Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Διπλωματική Εργασία Ακαδημαϊκό έτος 2019-2020

Αριθμητικές Μέθοδοι για την Εκτίμηση Εμφωλευμένων Μέσων Τιμών

Αναστάσιος Μπουτσίνης
ge14005
Επιβλέπων καθηγητής: Αντώνης Παπαπαντολέων

10 Ιουλίου 2020

Εισαγωγή

Η εργασία αυτή πραγματεύεται τον αριθμητικό υπολογισμό εμφωλευμένων μέσων τιμών, αντλώντας υλικό κατά κύριο λόγο από τη δημοσίευση των Giles και Haji-Ali, με τίτλο Multilevel Nested Simulation for Efficient Risk Estimation. Πιο συγκεκριμένα, αν θεωρήσουμε μια μονοδιάστατη τυχαία μεταβλητή X, η οποία δεσμεύεται από τις τιμές μιας πολυδιάστατης μεταβλητής Y, εστιάζουμε στην αριθμητική εκτίμηση της τιμής $\mathbb{P}[\mathbb{E}[X|Y] \geq 0] = \mathbb{E}[H(\mathbb{E}[X|Y])]$ (1), όπου με H συμβολίζουμε τη συνάρτηση Heaviside. Στα κεφάλαια που θα ακολουθήσουν, θα επιχειρήσουμε να αντιμετωπίσουμε το πρόβλημα που διατυπώσαμε μέσα από διαφορετικές θεωρήσεις, οι οποίες έχουν ως πρωταρχικό στόχο την ελαχιστοποίηση της ολικής πολυπλοκότητας, διατηρώντας ταυτόχρονα ένα σταθερό επιθυμητό επίπεδο σφάλματος. Παρά το γεγονός πως η πολυπλοκότητα θα ορισθεί επίσημα στο πρώτο κεφάλαιο της εργασίας αυτής, για τις ανάγκες της εισαγωγής μας, μπορούμε να τη φανταστούμε σαν το σύνολο των ενεργειών στις οποίες οφείλουμε να προβούμε, προκειμένου να οδηγηθούμε στην εύρεση της ζητούμενης ποσότητας. Τα βασικά μαθηματικά εργαλεία στα οποία θα στηριχτούμε, είναι τα αριθμητικά σχήματα Monte Carlo και Multilevel Monte Carlo, ενώ για την πλήρη κατανόηση των αποδείξεων απαιτούνται βασικές γνώσεις πιθανοτήτων και μια εξοικείωση με τον ασυμπτωτικό συμβολισμό Bachmann-Landeau. Τέλος, όσον αφορά τη δομή, η εργασία μας χωρίζεται σε τέσσερα βασικά κεφάλαια, τα οποία θα παρουσιάσουμε συνοπτικά στο παρακάτω κείμενο.

Στο πρώτο κεφάκαιο εισάγουμε το αναγκαίο μαθηματικό υπόβαθρο για τη μετέπειτα ανάλυσή μας. Ειδικότερα, σε ένα αρχικό στάδιο, ορίζουμε την απλή μέθοδο Monte Carlo ως το βασικό αριθμητικό σχήμα προσδιορισμού αναμενόμενων τιμών και εξετάζουμε κάποιες από τις βασικές ιδιότητες του εκτιμητή της όπως είναι η αμεροληψία και η κανονικότητα. Σε αυτή μας την προσπάθεια, αξιοποιούμε μια σειρά θεμελιωδών μεγεθών όπως είναι η ολική πολυπλοκότητα της μεθόδου και το μέσο τετραγωνικό της σφάλμα. Στη συνέχεια, επικεντρωνόμαστε στη μέθοδο Multilevel Monte Carlo, η οποία γενικεύει με ένα φυσιολογικό τρόπο τα αποτελέσματα και τις έννοιες που είδαμε μέσα από το πρώτο αριθμητικό μας σχήμα. Είναι σημαντικό να τονίσουμε πως η νέα αυτή μέθοδος έχει ως κύριο στόχο της την αντιμετώπιση βασικών προβλημάτων του απλού σχήματος Monte Carlo και ειδικότερα τον περιορισμό του υψηλού υπολογιστικού κόστους που το συνοδεύει. Έχοντας, ολοκληρώσει την παρουσίαση των κύριων σημείων της μεθόδου Multilevel Monte Carlo διατυπώνουμε το θεώρημα ανάλυσης πολυπλοκότητας του Giles, το οποίο μας προσδιορίζει για ένα σταθερό επίπεδο σφάλματος και με χρήση κατάλληλων συνθηκών ελέγχου το ολικό κόστος(πολυπλοκότητα) της μεθόδου μας. Τέλος, αποδεικνύουμε πως η ιδιότητα της κανονικότητας του Monte Carlo εκτιμητή μεταφέρεται στη μέθοδο Multilevel Monte Carlo, λαμβάνοντας υπόψη μας το σχετικό αποτέλεσμα των Collier, Haji-Ali, Nobile, Von Schwerin και Tempone.

Το δεύτερο κεφάλαιο ξεκινά με τη μελέτη ενός προβλήματος από το χώρο των χρηματοοικονομικών, το οποίο αποτέλεσε σε σημαντικό βαθμό το κίνητρο για την ανάπτυξη της βασικής μας μεθοδολογίας. Το πρόβλημα αυτό, εξετάζει από την οπτική γωνία ενός παρατηρητή στο χρόνο $t = 0$, την πιθανότητα η απώλεια στην αξία ενός χρτοφυλακίου σε κάποιο μελλοντικό χρόνο, να ξεπεράσει ένα δοσμένο φράγμα $c > 0$. Όπως θα δούμε αναλυτικά, η φύση του ζητούμενου μεγέθους μας οδηγεί στη χρήση ενός εκτιμητή, ο οποίος περιλαμβάνει την εφαρμογή ενός σχήματος Monte Carlo σε δύο επίπεδα, ένα εσωτερικό και ένα εξωτερικό. Σε ένα πρώτο στάδιο, θεωρούμε πως η διαδικασίας προσομοίωσης είναι ομοιόμορφη, δηλαδή υποθέτουμε πως για την κάθε εξωτερική τιμή που εξετάζουμε, το πλήθος του εσωτερικού δείγματος παραμένει σταθερό. Κάτω από το συγκεκριμένο πλαίσιο, αναζητούμε βέλτιστες τιμές των παραμέτρων του εσωτερικού και εξωτερικού πλήθους δείγματος και αξιοποιώντας το θεώρημα των Gordy και Juneja, οδηγούμαστε σε μια πρώτη έκφραση για την πολυπλοκότητα του αριθμητικού μας σχήματος. Εν συνεχεία, διατηρώντας τη λογική του ομοιόμορφου δείγματος, προσεγγίζουμε το αρχικό μας πρόβλημα μέσα από μια διαφορετική θεώρηση, σύμφωνα με την οποία η ζητούμενη πιθανότητα λαμβάνει τελικά μια μορφή σαν αυτή που εμφανίζεται στη σχέση (1), με αποτέλεσμα το πρόβλημά μας να μεταπίπτει στη γενική μελέτη εμφωλευμένων μέσων τιμών. Τέλος, στο κεφάλαιο αυτό, προχωρούμε σε μια προσπάθεια κατάλληλης προσαρμογής του εσωτερικού δείγματος, με βάση τις εξωτερικές τιμές δέσμευσης και η νέα μέθοδος που προκύπτει φέρει μειωμένο ολικό κόστος σε σχέση με το εμφωλευμένο ομοιόμορφο αριθμητικό σχήμα.

Στο τρίτο κεφάλαιο της εργασίας μας θα επιχειρήσουμε να προσδιορίσουμε τη ζητούμενη μέση τιμή εφαρμόζοντας τη μέθοδο Multilevel Monte Carlo. Αξίζει να αναφέρουμε πως στο συγκεκριμένο αριθμητικό σχήμα, το ρόλο της μεταβλητής διακριτοποίησης παίζει η Heaviside συνάρτηση της εκτιμήτριας που χρησιμοποιούμε για την εκτέλεση της εσωτερικής μεθόδου Monte Carlo. Μέσα από την ανάπτυξη της συγκεκριμένης μεθοδολογίας, στοχεύουμε να μειώσουμε περαιτέρω το ολικό υπολογιστικό κόστος, στηριζόμενοι στο θεώρημα ανάλυσης πολυπλοκότητας του Giles. Ειδικότερα, με τη βοήθεια μιας σειράς λημμάτων και λογικών υποθέσεων, οδηγούμαστε σε μια έκφραση για την τάξη της ολικής πολυπλοκότητας, μέσα από την οποία συμπεραίνουμε πως η επιθυμητή μείωση έχει πράγματι επιτευχθεί. Σε ένα επόμενο βήμα, εξετάζουμε κατά πόσο η συνύπαρξη του σχήματος Multilevel Monte Carlo και του προσαρμοσμένου εσωτερικού δείγματος που αναλύσαμε στο προηγούμενο κεφαλαιο μπορεί να επιφέρει βελτίωση στην αριθμητική πολυπλοκότητα του προβλήματός μας. Κάτω από το πλάισιο αυτό, ορίζουμε μια έκφραση για το πλήθος του εσωτερικού δείγματος, ως συνάρτηση του επιπέδου μελέτης του σχήματος Multilevel Monte Carlo, αλλά και της μέσης τιμής και διασποράς της μεταβλητής $X(Y)$. Το βασικό πρόβλημα που καλούμαστε να αντιμετωπίσουμε στην συγκεκριμένη προσέγγιση, είναι πως οι παραπάνω δείκτες δεν είναι a-priori γνωστοί και πρέπει να εκτιμηθούν κατάλληλα. Στο κεφάλαιο που εξετάζουμε, ξεπερνάμε τη δυσκολία αυτή θεωρώντας πως έχουμε πλήρη γνώση των αναγκαίων τιμών, μεταφέροντας έτσι τη γενική και σαφώς πιο σύνθετη περίπτωση στο κεφάλαιο που ακολουθεί. Τελικά λοιπόν, μέσω της χρήσης κατάλληλων θεωρημάτων και αξιοποιώντας μια γενικευμένη μορφή του θεωρήματος πολυπλοκότητας του Giles, καταλήγουμε πως το ολικό κόστος της νέας μας μεθόδου έχει μειωθεί.

Όπως προαναφέραμε, στο τελευταίο αυτό κεφάλαιο, θα μελετήσουμε το ολικό υπολογιστικό κόστος της μεθόδου Multilevel Monte Carlo με προσαρμοσμένο εσωτερικό δείγμα, εκτιμώντας τη μέση τιμή και τη διασπορά της μταβλητής $X(Y)$, δοθέντος της Y. Για το σκοπό αυτό, προχωράμε στην κατασκευή ενός αλγορίθμου με μεταβλητές εισόδου το επίπεδο του σχήματος Multilevel Monte Carlo και την τιμή y της μεταβλητής Y που εξετάζουμε. Ο αλγόριθμος αυτός, με τη χρήση κατάλληλων εκτιμητών, προσδιορίζει τους ζητούμενους δείκτες της μέσης τιμής και της διασποράς και μέσα από ελέγχους υποθέσεων οδηγεί στην εύρεση του πλήθους του εσωτερικού δείγματος. Παρά το γεγονός πως η τιμή του εσωτερικού δείγματος που εξάγεται από τον αλγόριθμό μας δεν είναι απόλυτα ισοδύναμη με την αντίστοιχη τιμή του προηγούμενου κεφαλάιου, ικανοποιεί κάποιες από τις βασικές της ιδιότητες οι οποίες μας οδηγούν στην επιθυμητή τάξη για την ολική πολυπλοκότητα. Το κεφάλαιο αυτό ολοκληρώνεται με μια ιδιαίτερα τεχνική απόδειξη από τη δημοσίευση των Giles και Haji-Ali, η οποία παρέχει τα αναγκαία αποτελέσματα για την εύρεση του ολικού αριθμητικού κόστους της μεθόδου μας.

Κλείνοντας την εισαγωγή μας, είναι σημαντικό να τονίσουμε πως έχει δοθεί ιδιαίτερη έμφαση στον τρόπο αναπαραγωγής αποδείξεων από δημοσιευμένα papers, καθώς παρουσιάζονται αναλυτικά όλα τα βήματα που απαιτούνται. Κάτι τέτοιο έχει γίνει με σκοπό τη διευκόλυνση του αναγνώστη, ενώ η εργασία έχει εμπλουτιστεί με διάφορες παρατηρήσεις, οι οποίες αποσκοπούν στην εμβάθυνση γύρω από τα βασικά αντικείμενα μελέτης.

