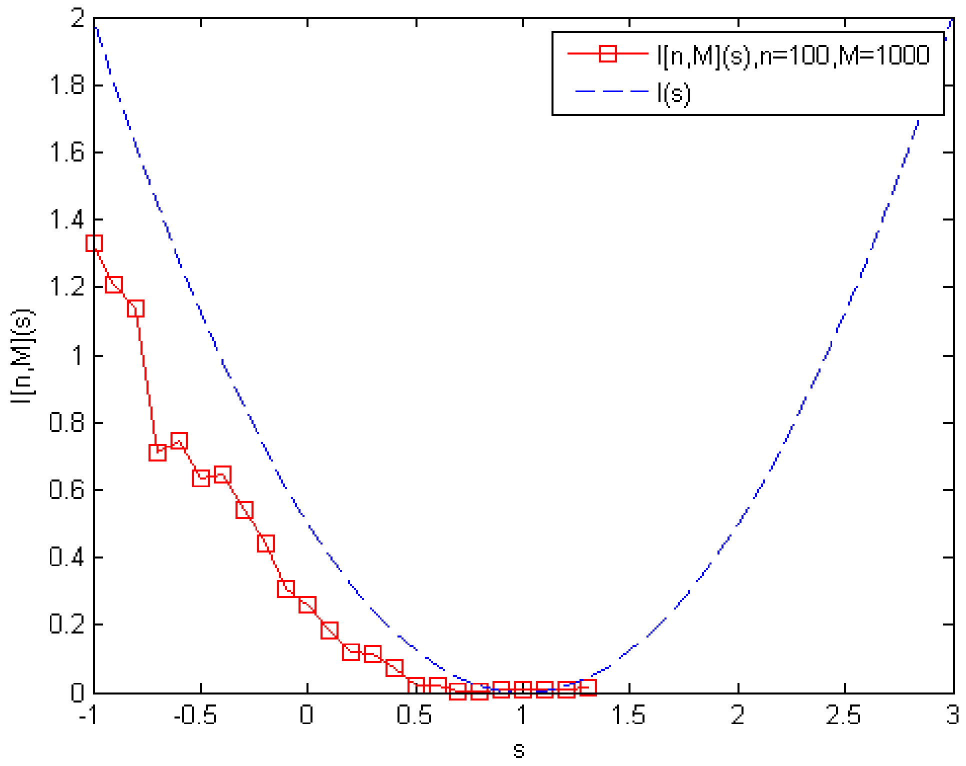


2.2.2.2 Κανονική Κατανομή $\mathcal{N}(\mu, \sigma^2)$

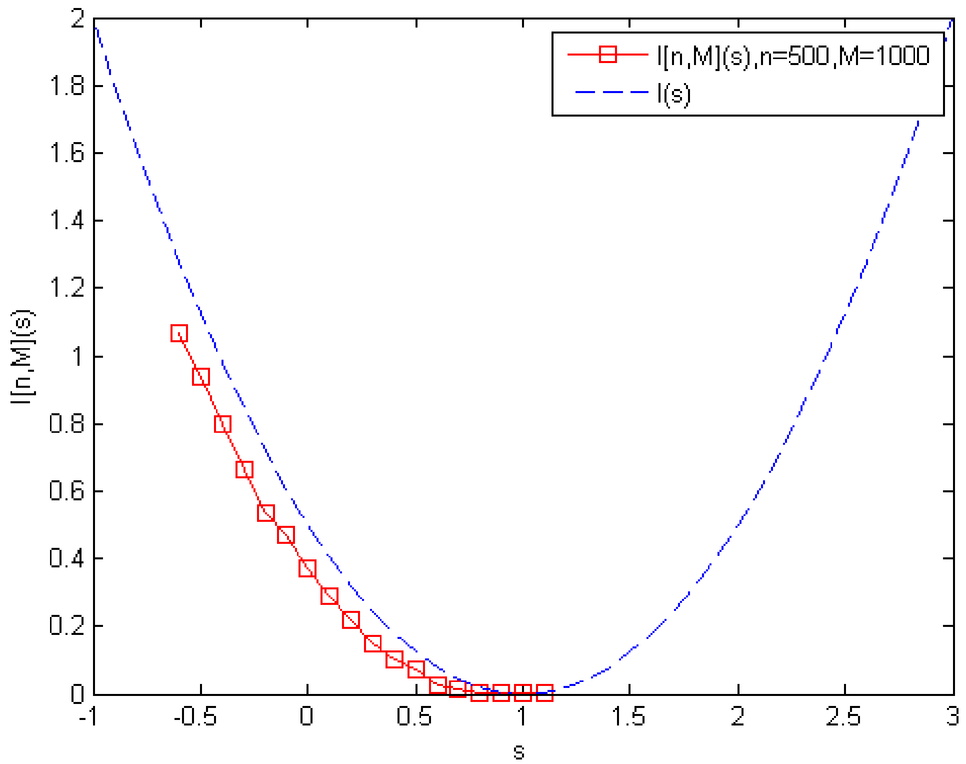
Δειγματοληψία σημαντικότητας (IS) δειγματικού μέσου ανεξάρτητων και όμοια κατανεμημένων τ. μ. κανονικής κατανομής $\mathcal{N}(1,1)$ για n=50 και μέγεθος δείγματος M=1000



Για n=100 και M=1000:

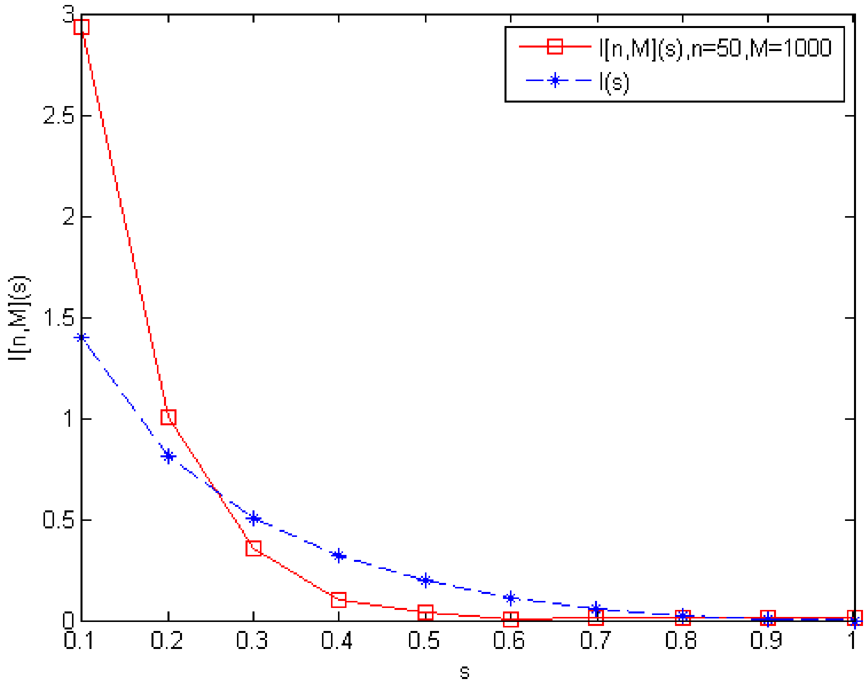


Για n=500 και M=1000:

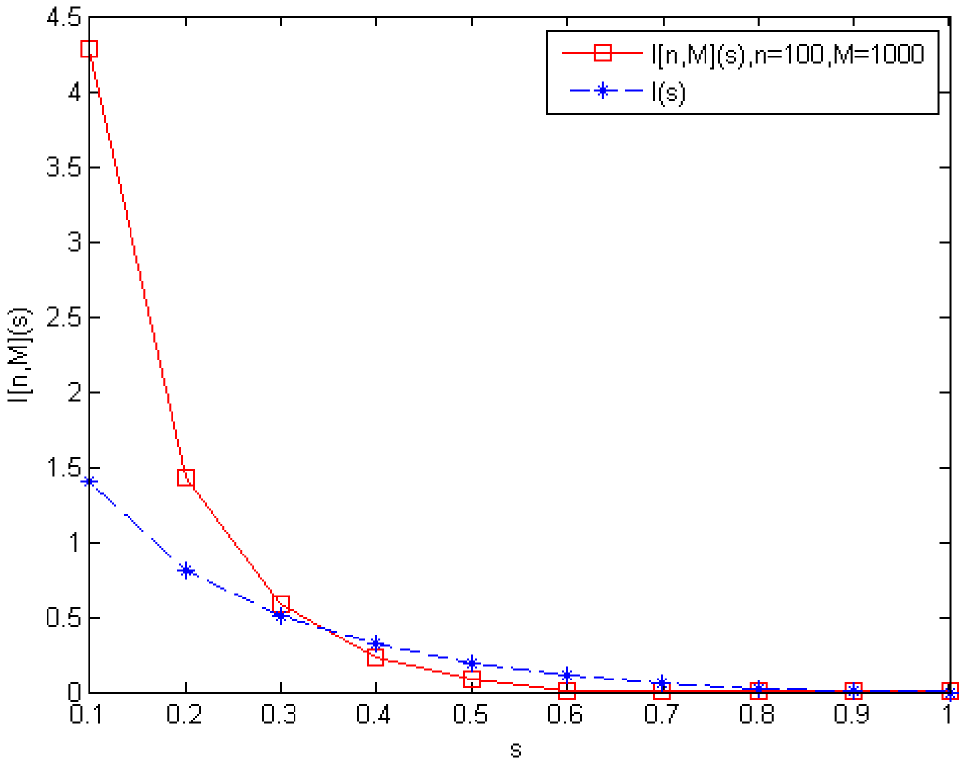


2.2.2.3 Εκθετική Κατανομή $\mathcal{E}(\boldsymbol{\lambda})$

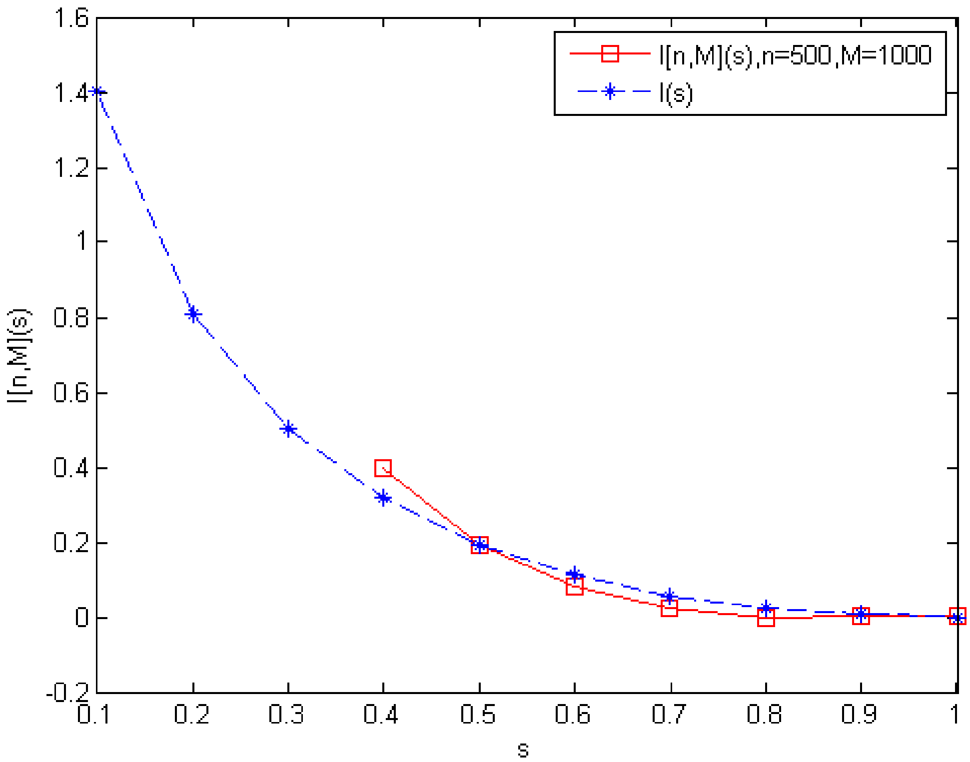
Δειγματοληψία σημαντικότητας (IS) δειγματικού μέσου ανεξάρτητων και όμοια κατανεμημένων τ. μ. εκθετικής κατανομής $\mathcal{E}(1)$ για n=50 και μέγεθος δείγματος M=1000



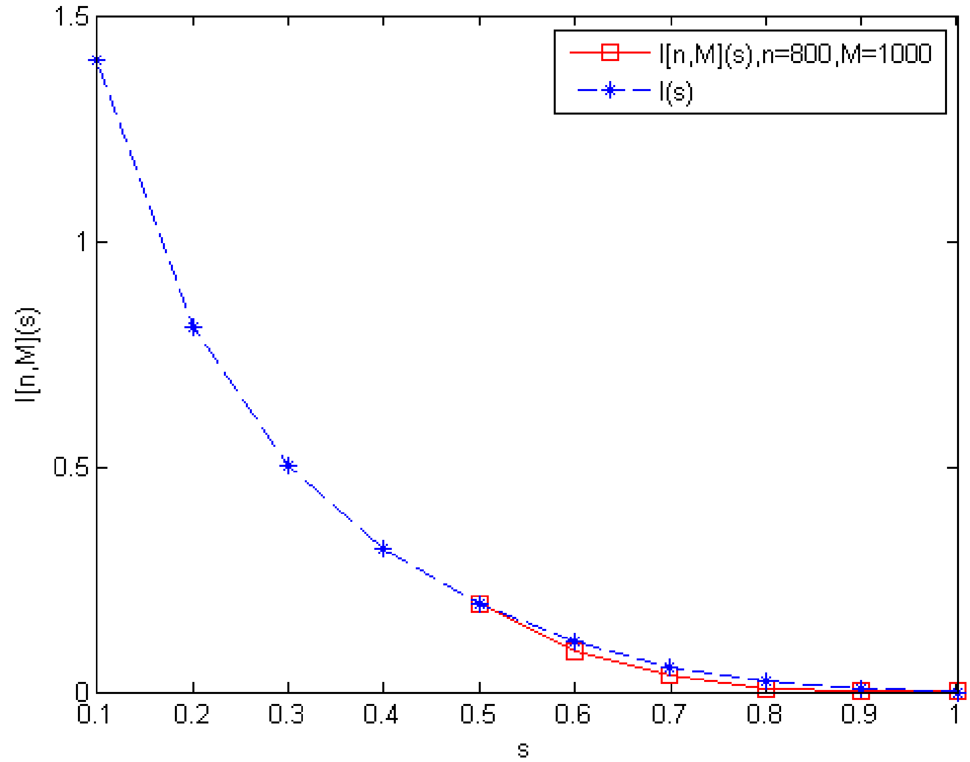
Για n=100 και M=1000:



Για n=500 και M=1000:

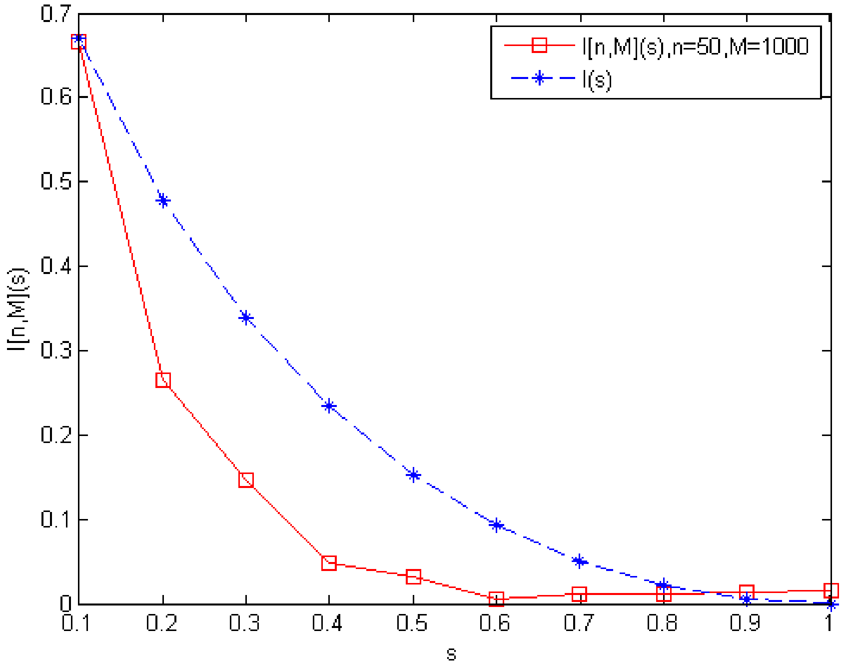


Για n=800 και M=1000:

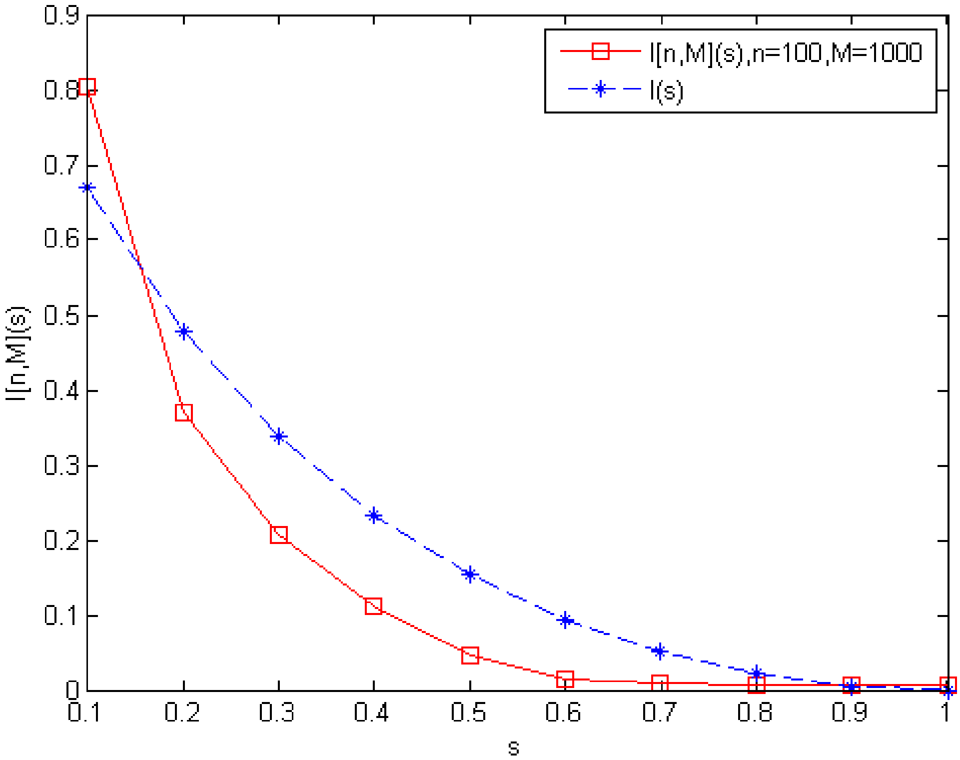


2.2.2.4 Κατανομή Poisson $\mathcal{P}(\lambda)$

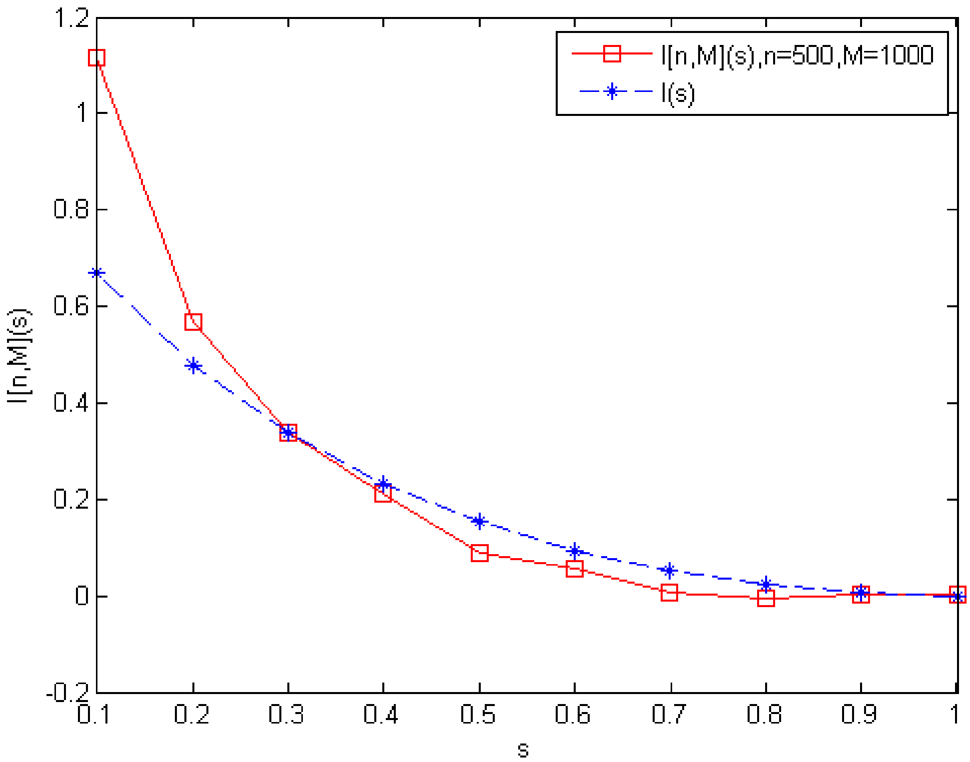
Δειγματοληψία σημαντικότητας (IS) δειγματικού μέσου ανεξάρτητων και όμοια κατανεμημένων τ. μ. εκθετικής κατανομής $\mathcal{P}(1)$ για n=50 και μέγεθος δείγματος M=1000



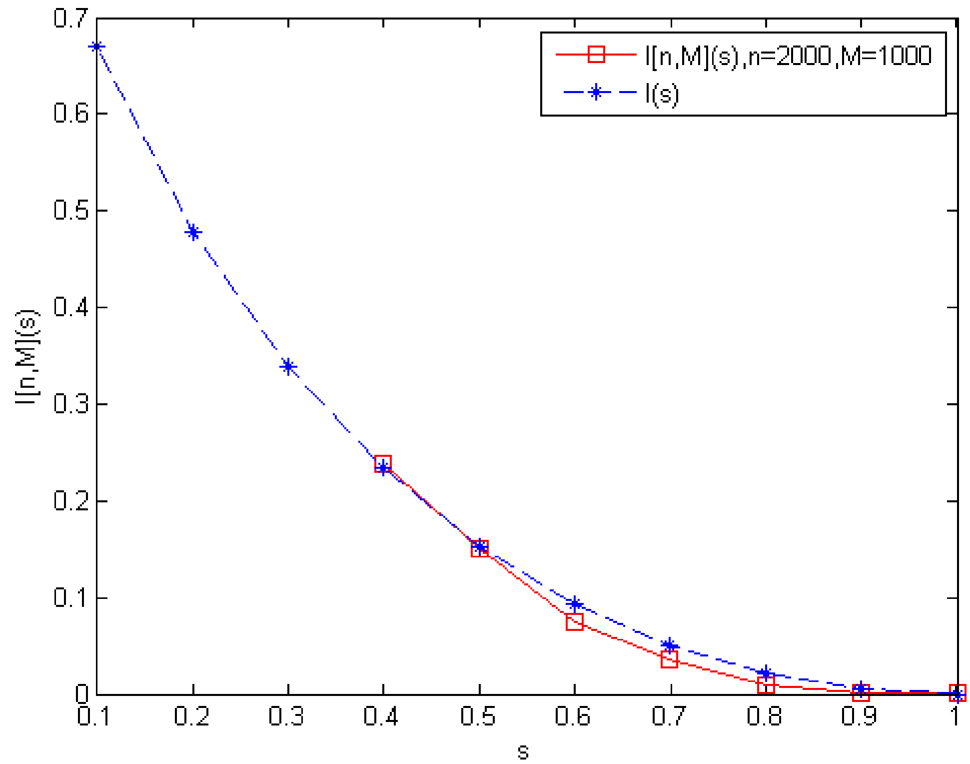
Για n=100 και M=1000:



Για n=500 και M=1000:

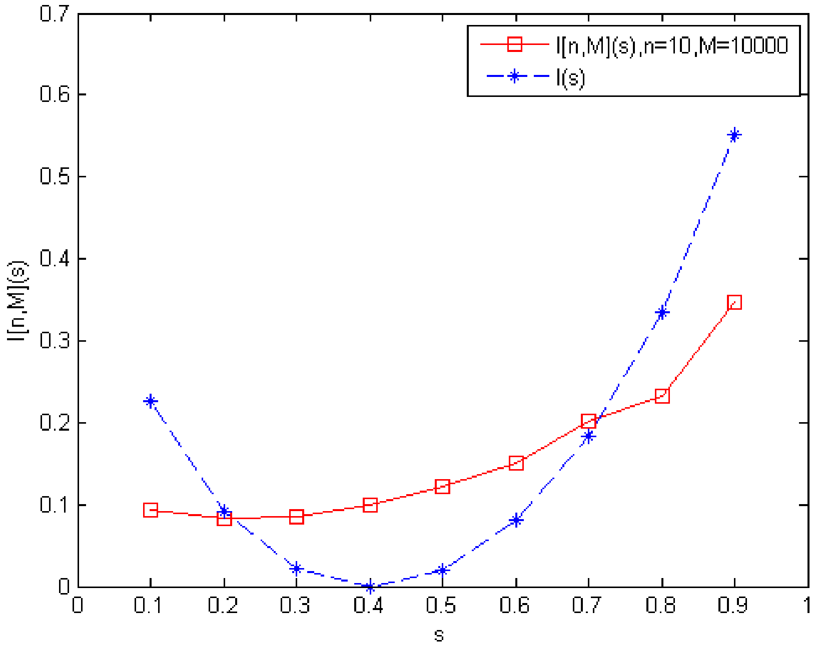


Για n=2000 και M=1000:

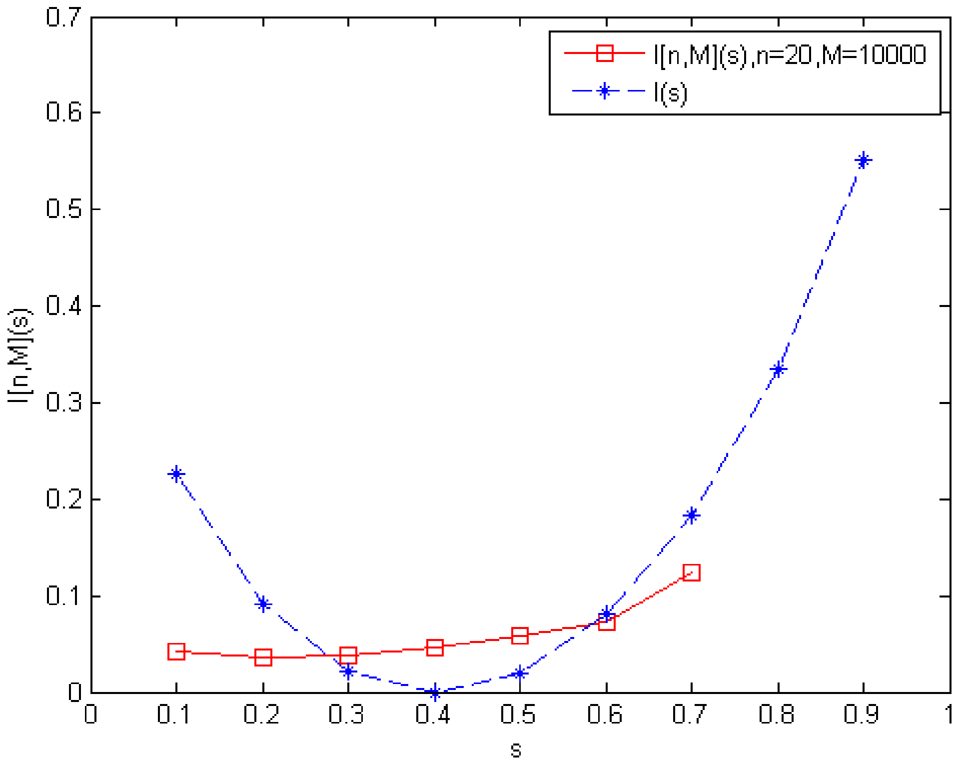


2.2.2.5 Κατανομή Bernoulli $\mathcal{B}(p)$

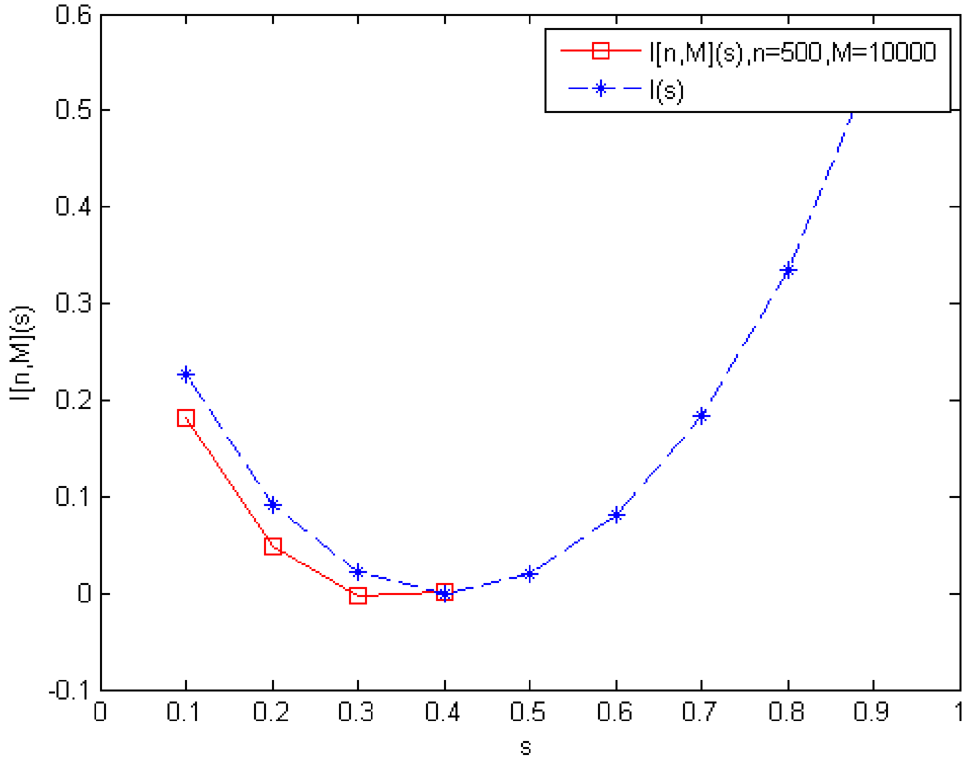
Δειγματοληψία σημαντικότητας (IS) δειγματικού μέσου ανεξάρτητων και όμοια κατανεμημένων τ. μ. κατανομής Bernoulli $\mathcal{B}(0.4)$ για n=10 και μέγεθος δείγματος M=10000



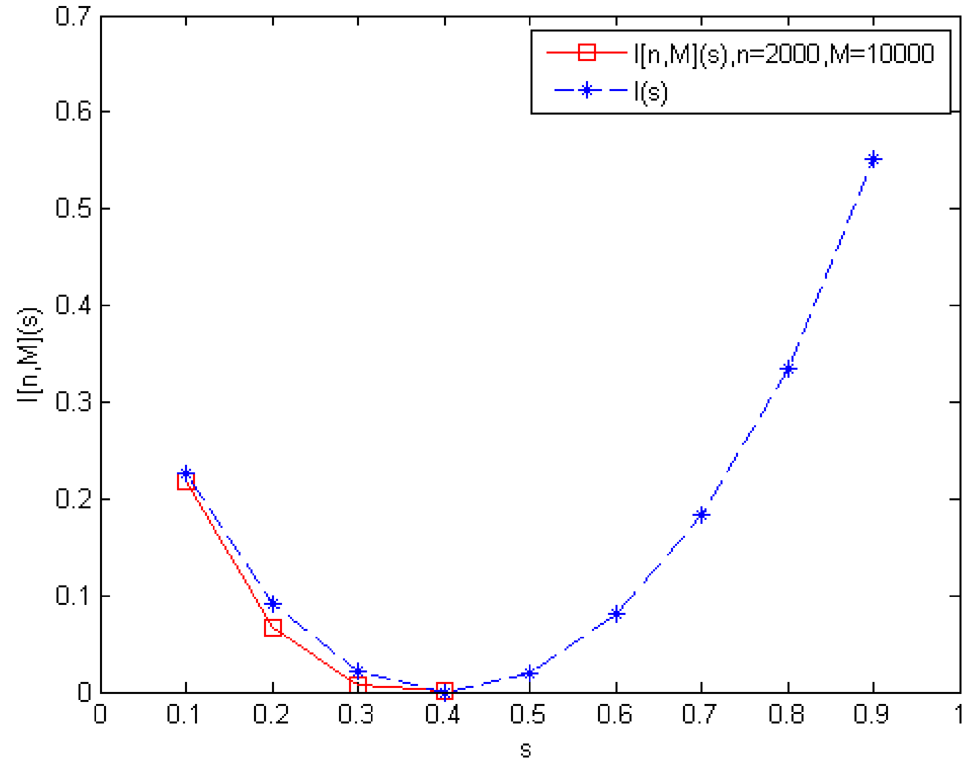
Για n=20 και M=10000:



Για n=500 και M=10000:



Για n=2000 και M=10000:



2.3 *Εφαρμογές των μεγάλων αποκλίσεων*

2.3.1 Φυσικές εφαρμογές

2.3.1.1 Συστήματα σε ισορροπία (Equilibrium systems)

Ο τομέας της στατιστικής μηχανικής συστημάτων σε ισορροπία, όπως ενσωματώνεται από τη θεωρία στατιστικών συνόλων των Boltzmann και Gibbs, μπορεί να θεωρηθεί εκ των υστέρων ως μία θεωρία μεγάλης απόκλισης πολύπλοκων συστημάτων σε ισορροπία. Αυτό γίνεται εμφανές με τη θεώρηση ότι το θερμοδυναμικό όριο είναι ένα όριο μεγάλης απόκλισης, ότι η εντροπία είναι το ισοδύναμο μίας συνάρτησης ρυθμού και η ελεύθερη ενέργεια το ισοδύναμο μίας ανηγμένης αθροιστικής γεννήτριας συνάρτησης (SCGF). Εξάλλου, ο μετασχηματισμός Legendre της θερμοδυναμικής, ο οποίος συνδέει την εντροπία και την ελεύθερη ενέργεια, δεν είναι τίποτα άλλο παρά ο μετασχηματισμός Legendre – Fenchel, ο οποίος συνδέει την συνάρτηση ρυθμού και την ανηγμένη αθροιστική γεννήτρια συνάρτηση στο θεώρημα Gärtner – Ellis και στο θεώρημα Varadhan.

2.3.1.2 Χαοτικά συστήματα και πολύπλοκα στοιχεία fractal γεωμετρίας (Chaotic systems and multifractals)

Ο επονομαζόμενος θερμοδυναμικός φορμαλισμός των δυναμικών συστημάτων, μπορεί επίσης να ερμηνευθεί εκ νέου και εκ των υστέρων ως μία εφαρμογή της θεωρίας μεγάλων αποκλίσεων για τη μελέτη χαοτικών συστημάτων. Υπάρχουν δύο μεγέθη σε αυτή τη θεωρία, τα οποία παίζουν το ρόλο της αθροιστικής γεννήτριας συνάρτησης, τα οποία είναι η τοπολογική πίεση (topological pressure) και η δομική συνάρτηση (structure function). Ο μετασχηματισμός Legendre, ο οποίος εμφανίζεται σε αυτή τη θεωρία είναι ανάλογος αυτού που έχουμε στη θεωρία μεγάλων αποκλίσεων.

2.3.1.3 Συστήματα σε μη ισορροπία (Nonequilibrium systems)

Η θεωρία μεγάλων αποκλίσεων γίνεται ο δεδομένος φορμαλισμός, ο οποίος χρησιμοποιείται για τη μελέτη μη ισορροπημένων συστημάτων, μοντελοποιημένα από στοχαστικές διαφορικές εξισώσεις (SDEs) και διαδικασίες Markov (Markov processes) γενικά. Όντως, η θεωρία των μεγάλων αποκλίσεων βιώνει σήμερα ένα είδος «αναγέννησης» στη φυσική και στα μαθηματικά, ως αποτέλεσμα της αυξανόμενης εφαρμογής της στη στατιστική μηχανική συστημάτων σε μη ισορροπία. Πολλές αρχές μεγάλης απόκλισης εμφανίζονται σε αυτή τη κατηγορία: Αρχές μεγάλης απόκλισης για τις διακυμάνσεις ή για την απασχόληση μοντέλων αλληλεπιδρόντων σωματιδίων, όπως η διαδικασία απαγόρευσης (exclusion process), η διαδικασία μηδενικού εύρους (zero-range process) και πολλές παραλλαγές τους, καθώς επίσης αρχές μεγάλης απόκλισης για μεγέθη που έχουν σχέση με τη παραγωγή εντροπίας για μη ισορροπημένα συστήματα με Στ.Δ.Ε (Στοχαστικές Διαφορικές Εξισώσεις)

2.3.1.4 Σχέσεις διακύμανσης (Fluctuation relations)

Διάφορες αρχές μεγάλης απόκλισης έχουν μελετηθεί τα τελευταία χρόνια για λογαριασμό των σχέσεων διακύμανσης. Για να απεικονίσουμε αυτά τα αποτελέσματα, θεωρούμε τη πρόσθετη διαδικασία S_T, η οποία θεωρούμε ότι παραδέχεται μία αρχή μεγάλης απόκλισης με συνάρτηση ρυθμού $I(s)$. Σε πολλές περιπτώσεις, είναι ενδιαφέρον όχι μόνο να γνωρίζουμε ότι η S_T δέχεται μία αρχή μεγάλης απόκλισης αλλά ότι γνωρίζουμε πως πιθανές, θετικές διακυμάνσεις της S_T συγκρίνονται με τις αρνητικές διακυμάνσεις. Γι' αυτό το σκοπό, είναι σύνηθες να μελετούμε το λόγο

$$\frac{p_{S_T(s)}}{p_{S_T}(-s)} \tag{2.26}$$

ο οποίος γίνεται

$$\frac{p_{S_T(s)}}{p_{S_T}(-s)} \approx e^{T[I(-s)-I(s)]} \tag{2.27}$$

εάν θεωρήσουμε μία αρχή μεγάλης απόκλισης για την S_T. Σε πολλές περιπτώσεις, η διαφορά $I(-s) - I(s)$ είναι γραμμική στο s, και τότε λέμε ότι η S_T ικανοποιεί μία συμβατική σχέση διακύμανσης (conventional fluctuation relation), ενώ εάν είναι μη – γραμμική στο s, τότε μπορούμε να πούμε ότι η S_T ικανοποιεί μία παρατεταμένη σχέση διακύμανσης (extended fluctuation relation). Άλλοι τύποι σχέσεων διακύμανσης έχουν οριστεί επιπλέον από αυτές. Υπάρχει μία λίστα από πολλά φυσικά συστήματα για τα οποία έχουν παρατηρηθεί τέτοιες σχέσεις διακύμανσης.

2.3.2 Χρηματοοικονομικά και ασφαλιστικά μαθηματικά

2.3.2.1 Το κλασικό πρόβλημα της χρεοκοπίας (The classical ruin problem)

Το ασφαλιστικό μοντέλο (the insurance model)

Θεωρούμε μία ασφαλιστική εταιρία κερδίζοντας ασφάλιστρα με έναν σταθερό ρυθμό p ανά μονάδα χρόνου, και ικανοποιώντας αξιώσεις (απαιτήσεις), οι οποίες εμφανίζονται από τη λειτουργία μίας διαδικασίας Poisson με ένταση λ. Συμβολίζουμε με N_t τον αριθμό των αξιώσεων που εμφανίζονται στο διάστημα $[0, t]$, T_n, $n \geq 1$, οι χρόνοι άφιξης της αξίωσης, και με $\xi_1 = T_1, \xi_n = T_n - T_{n-1}, n \geq 2$, οι ενδιάμεσοι χρόνοι άφιξης, οι οποίοι είναι τότε ανεξάρτητοι, όμοια και εκθετικά κατανεμημένοι με πεπερασμένη μέση τιμή $E\xi_1 = 1/\lambda$. Το μέγεθος της νιοστής αξίωσης συμβολίζεται με Y_n, και θεωρούμε ότι τα μεγέθη αξιώσεων Y_n, $n \in \mathbb{N}^*$, είναι (θετικά) ανεξάρτητα και όμοια κατανεμημένα, και ανεξάρτητα από τη διαδικασία Poisson. Αρχίζοντας από ένα αρχικό απόθεμα $x > 0$, η διαδικασία κινδύνου αποθέματος $X_t = X_t^x, t \geq 0$, της ασφαλιστικής εταιρίας δίνεται από:

$$X_t^x = x + pt - \sum_{i=1}^{N_t} Y_i. \tag{2.28}$$

Η πιθανότητα χρεοκοπίας με πεπερασμένο ορίζοντα είναι

$$\psi(x) = \mathbb{P}[\tau_x < \infty],$$

όπου $\tau_x = inf\{t \geq 0 : X_t^x < 0\}$ είναι ο χρόνος της χρεοκοπίας. Ενδιαφερόμαστε για την εκτίμηση της πιθανότητας χρεοκοπίας, συγκεκριμένα για μεγάλες τιμές του αρχικού αποθέματος.

Η εκτίμηση Cramer – Lundberg (Cramer – Lundberg estimate)

Η προσέγγιση Cramer – Lundberg αφορά την εκτίμηση της πιθανότητας χρεοκοπίας $\psi(x)$, και είναι ένα από τα πιο διάσημα αποτελέσματα της θεωρίας κινδύνου. Υπάρχουν διάφορες προσεγγίσεις για να παράγεις ένα τέτοιο αποτέλεσμα. Ακολουθούμε μία μέθοδο βασισμένη σε ορίσματα της θεωρίας μεγάλων αποκλίσεων και αλλαγής των μέτρων πιθανότητας.