Περιεχόμενα

1 Μέθοδοι Monte Carlo

1.1 Απλή Μέθοδος Monte Carlo

Η απλή μέθοδος Monte Carlo είναι ένα αριθμητικό σχήμα που χρησιμοποιείται για τον υπολογισμό αναμενόμενων τιμών τυχαίων μεταβλητών και στηρίζεται στο νόμο των μεγάλων αριθμών. Έστω λοιπόν $G : \Omega \to \mathbb{R}$ μια οποιαδήποτε μονοδιάστατη τυχαία μεταβλητή που δεν μπορούμε να προσομοιώσουμε και έστω πως επιθυμούμε να προσδιορίσουμε αριθμητικά την ποσότητα $\mathbb{E}[G] < \infty$. Για να το επιτύχουμε αυτό θα πρέπει καταρχάς να δημιουργήσουμε ένα τυχαίο δείγμα $\overline{G}_1, \overline{G}_2, \ldots, \overline{G}_M$ αριθμητικών προσεγγίσεων της μεταβλητής G, οι οποίες υποθέτουμε πως είναι ανεξάρτητες και φέρουν το ίδιο υπολογιστικό κόστος C. Αξίζει στο σημείο αυτό να αναφέρουμε πως σε πολλές περιπτώσεις οι αριθμητικές προσεγγίσεις της μεταβλητής G μπορούν να προκύψουν από την εφαρμογή ενός αριθμητικού σχήματος σε μια διαφορική εξίσωση, όπως είναι το σχήμα Euler. Για την ακολουθία τώρα που δημιουργήσαμε, από το νόμο των μεγάλων αριθμών παίρνουμε πως:

$$\frac{1}{M}\sum_{j=1}^{M} \overline{G}_j \xrightarrow[M\to\infty]{\mathbb{P}} \mathbb{E}\left[\overline{G}\right],$$

όπου στην περίπτωση που το αριθμητικό μας σχήμα συγκλίνει λαμβάνουμε ότι $\mathbb{E}\left[\overline{G}\right] \approx \mathbb{E}[G]$. Με βάση την παρατήρηση αυτή ορίζουμε τον εκτιμητή Monte Carlo της ποσότητας $\mathbb{E}[G]$ μέσω της σχέσης $I := \frac{1}{M}\sum_{j=1}^{M} \overline{G}_j$.

Στο σημείο αυτό θα παραθέσουμε κάποιους βασικούς ορισμούς και μια σειρά από ιδιότητες, μέσω παρατηρήσεων, οι οποίες θα μας φανούν ιδιαίτερα χρήσιμες στη συνέχεια της εργασίας αυτής. Έτσι έχουμε ότι:

Ορισμός 1.1.1

Έστω G μια τυχαία μεταβλητή και $(\overline{G}_j)_{j=1}^{M}$ ένα δείγμα αριθμητικών προσεγγίσεων της μεταβλητής αυτής. Έστω επιπλέον $I = \frac{1}{M}\sum_{j=1}^{M} \overline{G}_j$ η εκτιμήτρια Monte Carlo της ζητούμενης αναμενόμενης τιμής. Τότε ορίζουμε το μέσο τετραγωνικό σφάλμα της μεθόδου μέσα από τη σχέση:

$$MSE = \mathbb{E}\left[(I - \mathbb{E}[G])^2\right] \equiv \mathbb{E}\left[\epsilon^2\right].$$

Παρατήρηση 1.1.2

Ο εκτιμητής της μεθόδου Monte Carlo είναι αμερόληπτος. Πράγματι βλέπουμε ότι:

$$\mathbb{E}[I] = \mathbb{E}\left[\frac{1}{M}\sum_{j=1}^{M} \overline{G}_j\right] = \frac{1}{M}\sum_{j=1}^{M} \mathbb{E}\left[\overline{G}_j\right] \overset{\text{ισον.}}{=} \frac{1}{\cancel{M}}\cancel{M}\,\mathbb{E}\left[\overline{G}\right] = \mathbb{E}\left[\overline{G}\right] \approx \mathbb{E}[G].$$

Παρατήρηση 1.1.3

Λογω ανεξαρτησίας των αριθμητικών προσεγγίσεων για τη διασπορά του Monte Carlo εκτιμητή ισχύει ότι:

$$Var[I] = Var\left[\frac{1}{M}\sum_{j=1}^{M}\overline{G}_j\right] \overset{\text{ανεξ.}}{=} \frac{1}{M^2}\sum_{j=1}^{M} Var\left[\overline{G}_j\right] \overset{\text{ισον.}}{=} \frac{1}{M^2} M Var\left[\overline{G}\right] = \frac{Var[\overline{G}]}{M}.$$

Παρατήρηση 1.1.4

Για το μέσο τετραγωνικό σφάλμα παρατηρούμε ότι:

$$\begin{aligned} MSE &= \mathbb{E}\left[(I-\mathbb{E}[G])^2\right] \\ &= \mathbb{E}\left[(I-\mathbb{E}\left[\overline{G}\right]+\mathbb{E}\left[\overline{G}\right]-\mathbb{E}[G])^2\right] \\ &= \mathbb{E}\left[(I-\mathbb{E}\left[\overline{G}\right])^2\right]+2(\mathbb{E}\left[\overline{G}\right]-\mathbb{E}[G])\mathbb{E}\left[I-\mathbb{E}\left[\overline{G}\right]\right]+\left(\mathbb{E}\left[\overline{G}\right]-\mathbb{E}[G]\right)^2 \\ &= \mathbb{E}\left[(I-\mathbb{E}[I])^2\right]+\left(\mathbb{E}\left[\overline{G}\right]-\mathbb{E}[G]\right)^2 \\ &= Var[I]+\left(\mathbb{E}\left[\overline{G}\right]-\mathbb{E}[G]\right)^2. \end{aligned}$$

Από την παραπάνω σχέση συμπεραίνουμε πως το μέσο τετραγωνικό σφάλμα μπορεί να γραφτεί ως το άθροισμα της διασποράς του Monte Carlo εκτιμητή και ενός όρου σφάλματος διακριτοποίησης.

Παρατήρηση 1.1.5

Εκμεταλλευόμενοι το γεγονός πως οι $(\overline{G}_j)_{j=1}^{M}$ είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές και αξιοποιώντας το κεντρικό οριακό θεώρημα παίρνουμε ότι:

$$\sqrt{M}\left(\frac{\sum_{j=1}^{M}\overline{G}_j}{M}-\mathbb{E}\left[\overline{G}\right]\right) \underset{M\to\infty}{\longrightarrow} N\left(0, Var\left[\overline{G}\right]\right) \iff \frac{\sum_{j=1}^{M}\overline{G}_j/M-\mathbb{E}[\overline{G}]}{\sqrt{\frac{Var[\overline{G}]}{M}}} \underset{M\to\infty}{\longrightarrow} N(0,1) \overset{1.1.2}{\underset{1.1.3}{\iff}} \frac{I-\mathbb{E}[I]}{\sqrt{Var[I]}} \underset{M\to\infty}{\longrightarrow}$$
$$N(0,1) \iff \lim_{M\to\infty}\mathbb{P}\left[\frac{I-\mathbb{E}[I]}{\sqrt{Var[I]}}\leq z\right]=\Phi(z),$$

όπου με Φ έχουμε συμβολίσει τη συνάρτηση κατανομής πιθανότητας μιας τυπικής κανονικής τυχαίας μεταβλητής.

Ορισμός 1.1.6

Έστω M το πλήθος προσομοιώσεων της μεθόδου Monte Carlo και C το κόστος υπολογισμού της αριθμητικής προσέγγισης $\overline{G}$. Τότε το ολικό υπολογιστικό κόστος ή υπολογιστική πολυπλοκότητα της μεθόδου ορίζεται μέσω του τύπου $\mathcal{C} = M.C$.

Παρατήρηση 1.1.7

Το βασικό μειονέκτημα της απλής μεθόδου Monte Carlo εντοπίζεται ακριβώς στον παραπάνω ορισμό. Συγκεκριμένα, σε περιπτώσεις που το κόστος της αριθμητικής προσέγγισης είναι υψηλό και επιλέξουμε ένα μεγάλο πλήθος προσομοιώσεων για να επιτύχουμε σύγκλιση μέσα από το νόμο των μεγάλων αριθμών, το υπολογιστικό κόστος της μεθόδου μας λαμβάνει μια ιδιαίτερα υψηλή τιμή. Για να κρατήσουμε λοιπόν την υπολογιστική πολυπλοκότητα σε χαμηλά επίπεδα θα θέλαμε ιδανικά κάθε φορά που ο ένας από τους όρους M, C λαμβάνει υψηλή τιμή ο άλλος να λαμβάνει χαμηλή διατηρώντας ταυτόχρονα τόσο τη σύγκλιση του αριθμητικού σχήματος, όσο και τη σύγκλιση μέσω του νόμου των μεγάλων αριθμών. Αυτόν ακριβώς το σπουδαίο ρόλο επιτελεί η μέθοδος Multilevel Monte Carlo η οποία ακολουθεί στην επόμενη ενότητα.

1.2 Μέθοδος Multilevel Monte Carlo

Όπως είδαμε στην απλή μέθοδο Monte Carlo, το ολικό σφάλμα της μεθόδου είναι αποτέλεσμα τόσο της διαδικασίας διακριτοποίησης όσο και των εκτελούμενων προσομοιώσεων. Με άλλα λόγια, εξηγήσαμε πως για να επιτύχουμε σύγκλιση της μεθόδου και εξαγωγή του επιθυμητού αποτελέσματος απαιτείται μεγάλο πλήθος προσομοιώσεων (νόμος των μεγάλων αριθμών) αλλά και μεγάλο κόστος για τον προσδιορισμό της αριθμητικής μας προσέγγισης. Έτσι στην προσπάθεια μας να πάρουμε αποτέλεσμα με μια ικανοποιητική ακρίβεια οδηγούμαστε σε ένα δυσβάσταχτο υπολογιστικό κόστος, το οποίο προφανώς και θα θέλαμε να περιορίσουμε. Σε αυτά τα πλαίσια, γεννήθηκε η ιδέα της μεθόδου Multilevel Monte Carlo η οποία όπως φαίνεται καθαρά από το όνομά της είναι μια μέθοδος Monte Carlo που λειτουργεί σε πολλά επίπεδα.

Το βασικό ερώτημα που δημιουργείται στο σημείο αυτό είναι τι εννοούμε όταν λέμε πως η μέθοδος αυτή δρα σε πολλά επίπεδα. Αν ανατρέξουμε σε προηγούμενο ορισμό που δώσαμε, μπορούμε να δούμε πως το ολικό υπολογιστικό έργο της αρχικής μας μεθόδου δίνεται μέσω του τύπου $\mathcal{C} = M.C$, όπου M είναι το πλήθος των προσομοιώσεων μας και C το κόστος για τον προσδιορισμό της εκάστοτε αριθμητικής προσέγγισης. Αυτό που επιτελεί στην πραγματικότητα η νέα μας μέθοδος είναι να σπάει το παραπάνω κόστος σε επιμέρους τμήματα, δηλαδή $\mathcal{C} = \sum_{l=0}^{L} \mathcal{C}_l$. Κάθε μία από τις ποσότητες $\mathcal{C}_l$ δηλώνει ολικό κόστος στο $l-$οστό επίπεδο και κατ΄ επέκταση θα υπολογίζεται μέσω της σχέσης $\mathcal{C}_l = M_l.C_l$ με το πλήθος των προσομοιώσεων και το αριθμητικό κόστος να διαφοροποιούνται από το ένα επίπεδο στο άλλο.

Στο σημείο αυτό, είναι σημαντικό να αναφέρουμε πως τα επίπεδα διαμορφώνονται κατά τέτοιο τρόπο ώστε σε όσο πιο υψηλό εργαζόμαστε τόσο μεγαλύτερο να είναι το αριθμητικό υπολογιστικό κόστος που οφείλουμε να δαπανήσουμε. Για παράδειγμα , αν εστιάσουμε σε ένα πλαίσιο στο οποίο οι αριθμητικές προσεγγίσεις εξάγονται από την εφαρμογή ενός σχήματος Euler, καθώς μεταβαίνουμε προς πιο υψηλά επίπεδα θα πρέπει να αυξάνουμε τα σημεία της διαμέρισης εντός του διαστήματος μελέτης. Η ιδέα που χρησιμοποιούμε λοιπόν είναι απλή και είναι η παρακάτω: στα αρχικά επίπεδα που το υπολογιστικό κόστος της αριθμητικής προσέγγισης είναι μικρό χρησιμοποιούμε υψηλό πλήθος προσομοιώσεων σε αντίθεση με τα τελικά επίπεδα, όπου και ένα μικρό δείγμα μπορεί να οδηγήσει στο επιθυμητό αποτέλεσμα. Αξίζει να αναφέρουμε πως ο ισχυρισμός μας ότι στα υψηλά επίπεδα επαρκεί ένα μικρό πλήθος δείγματος για την ώρα φαίνεται αυθαίρετος, ωστόσο θα εξηγηθεί παρακάτω μέσω του εκτιμητή στον οποίο θα στηριχτούμε. Αν προς στιγμήν δεχτούμε την ιδέα που αναπτύξαμε παραπάνω, τότε αυτό που έχουμε κατορθώσει είναι να γράψουμε το ολικό κόστος με τη βοήθεια ενός αθροίσματος στο οποίο ο κάθε όρος προκύπτει μέσω του γινομένου μεταξύ ενός μικρού και ενός μεγάλου αριθμού και όχι δυο μεγάλων όπως είδαμε στην απλή μέθοδο Monte Carlo.

Είμαστε πλέον έτοιμοι να αναζητήσουμε τον εκτιμητή που θα χρησιμοποιήσουμε στη μέθοδο Multilevel Monte Carlo. Σίγουρα, με βάση τη μελέτη που προηγήθηκε ο εκτιμητής αυτός θα πρέπει να περιλαμβάνει αριθμητικές προσεγγίσεις πάνω σε όλα τα επίπεδα μελέτης και να διατηρεί τις καλές ιδιότητες που συναντήσαμε στην απλή μέθοδο Monte Carlo, όπως είναι η αμεροληψία. Με βάση αυτή τη λογική ορίζουμε τον εκτιμητή Multilevel Monte Carlo της ποσότητας $\mathbb{E}[G]$ πάνω στα $l = 0, \ldots, L$ επίπεδα μελέτης μέσω της σχέσης:

$$I := \frac{1}{M_0} \sum_{j=1}^{M_0} \overline{G}_j^0 + \sum_{l=1}^{L} \frac{1}{M_l} \sum_{j=1}^{M_l} \left(\overline{G}_j^l - \overline{G}_j^{l-1} \right),$$

όπου με $\overline{G}^l$ συμβολίζουμε την αριθμητική προσέγγιση της μεταβλητής G στο $l-$οστό επίπεδο μελέτης και με M_l το αντίστοιχο πλήθος προσομοιώσεων. Ετσι, στο κάθε επίπεδο χρησιμοποιούμε ουσιαστικά απλή μέθοδο Monte Carlo με τη λεπτομέρεια πως μετά το μηδενικό επίπεδο δεν εκτιμάμε την ποσότητα $\mathbb{E}\left[\overline{G}^l\right]$, αλλά την $\mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right]$. Μπορούμε εύκολα να διαπιστώσουμε τη σύγκλιση του παραπάνω εκτιμητή χρησιμοποιώντας το νόμο των μεγάλων αριθμών. Συγκεκριμένα, βλέπουμε ότι:

$$I = \frac{1}{M_0} \sum_{j=1}^{M_0} \overline{G}_j^0 + \sum_{l=1}^{L} \frac{1}{M_l} \sum_{j=1}^{M_l} \left(\overline{G}_j^l - \overline{G}_j^{l-1} \right) \xrightarrow[M_l \to \infty]{\mathbb{P}} \mathbb{E}\left[\overline{G}^0\right] + \sum_{l=1}^{L} \mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right] = \mathbb{E}\left[\overline{G}^L\right],$$

δηλαδή δείξαμε πως:

$I \underset{M_l \to \infty}{\xrightarrow{\mathbb{P}}} \mathbb{E}\left[\overline{G}^L\right]$.

Επειδή όμως στο τελικό επίπεδο το αριθμητικό κόστος είναι πολύ υψηλό συμπεραίνουμε πως $\mathbb{E}\left[\overline{G}^L\right] \approx \mathbb{E}[G]$ και κατ' επέκταση η μέθοδος μας συγκλίνει. Κατ' αναλογία τώρα με την απλή μέθοδο Monte Carlo για τη νέα μας μέθοδο ισχύουν τα ακόλουθα:

Ορισμός 1.2.1

Έστω G μια τυχαία μεταβλητή και $l = 0, 1, \ldots, L$ μια πεπερασμένη ακολουθία επιπέδων της μεθόδου Multilevel Monte Carlo. Αν $\overline{G}^l$ είναι η αριθμητική προσέγγιση της μεταβλητής G στο $l-$οστό επίπεδο και I ο εκτιμητής που εξετάσαμε παραπάνω, ορίζουμε το μέσο τετραγωνικό σφάλμα της μεθόδου μέσα από τη σχέση:

$$MSE = \mathbb{E}\left[(I - \mathbb{E}[G])^2\right] \equiv \mathbb{E}\left[\epsilon^2\right].$$

Ορισμός 1.2.2

Έστω M_l το πλήθος των προσομοιώσεων της μεθόδου Multilevel Monte Carlo στο $l-$οστό επίπεδο με $l = 0, \ldots, L$, C_0 το κόστος υπολογισμού της αριθμητικής προσέγγισης $\overline{G}^0$ και C_l το κόστος υπολογισμού της αριθμητικής προσέγγισης $\overline{G}^l - \overline{G}^{l-1}$, όπου $l = 1, 2, \ldots, L$. Τότε η υπολογιστική πολυπλοκότητα της μεθόδου Multilevel Monte Carlo ορίζεται ως:

$$\mathcal{C} = \sum_{l=0}^{L} M_l C_l.$$

Παρατήρηση 1.2.3

Ο εκτιμητής της μεθόδου Multilevel Monte Carlo είναι αμερόληπτος. Πράγματι έχουμε πως ισχύει ότι:

$$\begin{aligned}
\mathbb{E}[I] &= \mathbb{E}\left[\frac{1}{M_0}\sum_{j=1}^{M_0}\overline{G}_j^0 + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{j=1}^{M_l}\left(\overline{G}_j^l - \overline{G}_j^{l-1}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{M_0}\sum_{j=1}^{M_0}\overline{G}_j^0\right] + \sum_{l=1}^{L}\mathbb{E}\left[\frac{1}{M_l}\sum_{j=1}^{M_l}\left(\overline{G}_j^l - \overline{G}_j^{l-1}\right)\right] \\
&= \mathbb{E}\left[\overline{G}^0\right] + \sum_{l=1}^{L}\mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right] \\
&= \mathbb{E}\left[\overline{G}^0\right] + \sum_{l=1}^{L}\left(\mathbb{E}\left[\overline{G}^l\right] - \mathbb{E}\left[\overline{G}^{l-1}\right]\right) \\
&= \mathbb{E}\left[\overline{G}^L\right] \approx \mathbb{E}[G].
\end{aligned}$$

Παρατήρηση 1.2.4

Λόγω ανεξαρτησίας των αριθμητικών προσεγγίσεων μεταξύ των επιπέδων $l = 0, 1, \ldots, L$, αλλά και μέσα στο ίδιο επίπεδο, για τη διασπορά του εκτιμητή της μεθόδου Multilevel Monte Carlo παίρνουμε ότι:

$$\begin{aligned}
Var[I] &= Var\left[\frac{1}{M_0}\sum_{j=1}^{M_0}\overline{G}_j^0 + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{j=1}^{M_l}\left(\overline{G}_j^l - \overline{G}_j^{l-1}\right)\right] \\
&= \frac{1}{M_0^2}\sum_{j=1}^{M_0}Var\left[\overline{G}_j^0\right] + \sum_{l=1}^{L}\frac{1}{M_l^2}\sum_{j=1}^{M_l}Var\left[\overline{G}_j^l - \overline{G}_j^{l-1}\right] \\
&= \frac{1}{M_0^2}M_0 Var\left[\overline{G}^0\right] + \sum_{l=1}^{L}\frac{1}{M_l^2}M_l Var\left[\overline{G}^l - \overline{G}^{l-1}\right] \\
&= \frac{Var\left[\overline{G}^0\right]}{M_0} + \sum_{l=1}^{L}Var\left[\overline{G}^l - \overline{G}^{l-1}\right]/M_l.
\end{aligned}$$

Παρατήρηση 1.2.5

Για το μέσο τετραγωνικό σφάλμα της μεθόδου Multilevel Monte Carlo βλέπουμε ότι:

$$\begin{aligned}
MSE &= \mathbb{E}\Big[\Big(I - \mathbb{E}\big[G\big]\Big)^2\Big] \\
&= \mathbb{E}\left[\left(I - \mathbb{E}\left[\overline{G}^L\right] + \mathbb{E}\left[\overline{G}^L\right] - \mathbb{E}\big[G\big]\right)^2\right] \\
&= \mathbb{E}\left[\left(I - \mathbb{E}\left[\overline{G}^L\right]\right)^2\right] + 2\left(\mathbb{E}\left[\overline{G}^L\right] - \mathbb{E}\big[G\big]\right)\mathbb{E}\left[I - \mathbb{E}\left[\overline{G}^L\right]\right] + \left(\mathbb{E}\left[\overline{G}^L\right] - \mathbb{E}\big[G\big]\right)^2.
\end{aligned}$$

Δείξαμε όμως ότι $\mathbb{E}\big[I\big] = \mathbb{E}\left[\overline{G}^L\right]$, συνεπώς ισχύει ότι:

$$MSE = \mathbb{E}\Big[\Big(I - \mathbb{E}\big[I\big]\Big)^2\Big] + \left(\mathbb{E}\left[\overline{G}^L\right] - \mathbb{E}\big[G\big]\right)^2 = Var\big[I\big] + \left(\mathbb{E}\left[\overline{G}^L\right] - \mathbb{E}\big[G\big]\right)^2.$$

Βλέπουμε λοιπόν πως το μέσο τετραγωνικό σφάλμα μπορεί να προκύψει ως το άθροισμα της διασποράς του Multilevel Monte Carlo εκτιμητή και ενός όρου σφάλματος διακριτοποίησης στο τελικό επίπεδο μελέτης.

Παρατήρηση 1.2.6

Σε προγενέστερη ανάλυσή μας δηλώσαμε πως στα υψηλά επίπεδα l μπορούμε να επιτύχουμε ικανοποιητική ακρίβεια ακόμα και με ένα πολύ μικρό πλήθος προσομοιώσεων, δίχως να προβούμε σε κάποια αιτιολόγηση του ισχυρισμού αυτού. Έχοντας ωστόσο πλέον ορίσει τα βασικά εργαλεία της μεθόδου μας είμαστε σε θέση να προχωρήσουμε σε μια σχετική εξήγηση. Όπως έχουμε ήδη αναφέρει, ουσιαστικά στη μέθοδο Multilevel Monte Carlo εφαρμόζουμε απλή μέθοδο Monte Carlo σε πολλαπλά επίπεδα με αποτέλεσμα το ερώτημα που διατυπώσαμε παραπάνω να λαμβάνει την ακόλουθη μορφή: διαθέτοντας μόνο ένα μικρό δείγμα M_l μπορούμε να οδηγηθούμε σε μια ικανοποιητική αριθμητική προσέγγιση της τιμής $\mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right]$ στα υψηλά επίπεδα μελέτης; Στο σημείο αυτό θα ήταν χρήσιμο να θυμηθούμε πως η ακρίβεια της απλής μεθόδου Monte Carlo εξετάζεται από το μέσο τετραγωνικό σφάλμα, το οποίο στην περίπτωσή μας θα δίνεται μέσω της σχέσης:

$$MSE = \mathbb{E}\left[\left(\frac{1}{M_l}\sum_{j=1}^{M_l}\left(\overline{G}_j^l - \overline{G}_j^{l-1}\right) - \mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right]\right)^2\right] \equiv \mathbb{E}\left[\left(\widetilde{I} - \mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right]\right)^2\right].$$

Γνωρίζουμε όμως πως ο εκτιμητής μας είναι αμερόληπτος, συνεπώς λαμβάνουμε ότι $\mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right] = \mathbb{E}\big[\widetilde{I}\big]$

και η παραπάνω σχέση γράφεται τελικά με τον ακόλουθο τρόπο:

$$MSE = \mathbb{E}\left[\left(\widetilde{I} - \mathbb{E}\left[\widetilde{I}\right]\right)^2\right] = Var\left[\widetilde{I}\right] \overset{\text{ορ.}}{=} Var\left[\frac{1}{M_l}\sum_{j=1}^{M_l}\left(\overline{G}_j^l - \overline{G}_j^{l-1}\right)\right] \overset{1.1.3}{=} \frac{Var\left[\overline{G}^l - \overline{G}^{l-1}\right]}{M_l}.$$

Παρατηρούμε όμως πως οι αριθμητικές προσεγγίσεις $\overline{G}^l$ και $\overline{G}^{l-1}$ εκτιμούν την ίδια τυχαία μεταβλητή G, με αποτέλεσμα στα υψηλά επίπεδα l να ισχύει ότι $Var\left[\overline{G}^l - \overline{G}^{l-1}\right] \to 0$. Έτσι τελικά, με βάση τα παραπάνω αντιλαμβανόμαστε πως ακόμα και για ένα μικρό δείγμα M_l μπορούμε να επιτύχουμε μικρή τιμή του MSE ή με άλλα λόγια μπορούμε να επιτύχουμε μικρό σφάλμα για τη ζητούμενη ποσότητα στο υψηλό επίπεδο μελέτης μας.