Αρχικά, βλέπουμε ότι εύκολα, από το νόμο των μεγάλων αριθμών, ισχύει

$$\frac{1}{t}\sum_{i=1}^{N_t} Y_i \to \rho, \qquad t \to \infty,$$

όπου $\rho = \lambda\mathbb{E}[Y_1] > 0$ ερμηνεύεται ως ο μέσος όρος του ποσού της ασφάλειας ανά μονάδα χρόνου. Η φόρτωση ασφάλειας η παίζει ρόλο – κλειδί στη πιθανότητα χρεοκοπίας. Ορίζεται ως το σχετικό ποσό από το οποίο το ποσοστό ασφαλίστρου υπερβαίνει το ρ:

$$\eta = \frac{p-\rho}{\rho} \Leftrightarrow p = (1+\eta)\rho.$$

Ως εκ τούτου, $X_t^x/t \longrightarrow p - \rho = \rho\eta$ όπου t πηγαίνει στο άπειρο. Γι' αυτό το λόγο, εάν $\eta < 0$, $X_t^x \to -\infty$, και έχουμε $\psi(x) = 1$ όλα τα x. Για $\eta = 0$, μπορούμε να δείξουμε ότι $limsupX_t^x = -\infty$ έτσι ώστε $\psi(x) = 1$. Εν συνεχεία, κάνουμε το καθαρό κέρδος υπόθεση:

$$\eta = \frac{p-\lambda\mathbb{E}[Y_1]}{\lambda\mathbb{E}[Y_1]} > 0, \tag{2.29}$$

το οποίο εξασφαλίζει ότι η πιθανότητα της χρεοκοπίας είναι μικρότερη από 1.

Αφού η χρεοκοπία μπορεί να συμβεί μόνο κατά την άφιξη μίας ασφάλειας, για παράδειγμα, όταν το X κινείται προς τα κάτω, αρκεί να θεωρήσουμε τη χρονικά διακριτή διαδικασία ενσωματωμένη στα άλματα της διαδικασίας Poisson. Τότε ορίζουμε τη χρονικά διακριτή διαδικασία $X_{T_n}^x$, $n \geq 1$, έτσι ώστε

$$\psi(x) = \mathbb{P}[\sigma_x < \infty],$$

Όπου $\sigma_x = inf\{n \geq 1 : X_{T_n}^x < 0\} = inf\{n \geq 1 : S_n > x\}$, και $S_n = x - X_{T_n}^x$ είναι το καθαρό κέρδος πάνω από τη νιοστή ασφάλεια και δίνεται από το τυχαίο περίπατο:

$$S_n = Z_1 + \cdots + Z_n, \quad Z_i = Y_i - p\xi_i, \;\; i \in \mathbb{N}^*.$$

Οι τυχαίες μεταβλητές Z_i είναι ανεξάρτητες και όμοια κατανεμημένες και ικανοποιούν υπό τη συνθήκη του καθαρού κέρδους, $\mathbb{E}[Z_1] < 0$. Δηλώνουμε ότι λ_Z είναι η αθροιστική γεννήτρια συνάρτηση των Z_i, και βλέπουμε από την ανεξαρτησία των Y_i και ξ_i:

$$\lambda_Z(\theta) = \lambda_Y(\theta) + \lambda_\xi(-p\theta) = \lambda_Y(\theta) + \ln\left(\frac{\lambda}{\lambda + p\theta}\right), \quad \theta > -\frac{\lambda}{p},$$

όπου λ_Y είναι η αθροιστική γεννήτρια συνάρτηση των Y_i (αντίστοιχα ισχύει και για την λ_ξ και ξ_i). Για κάθε θ στη περιοχή του λ_Z, θεωρούμε την εκθετική αλλαγή του μέτρου για τη παράμετρο θ, και αφού σ_x είναι ένας χρόνος στάσης στη διήθηση του $(Z_1, \dots, Z_n)$, γράφουμε τη πιθανότητα χρεοκοπίας ως μία $\mathbb{E}_\theta$ εκτίμηση:

$$\psi(x) = \mathbb{P}[\sigma_x < \infty] = \mathbb{E}\left[1_{\sigma_x<\infty}\right] = \mathbb{E}_\theta\left[1_{\sigma_x<\infty}\exp\left(-\theta S_{\sigma_x} + \sigma_x \lambda_Z(\theta)\right)\right]. \qquad \textbf{(2.30)}$$

Θεωρούμε τώρα ότι Y έχει μία συνάρτηση ουράς κατανομής, η οποία είναι ασυμπτωτικά φραγμένη εκθετικά, δηλαδή αυτή η συνάρτηση μειώνεται ως το μηδέν εκθετικά ή και γρηγορότερα. Για παράδειγμα, υπάρχει $\bar{\theta} \in (0, \infty]$ τέτοιο ώστε $\lambda_y(\theta) < \infty$ για $\theta < \bar{\theta}$, και $\lambda_Y(\theta) \to \infty$ καθώς $\theta \nearrow \bar{\theta}$. Σε αυτή τη περίπτωση, η αθροιστική γεννήτρια συνάρτηση λ_Z των Z_i είναι πεπερασμένη στο $\left(-\frac{\lambda}{p}, \bar{\theta}\right)$, είναι διαφορίσιμη στο 0 με $\lambda'_Z(0) = \mathbb{E}[Z_1] < 0$ υπό τη συνθήκη καθαρού κέρδους (2.29). Εξάλλου, αφού $\mathbb{E}[Y_1] > 0$ και Y_1 είναι ανεξάρτητη της ξ_i, βλέπουμε ότι $\mathbb{P}[Z_1 > 0] > 0$, το οποίο συνεπάγεται ότι $\lambda_Z(\theta)$ πηγαίνει στο άπειρο καθώς το θ πηγαίνει στο $\bar{\theta}$. Από τη κυρτότητα της λ_Z και υπενθυμίζοντας ότι $\lambda_Z(0) = 0$, συμπεραίνουμε την ύπαρξη ενός μοναδικού $\theta_L > 0$ έτσι ώστε $\lambda_Z(\theta_L) = 0$. Αυτό το μοναδικό θετικό θ_L είναι η λύση της εξίσωσης Cramer – Lundberg:

$$\lambda_Y(\theta_L) + \ln\left(\frac{\lambda}{\lambda + p\theta_L}\right) = 0,$$

η οποία γράφεται ισοδύναμα:

$$\gamma_Y(\theta_L) = \frac{p\theta_L}{\lambda} \qquad \textbf{(2.31)}$$

όπου $\gamma_Y = \exp\left(\lambda_Y(\theta)\right) - 1 = \int e^{\theta y} F_Y(dy) - 1$ είναι η μετατοπισμένη ροπο-γεννήτρια συνάρτηση των Y_i, και F_Y είναι η συνάρτηση κατανομής των μεγεθών ασφάλειας Y_i. θ_L ονομάζεται συντελεστής προσαρμογής (adjustment coefficient) ή μερικές φορές εκθέτης Lundberg (Lundberg exponent). Τονίζουμε ότι από τη κυρτότητα της λ_Z, έχουμε $\lambda'_Z(\theta_L) > 0$. Γι' αυτό το λόγο, υπό τη $\mathbb{P}_{\theta_L}$, ο τυχαίος περίπατος έχει θετική απόκλιση $\mathbb{E}_{\theta_L}[Z_n] = \lambda'_Z(\theta_L) > 0$, και αυτό συνεπάγεται $\mathbb{P}_{\theta_L}[\sigma_x < \infty] = 1$. Για την επιλογή $\theta = \theta_L$, η σχέση (2.30) γίνεται

$$\psi(x) = \mathbb{E}_{\theta_L}\left[e^{-\theta_L S_{\sigma_x}}\right] = e^{-\theta_L x}\mathbb{E}_{\theta_L}\left[e^{-\theta_L (S_{\sigma_x} - x)}\right]. \qquad \textbf{(2.32)}$$

Σημειώνοντας ότι η υπέρβαση $S_{\sigma_x} - x$ είναι μη αρνητική, παίρνουμε την ανισότητα Lundberg στην πιθανότητα χρεοκοπίας:

$$\psi(x) \leq e^{-\theta_L x}, \qquad \forall x > 0. \qquad \textbf{(2.33)}$$

Εξάλλου, η υπέρβαση $R^x = S_{\sigma_x} - x$ έχει ένα όριο R^∞ (υπό την έννοια της ασθενούς σύγκλισης σε σχέση με τη πιθανότητα $\mathbb{P}_{\theta_L}$), όταν το x πηγαίνει στο άπειρο, και επίσης $\mathbb{E}_{\theta_L}\left[e^{-\theta_L(S_{\sigma_x}-x)}\right]$ συγκλίνει σε μια θετική σταθερά C. Τότε παίρνουμε τη κλασική προσέγγιση για μεγάλες τιμές του αρχικού αποθέματος:

$$\psi(x) \cong Ce^{-\theta_L x},$$

καθώς $x \to \infty$, το οποίο συνεπάγεται μία εκτίμηση τύπου μεγάλων αποκλίσεων

$$\lim_{x\to\infty} \frac{1}{x} \ln \psi(x) = -\theta_L. \qquad \textbf{(2.34)}$$

2.3.2.2 Πιθανότητες χρεοκοπίας και βέλτιστη επένδυση

Το μοντέλο ασφάλισης - χρηματοδότησης (The insurance – finance model)

Στη ρύθμιση του κλασικού μοντέλου που περιγράψαμε στη προηγούμενη παράγραφο, θεωρούμε το πρόσθετο χαρακτηριστικό ότι η ασφαλιστική εταιρεία μπορεί να επενδύσει σε κάποια χρηματιστηριακή αγορά, μοντελοποιημένη από μία γεωμετρική κίνηση Brown:

$$dS_t = bS_t dt + \sigma S_t dW_t,$$

όπου b, σ είναι σταθερές, $\sigma > 0$, και W είναι μία πρότυπη κίνηση Brown, ανεξάρτητη από το απόθεμα χρεοκοπίας X όπως ορίζεται στη σχέση (2.28). Σημειώνουμε ότι $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ είναι η διήθηση, η οποία δημιουργείται από τα X και S. Ο ασφαλιστής μπορεί να επενδύσει οποιαδήποτε χρονική στιγμή t ένα ποσό χρημάτων α_t στο χρηματιστήριο, και το υπόλοιπο σε ομόλογο (στο συγκεκριμένο μοντέλο δεν μας ενδιαφέρει αυτό). Το σύνολο $\mathcal{A}$ των αποδεκτών στρατηγικών επενδύσεων ορίζονται ως το σύνολο των $\mathbb{F}$-προσαρμοσμένων διαδικασιών $a = (a_t)$ έτσι ώστε $\int_0^t a_s^2\, ds < \infty$ σ.β. Δίνοντας ένα αρχικό κεφάλαιο $x \geq 0$, και έναν έλεγχο αποδεκτής επένδυσης a, η διαδικασία συγκέντρωσης πλούτου του ασφαλιστή μπορεί να γραφεί ως

$$V_t^{x,a} = X_t^x + \int_0^t \frac{a_u}{S_u} dS_u = x + pt - \sum_{i=1}^{N_t} Y_i + \int_0^t a_u(bdu + \sigma dW_u), \qquad t \geq 0.$$

Ορίζουμε τη πιθανότητα χρεοκοπίας απείρου χρόνου

$$\psi(x,a) = \mathbb{P}\left[\tau_{x,a} < \infty\right],$$

όπου $\tau_{x,a} = \inf\{t \geq 0 : V_t^{x,a} < 0\}$ είναι ο χρόνος χρεοκοπίας, εξαρτώμενος από τον αρχικό πλούτο x και τη στρατηγική επένδυσης a. Ενδιαφερόμαστε για την ελάχιστη πιθανότητα χρεοκοπίας του ασφαλιστή

$$\psi^*(x) = inf_{a\in\mathcal{A}}\psi(x,a).$$

Ασυμπτωτική εκτίμηση της πιθανότητας χρεοκοπίας

Το κύριο αποτέλεσμα είναι μία ασυμπτωτική εκτίμηση μεγάλης απόκλισης για την ελάχιστη πιθανότητα χρεοκοπίας όπου το αρχικό απόθεμα πηγαίνει στο άπειρο:

<u>Θεώρημα 3</u>: Έχουμε

$$\lim_{x\to\infty} \frac{1}{x} \ln \psi^*(x) = -\theta^*, \qquad \textbf{(2.35)}$$

όπου $\theta^* > 0$ είναι η μοναδική θετική λύση της εξίσωσης

$$\gamma_Y(\theta) = p\frac{\theta}{\lambda} + \frac{b^2}{2\sigma^2\lambda}. \qquad \textbf{(2.36)}$$

Εδώ $\gamma_Y(\theta) = \mathbb{E}\left[e^{\theta Y_1}\right] - 1$ είναι η μετατοπισμένη ροπο-γεννήτρια συνάρτηση του μεγέθους της ασφάλειας. Εξάλλου, η σταθερή στρατηγική $a^* = \frac{b^2}{\sigma^2\theta^*}$ είναι ασυμπτωτικά βέλτιστη υπό την έννοια ότι

$$\lim_{x\to\infty} \frac{1}{x} \ln \psi(x, a^*) = -\theta^*.$$

Τελικά, εάν $b \neq 0, \theta^* > \theta_L$ ο εκθέτης Lundberg.

Η εκτίμηση (2.36) είναι ανάλογη με την κλασική εκτίμηση Lundberg (2.34) χωρίς επένδυση. Ο εκθέτης είναι μεγαλύτερος από αυτόν του Lundberg, και γι’ αυτό παίρνουμε ένα πιο απότομο φράγμα στην ελάχιστη πιθανότητα χρεοκοπίας. Εξάλλου, η εμπορική στρατηγική, αποδίδοντας τη βέλτιστη ασυμπτωτική, εκθετική μείωση, συνίσταται στην εκμετάλλευση ενός σταθερού ποσού στη περιοχή κινδύνου. Αυτό το εκπληκτικό αποτέλεσμα, σε προφανή αντίφαση με τη κοινή άποψη ότι οι «πλούσιες» επιχειρήσεις μπορούν να επενδύουν περισσότερο από τις «φτωχές», εξηγείται από το γεγονός ότι η ελαχιστοποίηση της πιθανότητας χρεοκοπίας είναι ένα εξαιρετικά συντηρητικό κριτήριο.

Κεφάλαιο 3ο

Μεγάλες Αποκλίσεις για Στοχαστικές Διαφορικές Εξισώσεις

3.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα αναφερθούμε στις μεγάλες αποκλίσεις στο επίπεδο των στοχαστικών διαφορικών εξισώσεων (large deviations for stochastic differential equations) στο όριο μικρού θορύβου. Γενικά, μπορούμε να δούμε πώς η διάχυση X_ϵ η οποία λύνει τη Στ.Δ.Ε. (Στοχαστική Διαφορική Εξίσωση – SDE)

$$dX_\epsilon = a(X_\epsilon)dt + \epsilon dW,\ X_\epsilon(0) = x_0 \qquad \textbf{(3.1)}$$

συγκλίνει στην τροχιά $x_0(t)$ λύνοντας τη Συν.Δ.Ε. (Συνήθης Διαφορική Εξίσωση – ODE)

$$\frac{dx}{dt} = a\big(x(t)\big),\ x(0) = x_0 \qquad \textbf{(3.2)}$$

στο όριο του «μικρού θορύβου», $\epsilon \to 0$.