Έχοντας πλέον διατυπώσει του βασικούς ορισμούς της μεθόδου Multilevel Monte Carlo και αναλύσει κάποιες από τις κύριες ιδιότητές της θα επιχειρήσουμε να εξετάσουμε κατά πόσο υπάρχει ένα πλήθος προσομοιώσεων M_l το οποίο ελαχιστοποιεί την πολυπλοκότητα της μεθόδου μας και κατά αυτή την έννοια είναι βέλτιστο. Έτσι λοιπόν θα εργαστούμε κάτω από το ακόλουθο πλαίσιο:

Ορισμός του προβλήματος

Έστω C_l, $l = 1, 2, \ldots, L$ το κόστος υπολογισμού της αριθμητικής προσέγγισης $\overline{G}^l - \overline{G}^{l-1}$ και C_0 το κόστος υπολογισμού της αριθμητικής προσέγγισης $\overline{G}^0$. Έστω επιπλέον ότι γνωρίζουμε τη διασπορά της τυχαίας μεταβλητής $\overline{G}^l - \overline{G}^{l-1}$, $\forall\, l = 1, 2, \ldots, L$ την οποία συμβολίζουμε με V_l καθώς και τη διασπορά της τυχαίας μεταβλητής $\overline{G}^0$ την οποία συμβολίζουμε με V_0. Καλούμαστε τότε για σταθερό επίπεδο διασποράς του εκτιμητή Multilevel Monte Carlo (έστω V) να προσδιορίσουμε το βέλτιστο πλήθος προσομοιώσεων M_l ή αλλιώς την τιμή του M_l που ελαχιστοποιεί το ολικό υπολογιστικό κόστος της μεθόδου.

Λύση του προβλήματος

Με βάση τα εργαλεία που αναπτύξαμε έως τώρα βλέπουμε πως το συνολικό υπολογιστικό κόστος θα δίνεται από τη σχέση $\mathcal{C} = \sum_{l=0}^{L} M_l C_l$ και για τη διασπορά του εκτιμητή θα ισχύει ότι $Var[I] = \sum_{i=0}^{L} V_i/M_i = V$. Παρατηρούμε λοιπόν πως έχουμε να αντιμετωπίσουμε ένα πρόβλημα ελαχιστοποίησης υπό περιορισμό και για το λόγο αυτό θα στηριχτούμε στη μέθοδο του Lagrange. Ορίζουμε καταρχάς τη συνάρτηση του Lagrange μέσω του παρακάτω τύπου :

$$L(M_0, M_1, \ \ldots \ , M_L, \lambda) = \sum_{l=0}^{L} M_l C_l - \lambda\left(V - \sum_{i=0}^{L} V_i/M_i\right).$$

Τότε για την πρώτη παράγωγο της συνάρτησης αυτής ως προς την τυχούσα παράμετρο M_l παίρνουμε ότι:

$$\frac{\partial L}{\partial M_l} = 0 \Longrightarrow C_l - \lambda\frac{V_l}{M_l^2} = 0 \Longrightarrow \lambda\frac{V_l}{M_l^2} = C_l \Longrightarrow M_l^2 = \lambda\frac{V_l}{C_l} \overset{M_l>0}{\Longrightarrow} M_l = \sqrt{\lambda\frac{V_l}{C_l}}\ . \quad (1.1)$$

Επιπλέον, ισχύει και ο περιορισμός:

$$\sum_{i=0}^{L} V_i/M_i = V \overset{(1.1)}{\Longrightarrow} \sum_{i=0}^{L} V_i/\sqrt{\lambda V_i/C_i} = V \Longrightarrow \sum_{i=0}^{L}\sqrt{V_i C_i/\lambda} = V \Longrightarrow \sqrt{\lambda} = \frac{1}{V}\sum_{i=0}^{L}\sqrt{V_i C_i}\ . \quad (1.2)$$

Εισάγοντας τώρα το αποτέλεσμα (1.2) στη σχέση (1.1) λαμβάνουμε ότι:

$$M_l = \frac{1}{V}\sum_{i=0}^{L}\sqrt{V_i C_i}\sqrt{\frac{V_l}{C_l}} \Longrightarrow M_l = \frac{1}{V}\sqrt{\frac{V_l}{C_l}}\sum_{i=0}^{L}\sqrt{V_i C_i}\ .$$

Δεν πρέπει όμως να ξεχνάμε πως το M_l συμβολίζει πλήθος προσομοιώσεων, συνεπώς θα πρέπει να είναι ένας αυστηρά θετικός ακέραιος αριθμός. Αν λάβουμε την παρατήρηση αυτή υπόψη μας καταλήγουμε πως:

$$M_l = \left\lceil \frac{1}{V}\sqrt{\frac{V_l}{C_l}}\sum_{i=0}^{L}\sqrt{V_iC_i}\right\rceil,\ \forall\ l = 0, 1, \ldots, L.$$

Προτού προχωρήσουμε στην επόμενη ενότητα, αξίζει να αναφέρουμε πως ο παραπάνω τύπος δεν είναι απλά ένα σημαντικό θεωρητικό αποτέλεσμα αλλά βρίσκει εφαρμογή στην κατασκευή αλγορίθμου για τη μέθοδο Multilevel Monte Carlo, όπου προφανώς και αποσκοπούμε στην ελαχιστοποίηση του ολικού υπολογιστικού κόστους.

1.3 Θεώρημα Ανάλυσης Πολυπλοκότητας

Έχοντας πλέον ορίσει τη μέθοδο Multilevel Monte Carlo και μελετήσει ορισμένες από τις βασικές της ιδιότητες, στην ενότητα που ακολουθεί θα επικεντρωθούμε στο υπολογιστικό της κόστος. Το βασικό μας εργαλείο σε αυτή την προσπάθεια είναι το θεώρημα ανάλυσης πολυπλοκότητας του Giles, το οποίο θα διατυπώσουμε και θα αποδείξουμε αναλυτικά. Είναι σκόπιμο να αναφέρουμε πως η αξία του συγκεκριμένου θεωρήματος θα φανεί με καθαρό τρόπο στα κεφάλαια που θα επακολουθήσουν.

Θεώρημα 1.3.1(Giles)

Έστω $G : \Omega \to \mathbb{R}$ μια μονοδιάστατη τυχαία μεταβλητή και $\overline{G}^l$ η αριθμητική της προσέγγιση στο $l-$οστό επίπεδο μελέτης. Αν υπάρχουν ανεξάρτητοι εκτιμητές I_l βασισμένοι σε M_l το πλήθος προσομοιώσεις, η κάθε μία με κόστος (διακριτοποίησης) C_l και διασπορά V_l, και θετικές σταθερές $a, \beta, \gamma, c_1, c_2, c_3$ τέτοιες ώστε $a \geq \frac{1}{2}\min(\beta, \gamma)$ και

i) $\left|\mathbb{E}\left[\overline{G}^l - G\right]\right| \leq c_1 2^{-al}$

ii) $\mathbb{E}\Big[I_l\Big] = \begin{cases} \mathbb{E}\left[\overline{G}^0\right], & l = 0 \\ \mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right], & l > 0 \end{cases}$

iii) $V_l \leq c_2 2^{-\beta l}$

iv) $C_l \leq c_3 2^{\gamma l}$,

τότε υπάρχει θετική σταθερά c_4 ώστε $\forall\ \epsilon < e^{-1}$ να υπάρχουν τιμές L και M_l για τις οποίες ο multilevel εκτιμητής $I = \sum\limits_{l=0}^{L} I_l$ έχει μέσο τετραγωνικό σφάλμα με φράγμα:

$$MSE = \mathbb{E}\left[(I - \mathbb{E}[G])^2\right] < \epsilon^2$$

και υπολογιστικό κόστος με φράγμα:

$$\mathcal{C} \leq \begin{cases} c_4\epsilon^{-2}, & \beta > \gamma \\ c_4\epsilon^{-2}(\log\epsilon)^2, & \beta = \gamma \\ c_4\epsilon^{-2-(\gamma-\beta)/\alpha}, & \beta < \gamma \end{cases}.$$

Απόδειξη: Για την απόδειξη του θεωρήματος αυτού διακρίνουμε τις εξής περιπτώσεις:

α) β = γ

Θεωρούμε καταρχάς $L = \left\lceil \frac{\log(\sqrt{2}c_1\epsilon^{-1})}{a\log 2} \right\rceil$ και εξ' ορισμού θα πάρουμε την ακόλουθη σχέση:

$\frac{\log(\sqrt{2}c_1\epsilon^{-1})}{a\log 2} \leq L < \frac{\log(\sqrt{2}c_1\epsilon^{-1})}{a\log 2} + 1.$

Στο σημείο αυτό θα εργαστούμε με την κάθε ανισότητα χωριστά. Έτσι λοιπόν έχουμε ότι:

- $\frac{\log(\sqrt{2}c_1\epsilon^{-1})}{a\log 2} \leq L \Longrightarrow \log(\sqrt{2}c_1\epsilon^{-1}) \leq \log 2^{aL} \Longrightarrow \sqrt{2}c_1\epsilon^{-1} \leq 2^{aL} \Longrightarrow c_1 2^{-aL} \leq \frac{1}{\sqrt{2}}\epsilon$

- $L < \frac{\log(\sqrt{2}c_1\epsilon^{-1})}{a\log 2} + 1 \Longrightarrow \log 2^{(L-1)a} < \log(\sqrt{2}c_1\epsilon^{-1}) \Longrightarrow 2^{(L-1)a} < \sqrt{2}c_1\epsilon^{-1} \Longrightarrow c_1 2^{-aL} >$
$2^{-a}\frac{1}{\sqrt{2}}\epsilon.$

Τελικά λοιπόν δείξαμε πως:

$2^{-a}\frac{1}{\sqrt{2}}\epsilon < c_1 2^{-aL} \leq \frac{1}{\sqrt{2}}\epsilon. \quad (1.3)$

Βλέπουμε στο σημείο αυτό πως ισχύει ότι:

$$\begin{aligned}
\Big(\mathbb{E}\Big[I\Big] - \mathbb{E}\Big[G\Big]\Big)^2 &= \left(\mathbb{E}\left[\sum_{l=0}^{L} I_l\right] - \mathbb{E}\Big[G\Big]\right)^2 \\
&= \left(\sum_{l=0}^{L} \mathbb{E}\Big[I_l\Big] - \mathbb{E}\Big[G\Big]\right)^2 \\
&= \left(\sum_{l=1}^{L} \mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right] + \mathbb{E}\left[\overline{G}^0\right] - \mathbb{E}\Big[G\Big]\right)^2 \\
&= \left(\mathbb{E}\left[\overline{G}^L\right] - \mathbb{E}\Big[G\Big]\right)^2 \\
&= \left(\mathbb{E}\left[\overline{G}^L - G\right]\right)^2 \\
&\overset{(i)}{\leq} \Big(c_1 2^{-aL}\Big)^2 \\
&\leq \Big(\tfrac{1}{\sqrt{2}}\epsilon\Big)^2 \\
&= \frac{1}{2}\epsilon^2. \quad (1.4)
\end{aligned}$$

Ορίζουμε τώρα $M_l = \lceil 2\epsilon^{-2}(L+1)c_2 2^{-\gamma l} \rceil$ με αποτέλεσμα να λαμβανουμε την ακόλουθη ανισότητα:

$2\epsilon^{-2}(L+1)c_2 2^{-\gamma l} \leq M_l \Longrightarrow c_2 M_l^{-1} 2^{-\gamma l} \leq \frac{1}{2(L+1)}\epsilon^2. \quad (1.5)$

Για τη διασπορά του εκτιμητή I βλέπουμε πως:

$Var[I] \overset{\text{ανεξ.}}{=} \sum_{l=0}^{L} M_l^{-1} V_l \overset{(iii)}{\leq} \sum_{l=0}^{L} c_2 M_l^{-1} 2^{-\beta l} \overset{\beta\equiv\gamma}{=} \sum_{l=0}^{L} c_2 M_l^{-1} 2^{-\gamma l} \overset{(1.5)}{\leq} \sum_{l=0}^{L} \frac{1}{2(L+1)}\epsilon^2 = \frac{1}{2}\epsilon^2. \quad (1.6)$

Ισχύει τώρα ότι:

$$\begin{aligned} MSE = \mathbb{E}\left[(I-\mathbb{E}[G])^2\right] &= \mathbb{E}\left[I^2+\mathbb{E}[G]^2-2I\mathbb{E}[G]\right] \\ &= \mathbb{E}\left[I^2\right]-2\mathbb{E}\big[I\big]\mathbb{E}\big[G\big]+\mathbb{E}\big[G\big]^2 \\ &= \mathbb{E}\left[I^2\right]-\mathbb{E}\big[I\big]^2+\Big(\mathbb{E}\big[I\big]^2-2\mathbb{E}\big[I\big]\mathbb{E}\big[G\big]+\mathbb{E}\big[G\big]^2\Big) \\ &= Var\big[I\big]+\big(\mathbb{E}\big[I\big]-\mathbb{E}\big[G\big]\big)^2 \\ &\underset{(1.6)}{\overset{(1.4)}{\leq}} \frac{1}{2}\epsilon^2+\frac{1}{2}\epsilon^2 \\ &= \epsilon^2. \end{aligned}$$

Αυτό που απομένει να κάνουμε είναι να κατασκευάσουμε το επιθυμητό φράγμα για το υπολογιστικό κόστος $\mathcal{C}$. Έτσι λοιπόν βλέπουμε ότι:

$$\mathcal{C}=\sum_{l=0}^{L}C_lM_l \overset{(iv)}{\leq} \sum_{l=0}^{L}c_3M_l2^{\gamma l}.$$

Όμως εύκολα παίρνουμε πως $M_l<2\epsilon^{-2}(L+1)c_2 2^{-\gamma l}+1$, συνεπώς:

$$\mathcal{C}<\sum_{l=0}^{L}\left(c_3 2\epsilon^{-2}(L+1)c_2 2^{-\gamma l}2^{\gamma l}+c_3 2^{\gamma l}\right)=c_3\left(2\epsilon^{-2}(L+1)^2c_2+\sum_{l=0}^{L}2^{\gamma l}\right). \quad (1.7)$$

Παρατηρούμε τώρα πως ισχύει ότι:

- $L<\dfrac{\log(\sqrt{2}c_1\epsilon^{-1})}{a\log 2}+1 \Longrightarrow L<\dfrac{\log(\sqrt{2}c_1)}{a\log 2}+\dfrac{\log(\epsilon^{-1})}{a\log 2}+1 \Longrightarrow L+1<\dfrac{\log(\sqrt{2}c_1)}{a\log 2}+\dfrac{\log(\epsilon^{-1})}{a\log 2}+2$

Από την υπόθεσή μας όμως έχουμε πως:

$$\epsilon<e^{-1}\Longrightarrow \log\epsilon<\log e^{-1}\Longrightarrow \log\epsilon<-\log e\Longrightarrow \log\epsilon<-1\Longrightarrow -\log\epsilon>1\Longrightarrow 1<\log\epsilon^{-1},$$

οπότε:

$$L+1<\frac{\log(\sqrt{2}c_1)}{a\log 2}\log\epsilon^{-1}+\frac{\log(\epsilon^{-1})}{a\log 2}+2\log\epsilon^{-1}\leq\left\{\max\left(0,\frac{\log(\sqrt{2}c_1)}{a\log 2}+\frac{1}{a\log 2}+2\right)\right\}\log\epsilon^{-1}\equiv$$
$$c_5\log\epsilon^{-1}\Longrightarrow (L+1)^2\leq c_5^2\log^2\epsilon. \quad (1.8)$$

- $\sum_{l=0}^{L}2^{\gamma l}=\sum_{l=0}^{L}\left(2^L2^{l-L}\right)^{\gamma}=2^{\gamma L}\sum_{l=0}^{L}2^{-\gamma(L-l)}=2^{\gamma L}\sum_{l=0}^{L}2^{-\gamma l}<2^{\gamma L}\sum_{l=0}^{\infty}2^{-\gamma l}=2^{\gamma L}\frac{1}{1-2^{-\gamma}}=2^{\gamma L}\frac{2^{\gamma}}{2^{\gamma}-1}$

Δείξαμε προηγουμένως όμως ότι:

$2^{-a}\frac{1}{\sqrt{2}}\epsilon<c_1 2^{-aL}\Longrightarrow 2^{-aL}>\frac{1}{\sqrt{2}c_1}2^{-a}\epsilon\Longrightarrow 2^{\gamma L}<\left(\frac{1}{\sqrt{2}c_1}\right)^{-\gamma/a}2^{\gamma}\epsilon^{-\gamma/a}$, οπότε παίρνουμε:

$$\sum_{l=0}^{L}2^{\gamma l}<\left(\frac{1}{\sqrt{2}c_1}\right)^{-\gamma/a}\frac{2^{2\gamma}}{2^{\gamma}-1}\epsilon^{-\gamma/a}.$$

Επιπλέον, από την υπόθεσή μας ικανοποιείται η σχέση $a\geq\frac{1}{2}\min(\beta,\gamma)\overset{\beta=\gamma}{\Longrightarrow} a\geq\frac{1}{2}\gamma\Longrightarrow\frac{\gamma}{a}\leq 2\overset{\epsilon<1}{\Longrightarrow}$ $\epsilon^{-\gamma/a}\leq\epsilon^{-2}$, άρα τελικά:

$\sum_{l=0}^{L} 2^{\gamma l} < \left(\frac{1}{\sqrt{2}c_1}\right)^{-\gamma/a} \frac{2^{2\gamma}}{2^{\gamma}-1}\epsilon^{-2}. \quad (1.9)$

Με τη βοήθεια τώρα των σχέσεων (1.8) και (1.9) από τη σχέση (1.7) λαμβάνουμε ότι:

$\mathcal{C} \leq c_3 \left(2\epsilon^{-2}\log^2\epsilon\, c_2 c_5^2 + \frac{2^{2\gamma}}{2^{\gamma}-1}\left(\frac{1}{\sqrt{2}c_1}\right)^{-\gamma/a}\epsilon^{-2}\right)$ και αφού $\log\epsilon^{-1} > 1 \Longrightarrow -\log\epsilon > 1 \Longrightarrow \log^2\epsilon > 1$ έχουμε πως:

$\mathcal{C} \leq c_3 \left(2\epsilon^{-2}\log^2\epsilon\, c_2 c_5^2 + \frac{2^{2\gamma}}{2^{\gamma}-1}\left(\frac{1}{\sqrt{2}c_1}\right)^{-\gamma/a}\epsilon^{-2}\log^2\epsilon\right)$, οπότε:

$\mathcal{C} \leq c_4\epsilon^{-2}(\log\epsilon)^2$, με $c_4 = 2c_3c_2c_5^2 + \frac{2^{2\gamma}}{2^{\gamma}-1}\left(\frac{1}{\sqrt{2}c_1}\right)^{-\gamma/a}$, όπου ισχύει πως:

$c_5 = \left\{\max\left(0, \frac{\log(\sqrt{2}c_1)}{a\log 2} + \frac{1}{a\log 2} + 2\right)\right\}.$

β) β > γ

Ορίζουμε το L όπως και στην προηγούμενη περίπτωση, δηλαδή $L = \left\lceil \frac{\log(\sqrt{2}c_1\epsilon^{-1})}{a\log 2}\right\rceil$ με αποτέλεσμα να έχουμε τη σχέση:

$$2^{-a}\frac{1}{\sqrt{2}}\epsilon < c_1 2^{-aL} \leq \frac{1}{\sqrt{2}}\epsilon.$$

Τότε με τον τρόπο που μελετήσαμε στο (α) παίρνουμε πως ισχύει ότι $(\mathbb{E}[I] - \mathbb{E}[G])^2 \leq \frac{1}{2}\epsilon^2$.

Το σημείο διαφοροποίσης σε σχέση με την περίπτωση που εξετάσαμε εντοπίζεται στον ορισμό του πλήθους των προσομοιώσεων M_l. Συγκεκριμένα θεωρούμε $M_l = \left\lceil 2\epsilon^{-2}c_2(1 - 2^{-(\beta-\gamma)/2})^{-1}2^{-l(\beta+\gamma)/2}\right\rceil$ και εξ' ορισμού παίρνουμε πως:

$2\epsilon^{-2}c_2(1 - 2^{-(\beta-\gamma)/2})^{-1}\, 2^{-l(\beta+\gamma)/2} \leq M_l \Longrightarrow M_l^{-1} \leq \frac{1}{2}\epsilon^2\frac{1}{c_2}(1 - 2^{-(\beta-\gamma)/2})\, 2^{l(\beta+\gamma)/2}. \quad (1.10)$

Για τη διασπορά της εκτιμήτριας I ισχύει ότι:

$Var[I] \overset{\text{ανεξ.}}{=} \sum_{l=0}^{L} M_l^{-1}V_l \leq \sum_{l=0}^{L} c_2 M_l^{-1}2^{-\beta l} \overset{(1.10)}{\leq} \sum_{l=0}^{L}\frac{1}{2}\epsilon^2\frac{1}{c_2}c_2(1 - 2^{-(\beta-\gamma)/2})\, 2^{l(\beta+\gamma)/2}2^{-\beta l}$, άρα τελικά:

$Var[I] \leq \frac{1}{2}\epsilon^2(1 - 2^{-(\beta-\gamma)/2})\sum_{l=0}^{L} 2^{l(\gamma-\beta)/2}.$

Βλέπουμε όμως ότι:

$\sum_{l=0}^{L} 2^{l(\gamma-\beta)/2} = \sum_{l=0}^{L}\left(2^{-(\beta-\gamma)/2}\right)^l < \sum_{l=0}^{\infty}\left(2^{-(\beta-\gamma)/2}\right)^l = \left(1 - 2^{-(\beta-\gamma)/2}\right)^{-1}. \quad (1.11)$

Με τη βοήθεια λοιπόν της σχέσης (1.11) καταλήγουμε πως:

$Var[I] < \frac{1}{2}\epsilon^2\,\cancel{\left(1 - 2^{-(\beta-\gamma)/2}\right)}\,\cancel{\left(1 - 2^{-(\beta-\gamma)/2}\right)^{-1}} = \frac{1}{2}\epsilon^2.$

Έτσι, όπως δείξαμε στο πρώτο μέρος της απόδειξης ισχύει ότι:

$MSE = Var[I] + (\mathbb{E}[I] - \mathbb{E}[G])^2 < \frac{1}{2}\epsilon^2 + \frac{1}{2}\epsilon^2 = \epsilon^2.$

Επιπλέον εξ' ορισμού παίρνουμε πως:

$M_l < 2\epsilon^{-2}c_2(1 - 2^{-(\beta-\gamma)/2})^{-1}2^{-l(\beta+\gamma)/2} + 1. \quad (1.12)$

Για το υπολογιστικό έργο τώρα ισχύει ότι:

$\mathcal{C} = \sum_{l=0}^{L} C_l M_l \leq \sum_{l=0}^{L} c_3 2^{\gamma l} M_l \overset{(1.12)}{\leq} \sum_{l=0}^{L} \left(c_3 2\epsilon^{-2} c_2 (1 - 2^{-(\beta-\gamma)/2})^{-1} 2^{-l(\beta+\gamma)/2} 2^{\gamma l} + c_3 2^{\gamma l} \right)$, συνεπώς:

$\mathcal{C} \leq c_3 \left(2\epsilon^{-2} c_2 \left(1 - 2^{-(\beta-\gamma)/2}\right)^{-1} \sum_{l=0}^{L} 2^{l(\gamma-\beta)/2} + \sum_{l=0}^{L} 2^{\gamma l} \right). \quad (1.13)$

Από τη σχέση αυτή παρατηρούμε πως έχουμε ήδη βρει ένα άνω φράγμα για το πρώτο άθροισμα, ενώ για το δεύτερο θα στηριχτούμε σε ένα αποτέλεσμα που πήραμε από το πρώτο σκέλος. Συγκεκριμένα, εφόσον και σε αυτή την περίπτωση $\min(\beta, \gamma) = \gamma$ ισχύει ότι:

$\sum_{l=0}^{L} 2^{\gamma l} < \frac{2^{2\gamma}}{2^{\gamma}-1} (\sqrt{2} c_1)^{\gamma/a} \epsilon^{-2}. \quad (1.14)$

Αξίζει να αναφέρουμε πως οι πράξεις που χρησιμοποιήθηκαν είναι ακριβώς οι ίδιες με εκείνες της πρώτης περίπτωσης και για το λόγο αυτό αποφασίσαμε να τις παραλείψουμε. Το τελικό λοιπόν αποτέλεσμα, με αντικατάσταση των σχέσεων (1.11) και (1.14) στη (1.13), είναι το ακόλουθο:

$\mathcal{C} \leq c_3 \left(2\epsilon^{-2} c_2 (1 - 2^{-(\beta-\gamma)/2})^{-2} + \frac{2^{2\gamma}}{2^{\gamma}-1} (\sqrt{2} c_1)^{\gamma/a} \epsilon^{-2} \right) \equiv c_4 \epsilon^{-2}$, δηλαδή δείξαμε ότι:

$\mathcal{C} \leq c_4 \epsilon^{-2}$, όπου $c_4 = 2c_3 c_2 (1 - 2^{-(\beta-\gamma)/2})^{-2} + c_3 \frac{2^{2\gamma}}{2^{\gamma}-1} (\sqrt{2} c_1)^{\gamma/a}.$

γ) β < γ

Για μια ακόμα φορά ορίζουμε το L με τη βοήθεια του τύπου $L = \left\lceil \dfrac{\log(\sqrt{2} c_1 \epsilon^{-1})}{a \log 2} \right\rceil$ απ' όπου παίρνουμε ότι :

$2^{-a} \dfrac{1}{\sqrt{2}} \epsilon < c_1 2^{-aL} \leq \dfrac{1}{\sqrt{2}} \epsilon$ καθώς και $(\mathbb{E}[I] - \mathbb{E}[G])^2 \leq \dfrac{1}{2}\epsilon^2.$

Επιπλέον ορίζουμε το πλήθος των προσομοιώσεων M_l μέσα από την ακόλουθη σχέση:

$M_l = \left\lceil 2\epsilon^{-2} c_2 2^{L(\gamma-\beta)/2} \left(1 - 2^{-(\gamma-\beta)/2}\right)^{-1} 2^{-l(\beta+\gamma)/2} \right\rceil.$

Εξ' ορισμού τότε λαμβάνουμε ότι:

$2\epsilon^{-2} c_2 2^{L(\gamma-\beta)/2} \left(1 - 2^{-(\gamma-\beta)/2}\right)^{-1} 2^{-l(\beta+\gamma)/2} \leq M_l$ ή ισοδύναμα:

$M_l^{-1} \leq \frac{1}{2}\epsilon^2 \frac{1}{c_2} 2^{-L(\gamma-\beta)/2} \left(1 - 2^{-(\gamma-\beta)/2}\right) 2^{l(\beta+\gamma)/2}. \quad (1.15)$

Για τη διασπορά του εκτιμητή μας βλέπουμε πως:

$Var[I] \overset{\text{ανεξ.}}{=} \sum_{l=0}^{L} M_l^{-1} V_l \leq \sum_{l=0}^{L} c_2 M_l^{-1} 2^{-\beta l} \overset{(1.15)}{\leq} \sum_{l=0}^{L} \frac{1}{2}\epsilon^2 2^{-L(\gamma-\beta)/2}\left(1-2^{-(\gamma-\beta)/2}\right) 2^{l(\beta+\gamma)/2} 2^{-\beta l}$, άρα:
$Var[I] \leq \frac{1}{2}\epsilon^2 2^{-L(\gamma-\beta)/2}\left(1-2^{-(\gamma-\beta)/2}\right)\sum_{l=0}^{L} 2^{l(\gamma-\beta)/2}$.

Βλέπουμε όμως πως ισχύει ότι:

$$\begin{aligned}
\sum_{l=0}^{L} 2^{l(\gamma-\beta)/2} &= \sum_{l=0}^{L} 2^{L(\gamma-\beta)/2} 2^{(l-L)(\gamma-\beta)/2} \\
&= 2^{L(\gamma-\beta)/2} \sum_{l=0}^{L} 2^{-(L-l)(\gamma-\beta)/2} \\
&= 2^{L(\gamma-\beta)/2} \sum_{l=0}^{L} 2^{-l(\gamma-\beta)/2} \\
&< 2^{L(\gamma-\beta)/2} \sum_{l=0}^{\infty} \left(2^{-(\gamma-\beta)/2}\right)^l \\
&\overset{\beta\leq\gamma}{=} 2^{L(\gamma-\beta)/2}\left(1-2^{-(\gamma-\beta)/2}\right)^{-1}. \quad (1.16)
\end{aligned}$$

Έτσι για το φράγμα της διασποράς καταλήγουμε πως:

$Var[I] < \frac{1}{2}\epsilon^2 2^{-L(\gamma-\beta)/2}\left(1-2^{-(\gamma-\beta)/2}\right) 2^{L(\gamma-\beta)/2}\left(1-2^{-(\gamma-\beta)/2}\right)^{-1} = \frac{1}{2}\epsilon^2$.

Όπως έχουμε τώρα ήδη εξηγήσει ισχύει ότι:

$MSE = Var[I] + (\mathbb{E}[I]-\mathbb{E}[G])^2 < \frac{1}{2}\epsilon^2 + \frac{1}{2}\epsilon^2 = \epsilon^2$

Στη συνέχεια, μπορούμε εύκολα να πάρουμε την ανισότητα:

$M_l < 2\epsilon^{-2} c_2 2^{L(\gamma-\beta)/2}\left(1-2^{-(\gamma-\beta)/2}\right)^{-1} 2^{-l(\beta+\gamma)/2} + 1$, ενώ για το υπολογιστικό κόστος ισχύει ότι:

$\mathcal{C} = \sum_{l=0}^{L} C_l M_l \leq \sum_{l=0}^{L} c_3 2^{\gamma l} M_l < c_3\left(2\epsilon^{-2} c_2 2^{L(\gamma-\beta)/2}\left(1-2^{-(\gamma-\beta)/2}\right)^{-1}\sum_{l=0}^{L} 2^{l(\gamma-\beta)/2} + \sum_{l=0}^{L} 2^{\gamma l}\right)$.

Ακριβώς όπως και στην περίπτωση (β) έχει εντοπιστεί άνω φράγμα για το πρώτο άθροισμα με αποτέλεσμα από τη σχέση (1.16) να παίρνουμε ότι:

$\mathcal{C} < c_3\left(2\epsilon^{-2} c_2 2^{L(\gamma-\beta)}\left(1-2^{-(\gamma-\beta)/2}\right)^{-2} + \sum_{l=0}^{L} 2^{\gamma l}\right)$.

Προηγουμένως όμως είδαμε πως:

$2^{-a}\frac{1}{\sqrt{2}}\epsilon < c_1 2^{-aL} \Longrightarrow 2^{-aL} > \frac{1}{\sqrt{2}c_1}2^{-a}\epsilon \overset{\gamma\geq\beta}{\Longrightarrow} 2^{L(\gamma-\beta)} < (\sqrt{2}c_1)^{(\gamma-\beta)/a} 2^{\gamma-\beta}\epsilon^{-(\gamma-\beta)/a}$, οπότε:

$\mathcal{C} < c_3\left(2\epsilon^{-2} c_2 (\sqrt{2}c_1)^{(\gamma-\beta)/a} 2^{\gamma-\beta}\epsilon^{-(\gamma-\beta)/a}\left(1-2^{-(\gamma-\beta)/2}\right)^{-2} + \sum_{l=0}^{L} 2^{\gamma l}\right)$.

Το μόνο που απομένει για να ολοκληρωθεί η απόδειξη είναι να βρούμε ένα άνω φράγμα για το ά-θροισμα που εμφανίζεται στην τελευταία ανισοτική σχέση. Έτσι λοιπόν έχουμε ότι:

$$\sum_{l=0}^{L} 2^{\gamma l} = 2^{\gamma L}\sum_{l=0}^{L} 2^{-\gamma l} < 2^{\gamma L}\sum_{l=0}^{\infty} \left(2^{-\gamma}\right)^{l} = 2^{\gamma L}\frac{1}{1-2^{-\gamma}} = 2^{\gamma L}\frac{2^{\gamma}}{2^{\gamma}-1}.$$

Όμως αξιοποιώντας για μια ακόμα φορά τη σχέση (1.3) έχουμε και:

$$2^{-aL} > \frac{1}{\sqrt{2}c_1}2^{-a}\epsilon \Longrightarrow 2^{\gamma L} < (\sqrt{2}c_1)^{\gamma/a}2^{\gamma}\epsilon^{-\gamma/a}.$$

Επιπλέον σύμφωνα με τις βασικές μας υποθέσεις ισχύει ότι:

$a \geq \frac{1}{2}\min(\beta,\gamma) \Longrightarrow a \geq \frac{1}{2}\beta \Longrightarrow \frac{\beta}{a} \leq 2 \Longrightarrow \epsilon^{-\beta/a} \leq \epsilon^{-2} \Longrightarrow 1 \leq \epsilon^{-2}\epsilon^{\beta/a}$, επομένως:

$2^{\gamma L} < (\sqrt{2}c_1)^{\gamma/a}2^{\gamma}\epsilon^{-\gamma/a}\epsilon^{-2}\epsilon^{\beta/a} \Longrightarrow 2^{\gamma L} < (\sqrt{2}c_1)^{\gamma/a}2^{\gamma}\epsilon^{-2-(\gamma-\beta)/a}$, οπότε :

$\sum_{l=0}^{L} 2^{\gamma l} < (\sqrt{2}c_1)^{\gamma/a}\frac{2^{2\gamma}}{2^{\gamma}-1}\epsilon^{-2-(\gamma-\beta)/a}$ οδηγώντας τελικά στο αποτέλεσμα :

$\mathcal{C} \leq c_4\epsilon^{-2-(\gamma-\beta)/a}$, όπου $c_4 = 2c_2c_3(\sqrt{2}c_1)^{(\gamma-\beta)/a}2^{\gamma-\beta}\left(1-2^{-(\gamma-\beta)/2}\right)^{-2} + c_3(\sqrt{2}c_1)^{\gamma/a}\frac{2^{2\gamma}}{2^{\gamma}-1}.$

Με βάση τα παραπάνω βλέπουμε πως πήραμε το επιθυμητό άνω φράγμα και για τις τρεις περιπτώσεις γεγονός που μας οδηγεί στην ολοκλήρωση της συγκεκριμένης απόδειξης. □

<u>Παρατήρηση 1.3.2</u>

Αν επιθυμήσουμε να εξετάσουμε ποια επίπεδα συνεισφέρουν περισσότερο στην κατανομή της ολικής υπολογιστικής πολυπλοκότητας της μεθόδου Multilevel Monte Carlo, θα αρκούσε να βρούμε τα επίπεδα εκείνα που έχουν ίδια τάξη υπολογιστικού κόστους με αυτή του $\mathcal{C}$. Προφανώς, επειδή η τάξη του ολικού υπολογιστικού κόστους διαφοροποιείται ανάλογα με τη σχέση των παραμέτρων β, γ διακρίνουμε τις ακόλουθες περιπτώσεις:

1) $\beta > \gamma$

Ισχυριζόμαστε πως κάτω από τη συγκεκριμένη υπόθεση το κυρίαρχο υπολογιστικό κόστος εμφανίζεται στα χαμηλότερα επίπεδα l. Συγκεκριμένα, για $l = 0$ χρησιμοποιώντας τα δεδομένα μας αλλά και τον ορισμό του M_l που μας οδήγησε στη ζητούμενη ακρίβεια παίρνουμε ότι:

- $C_l \leq c_3 2^{\gamma l} \Longrightarrow C_0 \leq c_3 2^{\gamma \cdot 0} \Longrightarrow C_0 \leq c_3 \Longrightarrow C_0 = \mathcal{O}(1)$
- $M_l = \left\lceil 2\epsilon^{-2}c_2(1-2^{-(\beta-\gamma)/2})^{-1}2^{-l(\beta+\gamma)/2}\right\rceil \Longrightarrow M_0 = \left\lceil 2\epsilon^{-2}c_2(1-2^{-(\beta-\gamma)/2})^{-1}\right\rceil.$

Από την τελευταία σχέση όμως εύκολα παίρνουμε ότι:

$$M_0 < 2\epsilon^{-2}c_2(1-2^{-(\beta-\gamma)/2})^{-1} + 1 \overset{\epsilon\to 0}{\Longrightarrow} M_0 \leq \left(2c_2(1-2^{-(\beta-\gamma)/2})^{-1}+1\right)\epsilon^{-2} \Longrightarrow M_0 = \mathcal{O}(\epsilon^{-2}).$$

Βλέπουμε λοιπόν πως για την πολυπλοκότητα του πρώτου επιπέδου ισχύει ότι :

$\mathcal{C}_0 = C_0 M_0 = \mathcal{O}(1)\mathcal{O}(\epsilon^{-2}) = \mathcal{O}(\epsilon^{-2}),$

δηλαδή εμφανίζεται ίδιας τάξης υπολογιστική πολυπλοκότητα με αυτή που έχουμε για την ολική πολυπλοκότητα $\mathcal{C}$.