Η θεωρία των μεγάλων αποκλίσεων για τις Στ.Δ.Ε., γνωστή και ως θεωρία Freidlin – Wentzell (Freidlin – Wentzell theory), μας δείχνει ότι η οικογένεια των διαδικασιών X_ϵ υπακούει μία αρχή μεγάλης απόκλισης με ρυθμό ϵ^{-2}, και μία καλή συνάρτηση ρυθμού.

Η πλήρης θεωρία Freidlin – Wentzell προχωράει ακόμη περισσότερο από το επίπεδο των Στ.Δ.Ε, ώστε να θεωρήσουμε διαταραχές μικρού θορύβου δυναμικών συστημάτων πολλών σύντομων διαταραχών από διαδικασίες Markov.

Θέλουμε να αποδείξουμε πρώτα μία αρχή μεγάλης απόκλισης για μία συγκριτικά απλή περίπτωση, και στη συνέχεια να τη μεταφέρουμε σε περισσότερο περίπλοκες διαδικασίες οι οποίες μπορούν να αναπαρασταθούν ως κατάλληλα συναρτησιακά της βασικής περίπτωσης. Γι’ αυτό τον λόγο, η βασική περίπτωση, η οποία θα χρησιμοποιηθεί είναι η διαδικασία Wiener (Wiener process) $W(t)$, με t περιορισμένο στο μοναδιαίο διάστημα $[0,1]$.

3.2 Μεγάλες Αποκλίσεις για τη διαδικασία Wiener

Ξεκινάμε με μία τυπική d-διάστατη διαδικασία Wiener W, και θεωρούμε τη διεύρυνσή της με έναν παράγοντα ϵ, $X_\epsilon(t) = \epsilon W(t)$. Υπάρχει ένας αριθμός τρόπων να αποδείξουμε ότι η διάχυση X_ϵ υπακούει μία αρχή μεγάλης απόκλισης καθώς $\epsilon \to 0$. Μία προσέγγιση αρχίζει με την απόδειξη μίας αρχής μεγάλης απόκλισης για τυχαίους περιπάτους συνεχούς χρόνου (continuous – time random walks), βασιζόμενη στο θεώρημα Gärtner – Ellis, και δείχνοντας ότι η σύγκλιση τέτοιων διαδικασιών στη διαδικασία Wiener είναι αρκετά γρήγορη, η αρχή μεγάλης απόκλισης εξακολουθεί να ισχύει. Για μεγαλύτερη κατανόηση, $\|w\|_\infty$ θα συμβολίζουμε τη νόρμα supremum στο χώρο των συνεχών καμπυλών στο μοναδιαίο διάστημα, $\boldsymbol{C}\big([0,1],\mathbb{R}^{\boldsymbol{d}}\big)$.

<u>Ορισμός 4</u>: (Χώροι Cameron – Martin) Ο χώρος Cameron – Martin H_T συνίσταται από όλα τα συνεχή απλά μονοπάτια $x \in \boldsymbol{C}\big([0,T],\mathbb{R}^{\boldsymbol{d}}\big)$ όπου $x(0)=0$, x είναι απόλυτα συνεχές, και η παράγωγος Radon – Nikodym του $\dot{x}$ είναι τετραγωνικά ολοκληρώσιμη.

<u>Λήμμα 1</u>: Οι χώροι Cameron – Martin είναι χώροι Hilbert, με νόρμα $\|x\|_{CM} = \int_0^T |\dot{x}(t)|^2 dt$.

<u>Ορισμός 5</u>: Η δράση της διαδικασίας Wiener μίας συνεχούς συνάρτησης $x \in \boldsymbol{C}([0,t],\mathbb{R}^d)$ είναι

$$J_T(x) \equiv \frac{1}{2}\|x\|_{CM}^2 \qquad \textbf{(3.3)}$$

εάν $x \in H_T$, και ∞ σε διαφορετική περίπτωση. Συγκεκριμένα,

$$J_1(x) \equiv \frac{1}{2}\int_0^1 |\dot{x}(t)|^2 dt \qquad \textbf{(3.4)}$$

Για κάθε $j>0$, έστω $L_T(j) = \{x : J_T(x) \le j\}$.

<u>Πρόταση 2</u>: Φτιάχνουμε μία συνάρτηση $f \in H_1$, και έστω $Y_\epsilon = X_\epsilon - f$. Τότε $\mathcal{L}(Y_\epsilon) = v_\epsilon$ είναι απολύτως συνεχής σε σχέση με $\mathcal{L}(X_\epsilon) = \mu_\epsilon$, και η παράγωγος Radon – Nikodym είναι

$$\frac{dv_\epsilon}{d\mu_\epsilon}(\epsilon w) = exp\left\{-\frac{1}{\epsilon}\int_0^1 \dot{w}(t)\cdot dW - \frac{1}{2\epsilon^2}\int_0^1 |\dot{w}(t)|^2 dt\right\} \qquad \textbf{(3.5)}$$

<u>Λήμμα 2</u>: Για οποιαδήποτε $\delta, \gamma, K > 0$, υπάρχει ένα $\epsilon_0 > 0$, εάν $\epsilon < \epsilon_0$,

$$\mathbb{P}(\|X_\epsilon - x\|_\infty \le \delta) \ge e^{-\frac{J_1(x)+\gamma}{\epsilon^2}} \qquad \textbf{(3.6)}$$

υπό τον όρο ότι $x(0)=0$ και $J_1(x) < K$.

<u>Απόδειξη</u>: Χρησιμοποιώντας τη Πρόταση 2,

$$\mathbb{P}(\|X_\epsilon - x\|_\infty \le \delta) = \mathbb{P}(\|Y_\epsilon - 0\|_\infty \le \delta) =$$

$$= \int_{\|\epsilon w\|_\infty < \delta} \frac{dv_\epsilon}{d\mu_\epsilon}(\epsilon w)\, d\mu_\epsilon(\epsilon w) = e^{-\frac{J_1(x)}{\epsilon^2}} \int_{\|\epsilon w\|_\infty < \delta} e^{-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW} d\mu_\epsilon(w)$$

Γνωρίζουμε ότι $\mathbb{P}(\|\epsilon W\|_\infty) < \delta) \to 1$ καθώς $\epsilon \to 0$. Άρα, εάν το ϵ είναι αρκετά μικρό, $\mathbb{P}(\|\epsilon W\|_\infty) < \delta) \geq 3/4$. Εφαρμόζοντας την ανισότητα Chebyshev στο ολοκλήρωμα,

$$\mathbb{P}\left(-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW \leq -\frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}\right)$$

$$\leq \quad \mathbb{P}\left(\left|\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW\right| \leq \frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}\right)$$

$$\leq \frac{\epsilon^2 \boldsymbol{E}\left[\left(\int_0^1 \dot{x}\cdot dW\right)^2\right]}{8\epsilon^2 J_1(x)} = \frac{\int_0^1 |\dot{x}|^2 dt}{8J_1(x)} = \frac{1}{4}$$

Χρησιμοποιώντας την ισομετρία Itô (Itô isometry). Γι’ αυτό το λόγο,

$$\mathbb{P}\left(e^{-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW} \geq e^{-\frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}}\right) \geq \frac{3}{4}$$

$$\int_{\|\epsilon w\|_\infty < \delta} e^{-\frac{1}{\epsilon}\int_0^1 \dot{x}\cdot dW} d\mu_\epsilon(w) > \frac{1}{2} e^{-\frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}}$$

$$\mathbb{P}(\|X_e - x\|_\infty \leq \delta) > \frac{1}{2} e^{-\frac{J_1(x)}{\epsilon^2} - \frac{2\sqrt{2}}{\epsilon}\sqrt{J_1(x)}}$$

Όπου ο δεύτερος όρος στο εκθετικό μπορεί να γίνει μικρότερος από οποιοδήποτε επιθυμητό γ παίρνοντας το ϵ αρκετά μικρό.

__Λήμμα 3__: Για κάθε $j > 0, \delta > 0$, έστω $U(j,\delta)$ είναι η ανοικτή δ γειτονιά του $L_1(j)$, για παράδειγμα, όλες οι τροχιές ερχόμενες μέσα στη δ μιας τροχιάς της οποίας η δράση είναι λιγότερη ή ίση της j. Τότε για κάθε $\gamma > 0$, υπάρχει ένα $\epsilon_0 > 0$ τέτοιο ώστε, εάν $\epsilon < \epsilon_0$ και

$$\mathbb{P}(X_\epsilon \notin U(j,\delta)) \leq e^{-\frac{s-\gamma}{\epsilon^2}} \qquad \textbf{(3.7)}$$

__Απόδειξη__: Βασικά, προσεγγίζοντας τη διαδικασία Wiener από μία συνεχής, τμηματικά γραμμική συνάρτηση, και δείχνοντας ότι η προσέγγιση είναι επαρκώς λεπτομερής. Διαλέγουμε έναν φυσικό αριθμό n, και έστω $Y_{n,\epsilon}(t)$ είναι η τμηματικά γραμμική τυχαία συνάρτηση η οποία συμπίπτει με X_ϵ στους χρόνους $0, 1/n, 2/n, \ldots 1$, για παράδειγμα,

$$Y_{n,\epsilon}(t) = X_\epsilon([tn]/n) + \left(t - \frac{[tn]}{n}\right) X_\epsilon([tn+1]/n)$$

Θα δούμε ότι, για αρκετά μεγάλο n, αυτή είναι εκθετικά κοντά στην X_ϵ. Πρώτα, φράσουμε τη πιθανότητα στην ανισότητα 3.7.

$$\mathbb{P}\big(X_\epsilon \notin U(j,\delta)\big) = \mathbb{P}\left(X_\epsilon \notin U(j,\delta), \left\|X_\epsilon - Y_{n,\epsilon}\right\|_\infty < \delta\right) + \mathbb{P}\left(X_\epsilon \notin U(j,\delta), \left\|X_\epsilon - Y_{n,\epsilon}\right\|_\infty \geq \delta\right)$$

$$\leq \quad \mathbb{P}\left(X_\epsilon \notin U(j,\delta), \left\|X_\epsilon - Y_{n,\epsilon}\right\|_\infty < \delta\right) + \mathbb{P}\left(\left\|X_\epsilon - Y_{n,\epsilon}\right\|_\infty \geq \delta\right)$$
$$\leq \quad \mathbb{P}\big(J_1(Y_{n,\epsilon}) > j\big) + \mathbb{P}\left(\left\|X_\epsilon - Y_{n,\epsilon}\right\|_\infty \geq \delta\right)$$

Η $J_1(Y_{n,\epsilon})$ μπορεί να παρθεί από τις προσαυξήσεις της διαδικασίας Wiener:

$$J_1\big(Y_{n,\epsilon}\big) = n\frac{\epsilon^2}{2}\sum_{i=1}^{n}\left|W\left(\frac{i}{n}\right) - W(\frac{i-1}{n})\right|^2 = \frac{\epsilon^2}{2}\sum_{i=1}^{dn}\xi_i$$

Όπου το ξ_i έχει την χ^2 κατανομή με έναν βαθμό ελευθερίας. Για επαρκώς μικρό ϵ,

$$\mathbb{P}\big(J_1\big(Y_{n,\epsilon}\big) > j\big) \leq \frac{1}{2}e^{-\frac{j-\gamma}{\epsilon^2}}$$

Για να εκτιμήσουμε τη πιθανότητα ότι η απόσταση ανάμεσα στα X_ϵ και $Y_{n,\epsilon}$ προσεγγίζει ή υπερβαίνει το δ, ξεκινάμε με την ιδιότητα ανεξάρτητης προσαύξησης της X_ϵ, και με το γεγονός ότι οι δύο διαδικασίες συμπίπτουν όταν $t = i/n$.

$$\mathbb{P}\left(\left\|X_\epsilon - Y_{n,\epsilon}\right\|_\infty \geq \delta\right)$$

$$\leq \sum_{i=1}^{n}\mathbb{P}\left(\max_{(i-1)/n\leq t\leq i/n}\left|X_\epsilon(t) - Y_{n,\epsilon}(t)\right| \geq \delta\right) = n\mathbb{P}\left(\max_{0\leq t\leq 1/n}\left|X_\epsilon(t) - Y_{n,\epsilon}(t)\right| \geq \delta\right)$$

$$= n\mathbb{P}\left(\max_{0\leq t\leq 1/n}|\epsilon W(t) - n\epsilon W(1/n)| \geq \delta\right) \leq n\mathbb{P}\left(\max_{0\leq t\leq 1/n}|\epsilon W(t)| \geq \delta/2\right)$$

$$\leq 4dn\mathbb{P}(W_1(1/n) \geq \frac{\delta}{2d\epsilon}) \leq 4dn\frac{2d\epsilon}{\delta\sqrt{2\pi n}}e^{-\frac{n\delta^2}{8d^2\epsilon^2}}$$

Εάν $n > \frac{4d^2 j}{\delta^2}$, τότε $\mathbb{P}\left(\left\|X_\epsilon - Y_{n,\epsilon}\right\|_\infty\right) \leq \frac{1}{2}e^{-\frac{j-\gamma}{\epsilon^2}}$, και έχουμε συνολικά

$$\mathbb{P}(X_\epsilon \notin U(j,\delta)) \leq e^{-\frac{j-\gamma}{\epsilon^2}}$$

όπως απαιτείται.

<u>**Θεώρημα 4**</u>**: (Θεώρημα Schilder - Schilder's Theorem)** Εάν W είναι μία d – διάστατη διαδικασία Wiener στο μοναδιαίο διάστημα, τότε η σχέση $X_\epsilon = \epsilon W$ υπακούει μία αρχή μεγάλης απόκλισης (LDP) στον $\boldsymbol{C}([0,1],\mathbb{R}^d)$, με ρυθμό ϵ^{-2} και καλή συνάρτηση ρυθμού $J_1(x)$, η αποτελεσματική Wiener δράση στο [0,1].