2) $\beta < \gamma$

Ισχυριζόμαστε πως για την περίπτωση αυτή το κυρίαρχο υπολογιστικό κόστος εμφανίζεται στα υψηλότερα επίπεδα l. Ειδικότερα, για $l = L$ αξιοποιώντας τη λογική που αναπτύξαμε προηγουμένως βλέπουμε ότι:

- $C_l \leq c_3 2^{\gamma l} \Longrightarrow C_L \leq c_3 2^{\gamma L}$

Στο σημείο αυτό θα πρέπει να εκφράσουμε την ποσότητα $2^{\gamma L}$ συναρτήσει του σφάλματος ϵ με σκοπό να προσδιορίσουμε την τάξη της πολυπλοκότητας C_L. Μελετώντας προσεκτικά την απόδειξή μας έχουμε πως:

$$L < \frac{\log(\sqrt{2} c_1 \epsilon^{-1})}{a \log 2} + 1 \Longrightarrow L < \log_2(\sqrt{2} c_1 \epsilon^{-1})^{1/a} + 1 \Longrightarrow 2^{\gamma L} < 2^{\gamma} (\sqrt{2} c_1)^{\gamma / a} \epsilon^{-\gamma / a}.$$

Με βάση λοιπόν το παραπάνω άνω φράγμα λαμβάνουμε ότι:

$$C_L \leq c_3 2^{\gamma L} \overset{\epsilon \to 0}{\Longrightarrow} C_L \leq c_3 2^{\gamma} (\sqrt{2} c_1)^{\gamma / a} \epsilon^{-\gamma / a} \Longrightarrow C_L = \mathcal{O}(\epsilon^{-\gamma / a}).$$

- $M_l = \left\lceil 2\epsilon^{-2} c_2 2^{L(\gamma - \beta)/2} \left(1 - 2^{-(\gamma - \beta)/2}\right)^{-1} 2^{-l(\beta + \gamma)/2} \right\rceil \Longrightarrow M_L = \left\lceil 2\epsilon^{-2} c_2 \left(1 - 2^{-(\gamma - \beta)/2}\right)^{-1} 2^{-\beta L} \right\rceil$

Για μια ακόμα φορά σύμφωνα με την απόδειξη που παραθέσαμε βλέπουμε πως ισχύει ότι:

$$\frac{\log(\sqrt{2} c_1 \epsilon^{-1})}{a \log 2} \leq L \Longrightarrow \log_2(\sqrt{2} c_1 \epsilon^{-1})^{1/a} \leq L \Longrightarrow 2^{-\beta L} \leq (\sqrt{2} c_1)^{-\beta / a} \epsilon^{\beta / a}.$$

Επιπλέον εύκολα προκύπτει ότι:

$M_L = \left\lceil 2\epsilon^{-2} c_2 \left(1 - 2^{-(\gamma - \beta)/2}\right)^{-1} 2^{-\beta L} \right\rceil < 2\epsilon^{-2} c_2 \left(1 - 2^{-(\gamma - \beta)/2}\right)^{-1} 2^{-\beta L} + 1$, συνεπώς για $\epsilon \to 0$ ισχύει πως:

$$M_L \leq 2 c_2 (\sqrt{2} c_1)^{-\beta / a} \left(1 - 2^{-(\gamma - \beta)/2}\right)^{-1} \epsilon^{-2 + \beta / a} + 1.$$

Δεν πρέπει όμως να ξεχνάμε πως από την υπόθεσή μας ικανοποιείται η συνθήκη $a \geq \frac{1}{2} \min(\beta, \gamma)$ ή διαφορετικά $a \geq \frac{1}{2}\beta$ μιας και κάτω από το πλαίσιο που εξετάζουμε ισχύει πως $\beta < \gamma$. Έτσι λοιπόν έχουμε ότι:

$a \geq \frac{1}{2}\beta \Longrightarrow \frac{\beta}{a} \leq 2 \Longrightarrow \epsilon^{\beta / a} \geq \epsilon^2 \Longrightarrow \epsilon^{-2 + \beta / a} \geq 1$ και καταλήγουμε πως:

$$M_L \leq \left(2 c_2 (\sqrt{2} c_1)^{-\beta / a} \left(1 - 2^{-(\gamma - \beta)/2}\right)^{-1} + 1\right) \epsilon^{-2 + \beta / a} \Longrightarrow M_L = \mathcal{O}(\epsilon^{-2 + \beta / a}).$$

Με βάση τα παραπάνω βλέπουμε πως για την πολυπλοκότητα του τελικού επιπέδου L ισχύει ότι:

$$\mathcal{C}_L = C_L M_L = \mathcal{O}(\epsilon^{-\gamma / a}) \mathcal{O}(\epsilon^{-2 + \beta / a}) = \mathcal{O}(\epsilon^{-2 - (\gamma - \beta)/a}),$$

δηλαδή εμφανίζεται ίδιας τάξης υπολογιστικό κόστος με αυτή του ολικού υπολογιστικού κόστους $\mathcal{C}$.

Παραμένοντας στο παρόν πλαίσιο, θα μελετήσουμε την ειδική περίπτωση όπου ικανοποιείται η σχέση $\beta = 2a$. Τότε, παρατηρούμε καταρχήν πως η ολική πολυπλοκότητα της μεθόδου φράσσεται ως εξής:

$$\mathcal{C} \leq c_4 \epsilon^{-2-(\gamma-\beta)/a} = c_4 \epsilon^{-2} \epsilon^{-\gamma/a} \epsilon^{\beta/a} = c_4 \epsilon^{-2} \epsilon^{-\gamma/a} \epsilon^{2} = c_4 \epsilon^{-\gamma/a} \Longrightarrow \mathcal{C} = \mathcal{O}(\epsilon^{-\gamma/a}),$$

ακριβώς όπως συμβαίνει και για το κόστος μιας αριθμητικής προσέγγισης στο τελικό επίπεδο μελέτης. Το αξιοσημείωτο είναι πως το συγκεκριμένο κόστος είναι βέλτιστο μιας και από τη σχέση $a \geq \frac{1}{2}\min(\beta,\gamma)$ λαμβάνουμε εύκολα πως ισχύει ότι:

$$\epsilon^{-\gamma/a} \leq \epsilon^{-2-(\gamma-\beta)/a} \ , \forall\ \beta \in \mathbb{R}_+ \text{ με } \beta < \gamma.$$

3) $\boldsymbol{\beta = \gamma}$

Για αυτή την περίπτωση η ύπαρξη του λογαρίθμου του σφάλματος μας δημιουργεί πρόβλημα στην εκτέλεση αναλυτικών πράξεων όπως πριν. Παρ' όλα αυτά σύμφωνα με τη δημοσίευση [1] ισχύει πως η ολική υπολογιστική πολυπλοκότητα διαμοιράζεται σχεδόν ισόποσα μεταξύ των διαφόρων επιπέδων.

1.4 Κανονικότητα Εκτιμητή Multilevel Monte Carlo

Στις ενότητες που προηγήθηκαν εξετάσαμε δυο βασικές αριθμητικές μεθόδους για την εκτίμηση αναμενόμενων τιμών: την απλή μέθοδο Monte Carlo και τη μέθοδο Multilevel Monte Carlo, την οποία και εισαγάγαμε με σκοπό τη μείωση της ολικής υπολογιστικής πολυπλοκότητας. Αξίζει να αναφέρουμε πως ο τρόπος με τον οποίο παρουσιάσαμε τη μέθοδο Multilevel Monte Carlo στη δεύτερη ενότητα έγινε με κύριο στόχο να την αναδείξουμε ως μια φυσιολογική γενίκευση της απλής μεθόδου Monte Carlo, όπως φαίνεται μέσα από τη μεταφορά βασικών ιδιοτήτων της δεύτερης. Βέβαια, ο προσεκτικός αναγνώστης θα έχει παρατηρήσει πως, στην μέχρι στιγμής μελέτη μας, υπάρχει μια ιδιότητα της αρχικής μεθόδου που δεν έχει μεταφερθεί στην πιο σύνθετη, η οποία δεν είναι άλλη από την κανονικότητα του Monte Carlo εκτιμητή (παρατήρηση 1.1.5). Ουσιαστικά, στην απλή μέθοδο Monte Carlo η ιδιότητα αυτή δεν είναι τίποτα παραπάνω από μια απλή εφαρμογή του κεντρικού οριακού θεωρήματος, το οποίο ως γνωστόν απαιτεί την ύπαρξη ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Αν τώρα κοιτάξουμε προσεκτικά το Multilevel Monte Carlo εκτιμητή μας διαπιστώνουμε πως απαρτίζεται από τυχαίες μεταβλητές που είναι ανεξάρτητες αλλά όχι ισόνομες, λόγω της ύπαρξης των διαφόρων επιπέδων, γεγονός που μας ωθεί να καταφύγουμε στην αξιοποίηση νέων μαθηματικών εργαλείων. Για να αντιμετωπίσουμε λοιπόν το πρόβλημα αυτό θα στηριχτούμε στο κεντρικό οριακό θεώρημα των Lindeberg και Feller το οποίο διατυπώνεται ως εξής:

Θεώρημα 1.4.1 (Lindeberg-Feller)

Έστω $(X_i)_{i\in\mathbb{N}}$ μια αριθμήσιμη οικογένεια ανεξάρτητων τυχαίων μεταβλητών (όχι κατ' ανάγκη ισόνομων) με $Var[X_i] < \infty$, $\forall\ i \in \mathbb{N}$. Θεωρούμε:

$$I_n = \sum_{i=1}^{n} X_i,$$

$$T_i = X_i - \mathbb{E}[X_i],\ i = 1, \ldots, n \ \text{ και}$$

$$s_n^2 = \sum_{i=1}^{n} \mathbb{E}\left[T_i^2\right].$$

Υποθέτουμε πως η ακόλουθη συνθήκη Lindeberg ικανοποιείται $\ \forall\ \epsilon > 0$:

$$\lim_{n\to\infty} s_n^{-2} \sum_{i=1}^{n} \mathbb{E}\left[T_i^2 \mathbb{1}_{|T_i|>\epsilon s_n}\right] = 0.$$

Τότε $\lim\limits_{n\to\infty} \mathbb{P}\left[\frac{I_n - \mathbb{E}[I_n]}{s_n} \leq z\right] = \Phi(z)$, όπου με Φ έχουμε συμβολίσει τη συνάρτηση κατανομής πιθανότητας μιας τυπικής κανονικής τυχαίας μεταβλητής.