Απόδειξη: Είναι εύκολο να δείξουμε ότι το Λήμμα 2 συνεπάγεται το κάτω φράγμα μεγάλης απόκλισης για ανοικτά σύνολα. Το δύσκολο μέρος είναι το άνω φράγμα. Επιλέγουμε οποιοδήποτε κλειστό σύνολο C και οποιοδήποτε $\gamma > 0$. Έστω $s = J_1(C) - \gamma$. Από το Λήμμα 3 και τη Πρόταση 3, το σύνολο $K = L_1(s) = \{x\colon J_1(x) \leq s\}$ είναι συμπαγές. Από κατασκευής, $C \cap K = \emptyset$. Άρα $\delta = inf_{x\in C, y\in K}\|x - y\|_\infty > 0$. Έστω U είναι η κλειστή δ – γειτονιά του K. Τότε χρησιμοποιώντας το Λήμμα 2

$$\mathbb{P}(X_\epsilon \in C) \leq \mathbb{P}(X_\epsilon \notin U)$$

$$\leq e^{-\frac{s-\gamma}{\epsilon^2}}$$

$$\leq e^{-\frac{J_1(C)-2\gamma}{\epsilon^2}}$$

$$\log \mathbb{P}(X_\epsilon \in C) \leq -\frac{J_1(C) - 2\gamma}{\epsilon^2}$$

$$\epsilon^2 \log \mathbb{P}(X_\epsilon \in C) \leq \quad -J_1(C) - 2\gamma$$

$$\lim_{\epsilon\to 0} sup\, \epsilon^2 \log \mathbb{P}(X_\epsilon \in C) \leq \ -J_1(C) - 2\gamma$$

Αφού γ ήταν αυθαίρετο, αυτό ολοκληρώνει την απόδειξη.

Παρατήρηση: Το τέχνασμα το οποίο χρησιμοποιήθηκε εδώ, σχετικά με τη θέσπιση αποτελεσμάτων όπως το Λήμμα 2 και 3, και τότε χρησιμοποιώντας συμπαγή επίπεδα σύνολα για να αποδείξουμε τις μεγάλες αποκλίσεις, δουλεύει πιο γενικά.

Πόρισμα 1: Το θεώρημα Schilder παραμένει αληθές για διαδικασίες Wiener στο $[0, T]$, για όλα τα $T > 0$, με συνάρτηση ρυθμού J_T, η αποτελεσματική Wiener δράση στο $[0, T]$.

Απόδειξη: Εάν W είναι μία διαδικασία Wiener στο [0,1], τότε, για κάθε T, $S(W) = \sqrt{T}W(t/T)$ είναι μία διαδικασία Wiener στο $[0, T]$. Αφού η απεικόνιση S είναι συνεχής από $\boldsymbol{C}([0,1], \mathbb{R}^d)$ στο $\boldsymbol{C}([0, T], \mathbb{R}^d)$, από την αρχή συστολής στη θεωρία των Μεγάλων Αποκλίσεων, η οικογένεια $\epsilon S(W)$ υπακούει σε μία αρχή μεγάλης απόκλισης με ρυθμό ϵ^{-2} και καλή συνάρτηση ρυθμού $J_T(x) = J_1(S^{-1}(x))$. (Σημειώνουμε ότι S είναι αντιστρέψιμη, άρα $S^{-1}(x)$ είναι μία συνάρτηση, όχι ένα σύνολο συναρτήσεων.) Αφού $x \in H_1$ εάν $y = S(x) \in H_T$, είναι εύκολο να ελέγξουμε ότι για $\dot{y}(t) = T^{-\frac{1}{2}}\dot{x}\left(\frac{t}{T}\right)$, δηλώνοντας ότι

$$\|y\|_{CM} = \int_0^T |\dot{y}(t)|^2 dt = \int_0^T |\dot{x}(t/T)|^2 \frac{dt}{T} = \|x\|_{CM}$$

το οποίο ολοκληρώνει την απόδειξη.

<u>Πόρισμα 2</u>: Το θεώρημα Schilder παραμένει αληθές για διαδικασίες Wiener στον $\mathbb{R}^+$, με καλή συνάρτηση ρυθμού J_∞, η οποία δίνεται από την αποτελεσματική Wiener δράση στον $\mathbb{R}^+$,

$$J_\infty(x) \equiv \frac{1}{2}\int_0^\infty |\dot{x}(t)|^2 dt \quad \textbf{(3.8)}$$

εάν $x \in H_\infty$, $J_\infty(x) = \infty$ σε διαφορετική περίπτωση.

<u>Απόδειξη</u>: Για κάθε φυσικό αριθμό n, έστω $\pi_n x$ είναι περιορισμός του x στο διάστημα $[0, n]$. Από το Πόρισμα 1, κάθε ένας από αυτούς υπακούει μία αρχή μεγάλης απόκλισης με συνάρτηση ρυθμού $\frac{1}{2}\int_0^n |\dot{x}(t)|^2 dt$. Τώρα εφαρμόζουμε το προβολικό οριακό θεώρημα για να πάρουμε ότι $J_\infty(x) = sup_n J_n(x)$, η οποία είναι ξεκάθαρα η εξίσωση 3.8, καθώς η ολοκληρωτέα ποσότητα είναι μη αρνητική.

Παρατήρηση: Το προβολικό οριακό θεώρημα (Projective Limit) είναι το εξής

Έστω $\Xi_1, \Xi_2, \ldots$ είναι μία αριθμήσιμη ακολουθία μετρικών χώρων, και έστω X_ϵ είναι μία τυχαία ακολουθία από αυτόν το χώρο. Εάν, για κάθε n, $X_\epsilon^n = \pi_n X_\epsilon$ υπακούει την αρχή μεγάλης απόκλισης (LDP) με καλή συνάρτηση ρυθμού J_n, τότε X_ϵ υπακούει την αρχή μεγάλης απόκλισης με καλή συνάρτηση ρυθμού

$$J(x) \equiv sup_n J_n(\pi_n x) \quad \textbf{(3.9)}$$

3.3 Μεγάλες Αποκλίσεις για τις Στ.Δ.Ε. (SDEs) με θόρυβο ανεξάρτητος καταστάσεως (State-Independent Noise)

Έχοντας εισάγει μία αρχή μεγάλης απόκλισης για την διαδικασία Wiener, είναι αναμενόμενο να πάρουμε μία αρχή μεγάλης απόκλισης για στοχαστικές διαφορικές εξισώσεις όπου η οδήγηση του θορύβου είναι ανεξάρτητη από τη κατάσταση της διαδικασίας διάχυσης (diffusion process).

<u>Ορισμός 6</u>: (Στοχαστικές Διαφορικές Εξισώσεις με μικρό θόρυβο ανεξάρτητος καταστάσεως)

Μία Στοχαστική Διαφορική Εξίσωση με μικρό θόρυβο ανεξάρτητος καταστάσεως είναι μία στοχαστική διαφορική εξίσωση της μορφής

$$dX_\epsilon = a(X_\epsilon)dt + \epsilon dW \quad \textbf{(3.10)}$$

$$X_\epsilon(0) = 0 \quad \textbf{(3.11)}$$

όπου α: $\mathbb{R}^d \mapsto \mathbb{R}^d$ είναι ομοιόμορφα Lipschitz συνεχής.

Σημειώνεται ότι οποιαδήποτε μη – τυχαία αρχική συνθήκη x_0 μπορεί να αντιμετωπιστεί από μία απλή αλλαγή των συντεταγμένων.

<u>Ορισμός 7</u>: (Αποτελεσματική Δράση: Θόρυβος ανεξάρτητος καταστάσεως) Η αποτελεσματική δράση μίας τροχιάς $x \in H_\infty$ είναι

$$J(x) \equiv \frac{1}{2}\int_0^t |\dot{x}(t) - a(x(t))|^2 dt \tag{3.12}$$

και $= \infty$ εάν $x \in \boldsymbol{C} \setminus H_\infty$.

<u>Λήμμα 4</u>: Η απεικόνιση $F : \boldsymbol{C}(\mathbb{R}^+, \mathbb{R}^d) \mapsto \boldsymbol{C}(\mathbb{R}^+, \mathbb{R}^d)$ δίνεται από

$$x(t) = w(t) + \int_0^t a\big(w(s)\big)ds \tag{3.13}$$

όταν $x = F(w)$ είναι συνεχής.

<u>Απόδειξη</u>: Αυτή η απόδειξη πηγαίνει με τον ίδιο τρόπο όπως η απόδειξη της ύπαρξης και της μοναδικότητας των Στ.Δ.Ε. Για οποιοδήποτε w_1, $w_2 \in \boldsymbol{C}(\mathbb{R}^+, \mathbb{R}^d)$, ορίζεται $x_1 = F(w_1), x_2 = F(w_2)$. Από την ιδιότητα Lipschitz του α,

$$|x_1(t) - x_2(t)| \leq \|w_1 - w_2\| + K_a \int_0^t |x_1(s) - x_2(s)|ds$$

(γράφοντας $|y(t)|$ για την Ευκλείδεια νόρμα των διανυσμάτων y, και $\|x\|$ για την supremum νόρμα των συνεχών καμπυλών). Από την ανισότητα Gronwall, τότε,

$$\|x_1 - x_2\| \leq \|w_1 - w_2\| e^{K_a T}$$

σε κάθε διάστημα $[0, T]$. Έτσι μπορούμε να βεβαιωθούμε ότι $\|x_1 - x_2\|$ είναι μικρότερη από οποιαδήποτε επιθυμητή ποσότητα, βεβαιώνοντας ότι $\|w_1 - w_2\|$ είναι επαρκώς μικρή, άρα F είναι συνεχής.

<u>Λήμμα 5</u>: Εάν $w \in H_\infty$, τότε $x = F(w)$ είναι στον H_∞.

<u>Θεώρημα 5</u>: (Θεώρημα Freidlin – Wentzell: Θόρυβος ανεξάρτητος καταστάσεως) Οι διαδικασίες Itô X_ϵ του ορισμού 6 υπακούουν στην αρχή μεγάλης απόκλισης με ρυθμό ϵ^{-2} και καλή συνάρτηση ρυθμού, η οποία δίνεται από την αποτελεσματική δράση $J(x)$

<u>Απόδειξη</u>: Για κάθε ϵ, $X_\epsilon = F(\epsilon W)$. Το πόρισμα 2 μας λέει ότι το ϵW υπακούει την αρχή μεγάλης απόκλισης με ρυθμό ϵ^{-2} και καλή συνάρτηση ρυθμού J_∞. Αφού F είναι συνεχής, από την αρχή συστολής, η X_ϵ υπακούει την αρχή μεγάλης απόκλισης, με ρυθμό, ο οποίος δίνεται από $J_\infty\big(F^{-1}(x)\big)$. Εάν $F^{-1}(x) \cap H_\infty = \emptyset$, αυτό είναι ∞. Απ' την άλλη μεριά, εάν $F^{-1}(x)$ δεν περιέχει καμπύλες στον H_∞, τότε $J_\infty\big(F^{-1}(x)\big) = J_\infty(F^{-1}(x) \cap H_\infty)$. Από το λήμμα 5, αυτό συνεπάγεται ότι $x \in H_\infty$. Για οποιαδήποτε καμπύλη $w \in F^{-1}(x) \cap H_\infty$, $\dot{x} = \dot{w} + a(x)$, ή $\dot{w} = \dot{x} - a(x)$. $J_\infty(w) = \int_0^\infty |\dot{x} - a(x)|^2 dt$ είναι η αποτελεσματική δράση της τροχιάς x (ορισμός 7).

3.4 Μεγάλες Αποκλίσεις για τις Στ.Δ.Ε. (SDEs) με θόρυβο εξαρτημένος καταστάσεως (State-Dependent Noise)

Εάν ο όρος διάχυσης στη στοχαστική διαφορική εξίσωση δεν εξαρτάται από τη κατάσταση της διαδικασίας, έχουμε μία πανομοιότυπη αρχή μεγάλης απόκλισης σε σχέση με τα αποτελέσματα της προηγούμενης παραγράφου. Παρ' όλα αυτά, η προσέγγιση πρέπει να τροποποιηθεί: Η απεικόνιση από τη W στην X_ϵ, ενώ εξακολουθούσε να είναι μετρήσιμη, δεν είναι πλέον απαραίτητα συνεχής, γι' αυτό το λόγο δεν μπορούμε να χρησιμοποιήσουμε την αρχή της συστολής όπως πριν.

Ορισμός 8: (Στ.Δ.Ε. με μικρό θόρυβο εξαρτημένος καταστάσεως) Μία Στ.Δ.Ε. με μικρό θόρυβο εξαρτημένος καταστάσεως είναι μία στοχαστική διαφορική εξίσωση της μορφής

$$dX_\epsilon = a(X_\epsilon)dt + \epsilon b(X_\epsilon)dW \quad \textbf{(3.14)}$$

$$X_\epsilon(0) = 0 \quad \textbf{(3.15)}$$

όπου α και b είναι ομοιόμορφα Lipschitz συνεχείς, και b δεν είναι ιδιόμορφο.

Ορισμός 9: (Αποτελεσματική Δράση: Θόρυβος εξαρτημένος καταστάσεως) Η αποτελεσματική δράση μιας τροχιάς $x \in H_\infty$ δίνεται από

$$J(x) \equiv \int_0^\infty L\big(x(t), \dot{x}(t)\big)dt \quad \textbf{(3.16)}$$

όπου

$$L(q,p) = \frac{1}{2}\big(p_i - a_i(q)\big)B_{ij}^{-1}(q)(p_j - a_j(q)) \quad \textbf{(3.17)}$$

και

$$B(q) = b(q)b^T(q) \quad \textbf{(3.18)}$$

με $J(x) = \infty$ εάν $x \in \boldsymbol{C} \setminus H_\infty$.

Θεώρημα 6: (Θεώρημα Freidlin – Wentzell: Θόρυβος εξαρτημένος καταστάσεως) Οι διαδικασίες X_ϵ υπακούουν μία αρχή μεγάλων αποκλίσεων με ρυθμό ϵ^{-2} και καλή συνάρτηση ρυθμού ίση με την αποτελεσματική δράση.

<u>Απόδειξη</u>: Είναι αξιοσημείωτα περίπλοκη απόδειξη σε σχέση με τη προηγούμενη (απόδειξη θεωρήματος 4). Παρ' όλα αυτά, μία αίσθηση αυτής της απόδειξης είναι ότι θεωρούμε μία προσεγγιστική διαδικασία Itô X_n, όπου $\alpha(X_t)$ και $b(X_t)$ αντικαθίστανται στην εξίσωση 3.14 από τα $\alpha(X_n([tn]/n))$ και $b(X_n([tn]/n))$ αντίστοιχα. Εδώ η απεικόνιση από τη W στη X_n είναι συνεχής, άρα δεν είναι τόσο δύσκολο να δείξουμε ότι οι τελευταίες υπακούουν μία αρχή μεγάλων αποκλίσεων με μία λογική συνάρτηση ρυθμού, και άρα είναι εκθετικά ισοδύναμες (στο n) στο X_ϵ.

ΠΑΡΑΡΤΗΜΑ

σ – άλγεβρες: Για να ορίσουμε μαθηματικά την έννοια του γεγονότος, ή της συλλογής γεγονότων, θα πρέπει να εισάγουμε την έννοια της *σ – άλγεβρας*.

Ορισμός 1: Έστω Ω κάποιο σύνολο. Μία σ – άλγεβρα $\mathcal{F}$ επάνω στο σύνολο Ω είναι μία οικογένεια υποσυνόλων του Ω με τις παρακάτω ιδιότητες:

1. $\emptyset \in \mathcal{F}$
2. $F \in \mathcal{F} \Longrightarrow F^c \equiv \Omega \setminus F \in \mathcal{F}$
3. $A_1, A_2, \dots \in \mathcal{F} \Longrightarrow A \coloneqq \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Οι σ – άλγεβρες συνήθως χρησιμοποιούνται για να περιγράψουν δομές πληροφορίας. Τα στοιχεία του συνόλου Ω μπορούν να θεωρηθούν σαν οι πιθανές καταστάσεις ενός συστήματος ή τα πιθανά αποτελέσματα ενός πειράματος. Στην θεωρία πιθανοτήτων πολλές φορές τα υποσύνολα του Ω ονομάζονται γεγονότα.

Η ελάχιστη σ – άλγεβρα που ορίζεται από ένα σύνολο Α, είναι η μικρότερη σ – άλγεβρα που περιέχει το σύνολο Α. Η άλγεβρα αυτή συνήθως συμβολίζεται σ(Α).

Άλγεβρα Borel: Ορίζουμε μία συγκεκριμένη σ – άλγεβρα η οποία είναι ιδιαίτερα χρήσιμη στην μελέτη της θεωρία πιθανοτήτων και της στοχαστικής ανάλυσης. Θα θεωρήσουμε $\Omega = \mathbb{R}$ και θα κατασκευάσουμε μία σ – άλγεβρα που αποτελείται από υποσύνολα του $\mathbb{R}$.

Ορισμός 2: Η άλγεβρα Borel, την οποία συμβολίζουμε σαν $\mathcal{B}$, είναι η ελάχιστη σ – άλγεβρα που περιέχει την κλάση C όλων των διαστημάτων της μορφής $(-\infty, x)$, τα οποία μπορεί να θεωρηθούν ως υποσύνολα της πραγματικής ευθείας. Τα στοιχεία του $\mathcal{B}$ ονομάζονται σύνολα Borel.

Η σ – άλγεβρα Borel $\mathcal{B}(\mathbb{R}^d)$ είναι η ελάχιστη σ – άλγεβρα η οποία περιέχει όλα τα παραλληλόγραμμα της μορφής $(a, b]$ ή με άλλα λόγια την σ – άλγεβρα που παράγεται από τα παραλληλόγραμμα.

Μετρήσιμος χώρος: Ορίζουμε την έννοια του μετρήσιμου χώρου.

Ορισμός 3: Έστω ένα σύνολο Ω και μία σ – άλγεβρα $\mathcal{F}$ που αποτελείται από τα υποσύνολα του Ω. Το ζεύγος $(\Omega, \mathcal{F})$ καλείται **μετρήσιμος χώρος**.

Μέτρο πιθανότητας: Το μέτρο πιθανότητας έχει να κάνει με το ερώτημα του πόσο εύκολα μπορεί να συμβεί κάποιο γεγονός. Το εύκολα θα σχετιστεί με έναν αριθμό ο οποίος θα αντιστοιχεί με κάποιο γεγονός που όπως είδαμε μέχρι τώρα στην μαθηματική γλώσσα μεταφράζεται σε κάποιο σύνολο που ανήκει σε μία κατάλληλα επιλεγμένη σ – άλγεβρα. Γι' αυτό εισάγουμε την έννοια συναρτήσεων που παίρνουν ένα σύνολο και το αντιστοιχούν σε κάποιο αριθμό. Τέτοιες συναρτήσεις ονομάζονται **μέτρα**. Στην θεωρία πιθανοτήτων αυτά το μέτρο ονομάζεται **μέτρο πιθανότητας**.

Ορισμό 4: Ένα μέτρο πιθανότητας P επάνω σε ένα μετρήσιμο χώρο $(\Omega, \mathcal{F})$ είναι μία απεικόνιση $P: \mathcal{F} \to [0,1]$ με τις ιδιότητες

1. $P(\emptyset) = 0, P(\Omega) = 1$
2. Αν $A_1, A_2, \ldots \in \mathcal{F}$ και τα $\{A_i\}$ είναι ανά δύο ξένα, τότε

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Από τον παραπάνω ορισμό είναι σαφές ότι ένα μέτρο είναι μία απεικόνιση που απεικονίζει ένα σύνολο σε ένα πραγματικό αριθμό. Κατά κάποιο τρόπο ένα μέτρο εκφράζει το μέγεθος ενός συνόλου.

Χώρος πιθανοτήτων: Ορίζουμε την έννοια του χώρου πιθανοτήτων.

Ορισμός 5: Ας θεωρήσουμε ένα σύνολο Ω, μία σ – άλγεβρα $\mathcal{F}$ επάνω σε αυτό και ένα μέτρο πιθανότητας P. Η τριάδα $(\Omega, \mathcal{F}, P)$ ονομάζεται **χώρος πιθανοτήτων**.

Μετρήσιμο σύνολο: Μία πολύ σημαντική έννοια είναι η έννοια του μετρήσιμου συνόλου. Η έννοια αυτή ορίζεται πάντα ως προς μία συγκεκριμένη σ – άλγεβρα.

Ορισμός 6: Τα υποσύνολα F του Ω που ανήκουν στη σ – άλγεβρα $\mathcal{F}$ αποκαλούνται $\mathcal{F}$ - μετρήσιμα.

***$\mathcal{F}$*-μετρήσιμη συνάρτηση:** Εισάγουμε την έννοια της μετρήσιμης συνάρτησης.

Ορισμός 7: Έστω ***Ω*** κάποιο σύνολο. Η συνάρτηση $Y : \Omega \to \mathbb{R}^d$ αποκαλείται $\mathcal{F}$ - μετρήσιμη αν

$$Y^{-1}(U) = \{\omega \in \Omega: Y(\omega) \in U\} \in \mathcal{F}$$

για κάθε ανοικτό σύνολο $U \in \mathbb{R}^d$.

Με άλλα λόγια λέμε ότι η συνάρτηση Y είναι $\mathcal{F}$-μετρήσιμη αν η αντίστροφη εικόνα ενός υποσυνόλου του $\mathbb{R}^d$, κάτω από την συνάρτηση αυτή, ανήκει στην σ – άλγεβρα $\mathcal{F}$.

Τυχαία μεταβλητή:

Ορισμό 8: Μία (πραγματική) **τυχαία μεταβλητή** είναι μία $\mathcal{F}$-μετρήσιμη συνάρτηση $X : \Omega \to \mathbb{R}^d$ όπου $(\Omega, \mathcal{F}, P)$ είναι ένας χώρος πιθανοτήτων.

Μπορούμε να θεωρήσουμε την τυχαία μεταβλητή X σαν μία μεταβλητή που η τιμή της εξαρτάται από την έκβαση ενός τυχαίου πειράματος. Ο ορισμός αυτός μας λέει ότι για κάθε κατάσταση παίρνουμε ένα πραγματικό διάνυσμα $X \in \mathbb{R}^d$ (για d=1 παίρνουμε έναν αριθμό).

Κατανομή μίας τυχαίας μεταβλητής: Κάθε τυχαία μεταβλητή X επάγει ένα μέτρο μ_X στο $\mathbb{R}^d$, το οποίο ορίζεται σύμφωνα με την σχέση

$$\mu_X(B) = P(X^{-1}(B))$$

για κάποιο σύνολο Borel B στο $\mathcal{B}(\mathbb{R}^d)$. Το μέτρο αυτό αποκαλείται η κατανομή της X.

Η σ – άλγεβρα που παράγεται από μία τυχαία μεταβλητή: Ας θεωρήσουμε ένα χώρο πιθανοτήτων $(\Omega, \mathcal{F}, P)$ και μία τυχαία μεταβλητή, έστω X. Από τις παραπάνω πληροφορίες βλέπουμε ότι η X είναι εξ’ ορισμού $\mathcal{F}$-μετρήσιμη. Όμως, θα υπάρχουν μικρότερες σ – άλγεβρες για τις οποίες η X θα είναι επίσης μετρήσιμη.

Ορισμός 9: Η μικρότερη σ – άλγεβρα για την οποία η τυχαία μεταβλητή X είναι μετρήσιμη, αποκαλείται **η σ – άλγεβρα που παράγεται από την τυχαία μεταβλητή** X και συμβολίζεται σ(X).

Με άλλα λόγια η σ(X) είναι η μικρότερη σ-άλγεβρα που περιέχει την πληροφορία σχετικά με τις τιμές που μπορεί να πάρει η X.

Ιδιαίτερα χρήσιμο για τον χαρακτηρισμό της σ-άλγεβρας ως προς την οποία είναι μετρήσιμη μία τυχαία μεταβλητή είναι το ακόλουθο λήμμα, γνωστό και ως λήμμα Doob – Dynkin:

<u>Θεώρημα 1</u>: Έστω $X, Y : \Omega \rightarrow \mathbb{R}^d$, δύο τυχαίες μεταβλητές. Η Y είναι μετρήσιμη ως προς την σ-άλγεβρα η οποία παράγεται από την τυχαία μεταβλητή X, $\sigma(X)$, αν και μόνο αν υπάρχει συνάρτηση g, μετρήσιμη κατά Borel, $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ τέτοια ώστε $Y = g(X)$.

<u>Στοχαστικές διαδικασίες</u>: Ορίζουμε την έννοια της στοχαστικής διαδικασίας.

<u>Ορισμός 10</u>: Μία στοχαστική διαδικασία είναι μία παραμετρισμένη συλλογή τυχαίων μεταβλητών $\{X_t\}_{t\in T}$ οι οποίες ορίζονται σε ένα χώρο πιθανοτήτων $(\Omega, \mathcal{F}, P)$ και παίρνουν τιμές στο $\mathbb{R}^d$.

Μία στοχαστική διαδικασία έχει δύο 'μεταβλητές', την t και την ω.

- Για κάθε $t \in T$ (το οποίο θεωρούμε δεδομένο και σταθερό) έχουμε μία τυχαία μεταβλητή
$$\omega \rightarrow X_t(\omega); \omega \in \Omega$$

- Θεωρώντας σταθερό $\omega \in \Omega$ θεωρούμε την συνάρτηση
$$t \rightarrow X_t(\omega); t \in T$$

  η οποία ονομάζεται **τροχιά** (path) της X_t.

Μία στοχαστική διαδικασία μπορεί να θεωρηθεί σαν ένα μέτρο πιθανότητας P σε έναν κατάλληλα ορισμένο χώρο μέτρου δηλαδή

$$\boxed{\text{Στοχαστική διαδικασία} \Leftrightarrow \text{Μέτρο Πιθανότητας}}$$

<u>Ανεξάρτητα γεγονότα</u>: Από βασική θεωρία πιθανοτήτων έχουμε

<u>Ορισμός 11</u>: Δύο γεγονότα A και B λέγονται ανεξάρτητα αν $P(A \cap B) = P(A)P(B)$.

Με άλλα λόγια δύο γεγονότα είναι ανεξάρτητα αν το ένα δεν επηρεάζει το άλλο.

<u>Ανεξάρτητες σ – άλγεβρες</u>: Θεωρούμε μία σ – άλγεβρα $\mathcal{F}$ επάνω σε ένα σύνολο Ω.

<u>Ορισμός 12</u>: Οι σ – υποάλγεβρες $\mathcal{F}_i$, $i \in I$ της $\mathcal{F}$ ονομάζονται **ανεξάρτητες** αν για κάθε υποσύνολο J του I και κάθε σύνολο $A_i \in \mathcal{F}_i$ έχουμε

$$P\left(\bigcap_{n\in J} A_n\right) = \prod_{n\in J} P(A_n)$$

<u>Ανεξάρτητες τυχαίες μεταβλητές</u>: Με βάση τον ορισμό των ανεξάρτητων αλγεβρών μπορούμε να ορίσουμε την έννοια των ανεξαρτήτων τυχαίων μεταβλητών.

<u>Ορισμός 13</u>: Οι τυχαίες μεταβλητές X_i, $i \in I$ ονομάζονται **ανεξάρτητες τυχαίες μεταβλητές** αν οι σ – άλγεβρες που παράγονται από τις τυχαίες μεταβλητές είναι ανεξάρτητες.

<u>Το γινόμετρο μέτρο</u>: Είναι πολύ σημαντικό να αναφέρουμε τη κατανομή που ακολουθούν ανεξάρτητες τυχαίες μεταβλητές. Συγκεκριμένα ας θεωρήσουμε τυχαίες μεταβλητές $X_i \in \mathbb{R}, i = 1, \dots, d$, οι οποίες είναι ανεξάρτητες μεταξύ τους και που η κάθε μία από αυτές επάγει το μέτρο πιθανότητας μ_{X_i}. Κατασκευάζουμε την τυχαία μεταβλητή $X = (X_1, \dots, X_d) \in \mathbb{R}^d$. Για να ορίσουμε το μέτρο πιθανότητας που επάγει στον $\mathbb{R}^d$ η τυχαία μεταβλητή X, πρέπει να ορίσουμε την έννοια του γινόμενου μέτρου (**product measure**).

<u>Ορισμός 14</u>: Αν θεωρήσουμε τους χώρους μέτρου $(\Omega_i, \mathcal{F}_i, \mu_i), i = 1, \dots, d$, μπορούμε να ορίσουμε το **γινόμενο μέτρο** $\mu_1 \otimes \dots \otimes \mu_d$ σύμφωνα με τον τύπο

$$(\mu_1 \otimes \dots \otimes \mu_d)(A_1 \times \dots \times A_d) = \mu_1(A_1) \dots \mu_d(A_d)$$

όπου $A_i \in \mathcal{F}_i$ και με $A_1 \times \dots \times A_d$ συμβολίζουμε το καρτεσιανό γινόμενο των συνόλων A_i. Το γινόμενο μέτρο ορίζεται επάνω στη **γινόμενη σ – άλγεβρα** $\boldsymbol{\otimes_i \mathcal{F}_i}$ η οποία είναι η μικρότερη σ – άλγεβρα που παράγεται από όλα τα σύνολα της μορφής $A_1 \times \dots \times A_k, k = 1, \dots d$ όπου με το γινόμενο εννοούμε εδώ το καρτεσιανό γινόμενο.

Το γινόμενο μέτρο είναι το μέτρο που αντιστοιχεί σε συλλογές ανεξαρτήτων τυχαίων μεταβλητών. Γι' αυτό το λόγο έχουμε το παρακάτω θεώρημα.

<u>Θεώρημα 2</u>: Έστω $X_i \in \mathbb{R}^d$ τυχαίες μεταβλητές που επάγουν τα μέτρα πιθανότητας $\mu_{X_i}, i = 1, \dots, d$. Οι X_i είναι ανεξάρτητες αν και μόνο αν η τυχαία μεταβλητή $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ επάγει το μέτρο (δηλαδή έχει κατανομή) $\mu_X = \mu_{X_1} \otimes \dots \otimes \mu_{X_d}$.

<u>Μέση τιμή</u>: Δίνουμε την έννοια της μέσης τιμής, τον ορισμό της και τις βασικές της ιδιότητες.

<u>Ορισμός 15</u>: Η ***μέση τιμή*** μιας τυχαίας μεταβλητής X, ή οποιασδήποτε συνάρτησης της ορίζεται ως

$$E[X] \coloneqq \int_{\Omega} X(\omega) dP(\omega) \coloneqq \int_{\mathbb{R}^d} x d\mu_X(x) \coloneqq \int_{\mathbb{R}^d} x \mu_X d(x)$$

$$E[g(x)] \coloneqq \int_{\Omega} g(X(\omega)) dP(\omega) \coloneqq \int_{\mathbb{R}^d} g(x) d\mu_X(x) \coloneqq \int_{\mathbb{R}^d} g(x) \mu_X d(x)$$

Ορισμένες φορές μας ενδιαφέρει η μέση τιμή μίας τυχαίας μεταβλητής επάνω σε ένα υποσύνολο A του συνόλου των γεγονότων. Αυτό το συμβολίζουμε ως $E(X; A) \coloneqq \int_A X dP(\omega)$.

Για τον μαθηματικό ορισμό της μέσης τιμής για μία οποιαδήποτε τυχαία μεταβλητή θα πρέπει να ακολουθήσουμε τα παρακάτω βήματα:

1. Αν η X είναι μία απλή τυχαία μεταβλητή στον χώρο πιθανοτήτων $(\Omega, \mathcal{F}, P)$, δηλαδή μία μεταβλητή της μορφής

$$X = \sum_{i=0}^{n} c_i \mathbf{1}_{A_i},\ A_i \cap A_j = \emptyset, \bigcup_{i=1}^{n} A_i = \Omega,$$

   τότε η μέση τιμή της X μπορεί να οριστεί σαν

$$E[X] \coloneqq \sum_{i=1}^{n} c_i P(A_i)$$

   Για το παραπάνω άθροισμα χρησιμοποιείται ο συμβολισμός

$$E[X] \coloneqq \int \Omega X(\omega) P(d\omega) \coloneqq \int_{\Omega} X dP$$

2. Για να ορίσουμε την μέση τιμή για γενικότερες τυχαίες μεταβλητές θα πρέπει να βασιστούμε στο γεγονός ότι κάθε θετική τυχαία μεταβλητή μπορεί να προσεγγιστεί από μία ακολουθία απλών μεταβλητών X_n, η οποία είναι μονότονη και συγκεκριμένα αύξουσα που τείνει προς το X καθώς $n \to \infty$, $(X_n \uparrow X)$.
3. Μπορούμε να πάρουμε την μέση τιμή της ακολουθίας μεταβλητών X_n που είναι απλές (οπότε η μέση τιμή δίνεται από τον παραπάνω ορισμό για απλές μεταβλητές). Αυτό οδηγεί στην ακολουθία πραγματικών αριθμών $E[X_n]$. Μπορούμε να δείξουμε πως $X_n(\omega) \to X(\omega)$ και $E[X_n] \to E[X]$ καθώς $n \to \infty$.
4. Μπορούμε να ορίσουμε τη μέση τιμή για μία οποιαδήποτε θετική τυχαία μεταβλητή σαν

$$E[X] = sup_Z E[Z]$$

   όπου παίρνουμε το sup επάνω στο σύνολο όλων των απλών τυχαίων μεταβλητών Z για τις οποίες ισχύει $Z \leq X$.

<u>Θεώρημα 3</u>: Η μέση τιμή έχει τις παρακάτω σημαντικές ιδιότητες

1. Η μέση τιμή είναι γραμμικός τελεστής δηλαδή αν X και Y είναι τυχαίες μεταβλητές και c_1 και c_2 σταθερές τότε
$$E[c_1X + c_2Y] = c_1E[X] + c_2E[Y]$$
2. Αν $X \leq Y$, σ.β. τότε $E[X] \leq E[Y]$.
3. Αν $X = c$, σ.β. όπου c σταθερά, τότε $E[X] = c$.
4. Αν $X \geq 0$ τότε $E[X] \coloneqq \int XdP = 0$ αν και μόνο αν $X = 0$ σχεδόν παντού (ως προς το μέτρο P).

<u>Θεώρημα 4 (Ανισότητα Chebychev)</u>: Έστω X τυχαία μεταβλητή. Ισχύει η ακόλουθη ανισότητα

$$P(|X(\omega)| \geq a) \leq \frac{E[X^2]}{a^2}$$

<u>Θεώρημα 5 (Ανισότητα Hölder)</u>: Έστω X, Y τυχαίες μεταβλητές. Ισχύει ότι

$$\{E[|XY|^r]\}^{1/r} \leq \{E[|X|^p]\}^{1/p}\{E[|Y|^q]\}^{1/q}$$

όπου $p, q, r > 0$ και $p^{-1} + q^{-1} = r^{-1}$, αρκεί φυσικά οι παραπάνω ποσότητες να ορίζονται. Στην ειδική περίπτωση όπου $r = 1, p = q = 2$ παίρνουμε την ανισότητα Cauchy – Schwartz.

<u>Η κίνηση Brown</u>: Μία από τις πιο σημαντικές στοχαστικές διαδικασίες είναι η κίνηση Brown. Η κίνηση Brown (η οποία ονομάζεται και **διαδικασία Wiener**). Η στοχαστική αυτή διαδικασία παίζει πολύ σημαντικό ρόλο στην θεωρία των στοχαστικών διαφορικών εξισώσεων.

<u>Ορισμός 16</u>: Η κίνηση Brown είναι μία στοχαστική διαδικασία B_t η οποία παίρνει τιμές στον $\mathbb{R}$ και έχει τις ακόλουθες ιδιότητες

- Αν $t_0 < t_1 < \cdots < t_n$ τότε οι τυχαίες μεταβλητές $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ είναι **ανεξάρτητες** (ανεξάρτητες μεταβολές)
- Αν $s, t \geq 0$, τότε
$$P(B_{s+t} - B_s \in A) = \int_A \frac{1}{(2\pi t)^{1/2}} \exp\left(-\frac{|x|^2}{2t}\right),$$
όπου A κάποιο σύνολο Borel, δηλαδή οι μεταβολές της κίνησης Brown είναι κατανεμημένες με την **κανονική κατανομή** (κατανομή Gauss).
- Οι τροχιές της κίνησης Brown είναι συνεχείς με πιθανότητα 1, δηλαδή η $t \to B_t$ είναι συνεχής συνάρτηση.

Οι τρείς αυτές ιδιότητες ορίζουν μία και μοναδική στοχαστική διαδικασία. Από τις ιδιότητες της κίνησης Brown μπορούμε να συνάγουμε τις ιδιότητες του μέτρου μ που αυτή επάγει (μέτρο Wiener)

$$\mu_{t_1,t_2,\ldots t_n}(A_1 \times A_2 \times \ldots \times A_n) = \int_{A_1} dx_1 \int_{A_2} dx_2 \ldots \int_{A_n} dx_n \prod_{i=1}^{n} p(t_i - t_{i-1}, x_{i-1}, x_i)$$

όπου $x_0 = x, t_1 = 0$, και

$$p(t,x,y) = \frac{1}{(2\pi t)^{1/2}} \exp\left(-\frac{|y-x|^2}{2t}\right)$$

Η παραπάνω ποσότητα είναι κατά κάποιο τρόπο η πιθανότητα να βρίσκεται η στοχαστική διαδικασία τις χρονικές στιγμές t_i στα υποσύνολα $A_i \in \mathcal{B}(\mathbb{R})$. Τα υποσύνολα αυτά μπορούν να θεωρηθούν διαστήματα του $\mathbb{R}$ οπότε η παραπάνω ποσότητα είναι ουσιαστικά η πιθανότητα να βρίσκεται η κίνηση Brown τις χρονικές στιγμές t_i σε συγκεκριμένα διαστήματα του $\mathbb{R}$.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 2^{nd} edition, 1998.
2. M. Capinski and E. Kopp. *Measure, Integral and Probability*. Springer Undergraduate Mathematics Series. Springer, New York, 2005.
3. P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley Series in Probability and Statistics. John Wiley, New York, 1997.
4. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley, New York, 1991.
5. R. S. Ellis. *Entropy*, *Large Deviations, and Statistical Mechanics*. Springer, New York, 1985.
6. J. Gärtner. On large deviations from the invariant measure. *Th. Prob. Appl.*, 22:24 – 39, 1977.
7. R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
8. S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261 – 286, 1966.
9. J. van Tiel. *Convex Analysis*: *An Introductory Text*. John Wiley, New York, 1984.
10. N. O'Connell. From laws of large numbers to large deviation principles. *Markov Proc. Rel. Fields*, 3(4), 1997.
11. W. Bryc. A remark on the connection between the large deviation principle and the central limit theorem. *Stat. Prob. Lett.*, 18(4):253 – 256, 1993.
12. M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.*, 36(2):183 – 212, 1983.
13. O. E. Lanford. Entropy and equilibrium states in classical statistical mechanics. In A. Lenard, editor, *Statistical Mechanics and Mathematical Problems*, volume 20 of *Lecture Notes in Physics*, pages 1 – 113, Berlin, 1973. Springer.
14. P. Ney. Dominating points and the asymptotics of large deviations for random walk on $\mathbb{R}^d$. *Ann. Prob.*, 11(1):158 – 167, 1983.
15. J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation and Estimation.* Wiley Series in Probability and Mathematical Statistics. Wiley Interscience, New York, 1990.
16. S. Asmussen. *Ruin Probabilities*, World Scientific, 2000.
17. P. Bares, R. Cont, L. Gardiol, R. Gibson and S. Gyger. A large deviation approach to portfolio management, *International Journal of Theoretical and Applied Finance,* 3, 617 – 639, 2000.
18. J. David Logan. *Εφαρμοσμένα Μαθηματικά*, Πανεπιστημιακές Εκδόσεις Κρήτης, 1999.
19. J. Yeh. *Real Analysis, Theory of Measure and Integration*, World Scientific, 2006.

20. D. Ruelle. *Thermodynamic Formalism*. Cambridge University Press, Cambridge, 2^{nd} edition, 2004.
21. H. Spohn. *Large Scale Dynamics of Interacting Particles*. Springer Verlag, Heidelberg, 1991.
22. R.J. Harris and G.M. Schütz. Fluctuation theorems for stochastic dynamics. *J. Stat. Mech.*, 2007(07):P07020, 2007.
23. R. Chetrite and K. Gawedzki. Fluctuation relations for diffusion processes. *Comm. Math. Phys.*, 282(2):469 – 518, 2008.
24. S. Asmussen and P.W. Glynn. *Stochastic Simulation*: *Algorithms and Analysis*. Stochastic Modelling and Applied Probability. Springer, New York, 2007.
25. J.A. Bucklew. *Introduction to Rare Event Simulation*. Springer, New York, 2004.
26. M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*, Springer, Second Edition, 1998.
27. Α. Ν. Γιαννακόπουλος. *Στοχαστική Ανάλυση και Εφαρμογές στη Χρηματοοικονομική*, 2003.
28. Α. Ν. Γιαννακόπουλος. *Αριθμητικές Μέθοδοι στα Χρηματοοικονομικά ΙΙΙ, Μέθοδοι προσομοίωσης (μέρος Α)*, 2008.

<u>ΕΥΧΑΡΙΣΤΙΕΣ</u>

Η παρούσα μεταπτυχιακή εργασία πραγματοποιήθηκε στο Τμήμα Μαθηματικών της Σχολής Θετικών Επιστημών του Ε.Κ.Π.Α. στα πλαίσια του προγράμματος μεταπτυχιακών σπουδών «Εφαρμοσμένα Μαθηματικά». Την τριμελή επιτροπή αποτέλεσαν ο Δρ. Ι. Στρατής Καθηγητής Ε.Κ.Π.Α. (επιβλέπων), ο Δρ. Α. Γιαννακόπουλος Καθηγητής Ο.Π.Α. και ο Δρ. Α. Μπουρνέτας Καθηγητής Ε.Κ.Π.Α., τους οποίους και ευχαριστώ για τις χρήσιμες συμβουλές τους.

Ευχαριστώ ιδιαίτερα τον Δρ. Α. Γιαννακόπουλο για την επιστημονική συνεργασία μας και τη χρήσιμη καθοδήγησή του κατά τη διάρκεια της διαδικασίας εκπόνησης της συγκεκριμένης εργασίας, καθώς επίσης και για τις πολύτιμες παρατηρήσεις του και τις πολύ ενδιαφέρουσες συζητήσεις μας πάνω σε ζητήματα στο τομέα των εφαρμοσμένων μαθηματικών, των πιθανοτήτων, της στοχαστικής ανάλυσης και άλλων επιστημονικών και κοινωνικών θεμάτων. Η παρουσία του υπήρξε καταλυτική για μένα ώστε να ολοκληρώσω αυτόν τον κύκλο σπουδών και ελπίζω σε μια μελλοντική συνεργασία.

Τέλος ευχαριστώ τον Δρ. Ι. Στρατή για την τιμή, την οποία μου έκανε να είναι επιβλέπων στην εκπόνηση της συγκεκριμένης εργασίας, αλλά και για το σύνολο των γνώσεων που μου μετέδωσε κατά τη διάρκεια των μαθημάτων, τα οποία παρακολούθησα στο συγκεκριμένο πρόγραμμα μεταπτυχιακών σπουδών.