Επιπλέον, στην προσπάθεια μας να δείξουμε την κανονικότητα του Multilevel Monte Carlo εκτιμητή θα κάνουμε χρήση μιας ταυτότητας που εμπεριέχεται στη δημοσίευση [2] και διατυπώνεται με τον ακόλουθο τρόπο:

Πρόταση 1.4.2

Έστω $\{a_l\}, \{b_l\}$ δυο μη αρνητικές και πεπερασμένες ακολουθίες ίσου μεγέθους και q ένας αυθαίρετος μη αρνητικός αριθμός. Τότε ισχύει η ακόλουθη ανισοτική σχέση:

$$\sum_l a_l^q b_l \leq \left(\sum_l a_l\right)^q \sum_l b_l.$$

Με βάση τα παραπάνω είμαστε έτοιμοι να μεταβούμε στο βασικό θεώρημα της συγκεκριμένης ενότητας το οποίο είναι το εξής:

Θεώρημα 1.4.3(Collier, Haji-Ali, Nobile, Von Schwerin, Tempone)

Έστω $G : \Omega \to \mathbb{R}$ μια μονοδιάστατη τυχαία μεταβλητή και $\overline{G}^l$ η αριθμητική της προσέγγιση στο l-οστό επίπεδο μελέτης. Υποθέτουμε επιπλέον πως οι τυχαίες μεταβλητές $\overline{G}^l - \overline{G}^{l-1}$ είναι μεταξύ τους ανεξάρτητες και ανεξάρτητες ως προς την τυχαία μεταβλητή $\overline{G}^0$. Ορίζουμε τη νέα μεταβλητή:

$$Y_l = \begin{cases} \left|\overline{G}^0 - \mathbb{E}\left[\overline{G}^0\right]\right|, & \text{όταν } l = 0 \\ \left|\overline{G}^l - \overline{G}^{l-1} - \mathbb{E}\left[\overline{G}^l - \overline{G}^{l-1}\right]\right|, & \text{όταν } l \geq 1 \end{cases}$$

και θεωρούμε πως πληρούνται οι συνθήκες:

- $c_1 \beta^{-q_3 l} \leq \mathbb{E}\left[Y_l^2\right]$, $\forall\, l \geq 0$
- $\mathbb{E}\left[Y_l^{2+\delta}\right] \leq c_2 \beta^{-\tau l}$, $\forall\, l \geq 0$

για κάποιο $\beta > 1$ και για αυστηρά θετικές σταθερές c_1, c_2, q_3, δ και τ. Επιλέγουμε το πλήθος των προσομοιώσεων σε κάθε επίπεδο ώστε για $q_2 > 0$ και για μια αυστηρά θετική ακολουθία $\{H_l\}_{l\geq 0}$ να ικανοποιείται η σχέση:

$$M_l \geq \beta^{-q_2 l} TOL^{-2} H_l^{-1} \left(\sum_{l=0}^{L} H_l\right), \forall\, l \geq 0,$$

όπου TOL είναι η παράμετρος σφάλματος της μεθόδου μας. Επιπλέον, επιλέγουμε τον αριθμό των επιπέδων L ώστε να ισχύει πως:

$L \leq \max\left(0, \frac{c \log(TOL^{-1})}{\log \beta} + C\right)$ για κάποιες σταθερές C, και $c > 0$.

Τέλος θετουμε $p = (1 + \delta/2)q_3 + (\delta/2)q_2 - \tau$. Αν ισχύει ότι είτε $p < 0$ είτε $p > 0$ με $c < \delta/p$, τότε :

$\lim_{TOL\to 0} \mathbb{P}\left[\frac{I-\mathbb{E}[I]}{\sqrt{Var[I]}} \leq z\right] = \Phi(z)$, όπου με I συμβολίζουμε το Multilevel Monte Carlo εκτιμητή μας.

Απόδειξη: Όπως ήδη αναφέραμε η απόδειξη του θεωρήματος αυτού θα στηριχτεί στο κεντρικό οριακό θεώρημα των Lindeberg και Feller ή αλλιώς στο θεώρημα 1.4.1. Έτσι, ορίζουμε καταρχάς τις ανεξάρτητες τυχαίες μεταβλητές $X_{l,j} := G_{l,j}/M_l$ με $l = 0, \ldots, L$ και $j = 1, \ldots, M_l$, όπου θεωρούμε πως:

$$G_{l,j} = \begin{cases} \overline{G}_j^l - \overline{G}_j^{l-1}, & \text{όταν } l \geq 1 \\ \overline{G}_j^0, & \text{όταν } l = 0 \end{cases}.$$

Τότε εύκολα βλέπουμε με βάση προγενέστερη ανάλυσή μας, πως ο εκτιμητής Multilevel Monte Carlo της ποσότητας $\mathbb{E}[G]$ δίνεται μέσω της σχέσης:

$$I = \frac{1}{M_0}\sum_{j=1}^{M_0} \overline{G}_j^0 + \sum_{l=1}^{L} \frac{1}{M_l}\sum_{j=1}^{M_l}\left(\overline{G}_j^l - \overline{G}_j^{l-1}\right) = \frac{1}{M_0}\sum_{j=1}^{M_0} G_{0,j} + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{j=1}^{M_l} G_{l,j} = \sum_{l=0}^{L}\sum_{j=1}^{M_l}\frac{G_{l,j}}{M_l}.$$

Κάνοντας χρήση τώρα του ορισμου της μεταβλητής X_l παίρνουμε ότι:

$$I = \sum_{l=0}^{L}\sum_{j=1}^{M_l}\frac{G_{l,j}}{M_l} = \sum_{l=0}^{L}\sum_{j=1}^{M_l} X_{l,j}. \quad (1.17)$$

Ορίζουμε στο σημείο αυτό για $l = 0, \ldots, L$ και $j = 1, \ldots, M_l$ τις νέες τυχαίες μεταβλητές:

$$T_{l,j} = X_{l,j} - \mathbb{E}[X_{l,j}] = \frac{G_{l,j}}{M_l} - \frac{\mathbb{E}[G_{l,j}]}{M_l} \equiv \frac{\Psi_{l,j}}{M_l}. \quad (1.18)$$

Τέλος θεωρούμε την ποσότητα s^2 η οποία δίνεται σύμφωνα με την ακόλουθη σχέση:

$$s^2 = \sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}[T_{l,j}^2] = \sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}\left[(X_{l,j} - \mathbb{E}[X_{l,j}])^2\right] = \sum_{l=0}^{L}\sum_{j=1}^{M_l} Var[X_{l,j}].$$

Όμως στην υπόθεσή μας θεωρήσαμε πως οι μεταβλητές διακριτοποίησης της μεταβλητής G είναι α-νεξάρτητες, γεγονός που μας οδηγεί στη σχέση:

$$s^2 = \sum_{l=0}^{L}\sum_{j=1}^{M_l} Var[X_{l,j}] = Var\left[\sum_{l=0}^{L}\sum_{j=1}^{M_l} X_{l,j}\right] \overset{(1.17)}{=} Var[I]. \quad (1.19)$$

Με βάση τα παραπάνω και αξιοποιώντας το θεώρημα των Lindeberg και Feller συμπεραίνουμε πως:

$$\lim_{TOL\to 0}\mathbb{P}\left[\frac{I-\mathbb{E}[I]}{\sqrt{Var[I]}} \leq z\right] = \Phi(z) \overset{(1.19)}{\Longleftrightarrow} \lim_{TOL\to 0}\mathbb{P}\left[\frac{I-\mathbb{E}[I]}{s} \leq z\right] = \Phi(z)$$

εφόσον ικανοποιείται $\forall\ \epsilon > 0$ η συνθήκη:

$$\lim_{TOL\to 0} s^{-2}\sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}\left[T_{l,j}^2 \mathbb{1}_{|T_{l,j}|>\epsilon s}\right] = 0 \underset{(1.19)}{\overset{(1.18)}{\Longleftrightarrow}} \lim_{TOL\to 0}\frac{1}{Var[I]}\sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}\left[\frac{\Psi_{l,j}^2}{M_l^2}\mathbb{1}_{\frac{|\Psi_{l,j}|}{M_l}>\epsilon\sqrt{Var[I]}}\right] = 0.$$

Από την υπόθεσή μας όμως μπορούμε εύκολα να δούμε πως $Y_{l,j} = |\Psi_{l,j}|$, οπότε αρκεί τελικά να δείξουμε ότι:

$$\lim_{TOL\to 0}\frac{1}{Var[I]}\sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}\left[\frac{Y_{l,j}^2}{M_l^2}\mathbb{1}_{\frac{Y_{l,j}}{M_l}>\epsilon\sqrt{Var[I]}}\right] = 0,\ \forall\ \epsilon > 0. \quad (1.20)$$

Έχοντας διασαφηνίσει τα παραπάνω προχωράμε τώρα στο κύριο μέρος της απόδειξής μας, όπου καλούμαστε να ικανοποιήσουμε τη σχέση (1.20). Συγκεκριμένα, βλέπουμε ότι:

$$\frac{1}{Var[I]}\sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}\left[\frac{Y_{l,j}^2}{M_l^2}\mathbb{1}_{\frac{Y_{l,j}}{M_l}>\epsilon\sqrt{Var[I]}}\right]=\frac{1}{Var[I]}\sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}\left[\frac{Y_{l,j}^2}{M_l^2}\mathbb{1}_{Y_{l,j}>\epsilon\sqrt{Var[I]}M_l}\right]\equiv F.$$

Αυτό που θα επιχειρήσουμε να κάνουμε τώρα είναι να γράψουμε το ενδεχόμενο που βρίσκεται εντός της δείκτριας με έναν διαφορετικό, αλλά πλήρως ισοδύναμο τρόπο. Συγκεκριμένα, παρατηρούμε πως ισχύει ότι:

$Y_{l,j}>\epsilon\sqrt{Var[I]}M_l \overset{\delta>0}{\iff} Y_{l,j}^{\delta}>\epsilon^{\delta}Var[I]^{\delta/2}M_l^{\delta} \iff Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta}>1$, άρα:

$$\mathbb{1}_{Y_{l,j}>\epsilon\sqrt{Var[I]}M_l}=\mathbb{1}_{Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta}>1}. \quad (1.21)$$

Θα δείξουμε τώρα πως για την αρχική μας δείκτρια ισχύει η ανισοτική σχέση:

$$\mathbb{1}_{Y_{l,j}>\epsilon\sqrt{Var[I]}M_l}\le Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta},$$

διακρίνοντας τις ακόλουθες περιπτώσεις:

(α) $Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta}\le 1$

Κάτω από την υπόθεση αυτή και με τη βοήθεια της σχέσης (1.21) εύκολα παίρνουμε ότι:

$$\mathbb{1}_{Y_{l,j}>\epsilon\sqrt{Var[I]}M_l}=0\le Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta}.$$

(β) $Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta}>1$

Για μια ακόμα φορά αξιοποιώντας την παραπάνω συνθήκη και τη σχέση (1.21) λαμβάνουμε ότι:

$$\mathbb{1}_{Y_{l,j}>\epsilon\sqrt{Var[I]}M_l}=1< Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta}.$$

Δείξαμε λοιπόν πως σε κάθε περίπτωση ισχύει η ζητούμενη σχέση η οποία μας οδηγεί στο ακόλουθο αποτέλεσμα:

$$\mathbb{E}\left[\frac{Y_{l,j}^2}{M_l^2}\mathbb{1}_{Y_{l,j}>\epsilon\sqrt{Var[I]}M_l}\right]\le\mathbb{E}\left[\frac{Y_{l,j}^2}{M_l^2}Y_{l,j}^{\delta}\,\epsilon^{-\delta}Var[I]^{-\delta/2}M_l^{-\delta}\right]=\epsilon^{-\delta}\frac{M_l^{-2-\delta}}{Var[I]^{\delta/2}}\mathbb{E}\left[Y_{l,j}^{\delta+2}\right].$$

Με βάση το παραπάνω άνω φράγμα για την ποσότητα F βλέπουμε πως ισχύει ότι:

$F=\frac{1}{Var[I]}\sum_{l=0}^{L}\sum_{j=1}^{M_l}\mathbb{E}\left[\frac{Y_{l,j}^2}{M_l^2}\mathbb{1}_{Y_{l,j}>\epsilon\sqrt{Var[I]}M_l}\right]\overset{\text{ισον.}}{\le}\frac{\epsilon^{-\delta}}{Var[I]^{1+\delta/2}}\sum_{l=0}^{L}M_l^{-2-\delta}\sum_{j=1}^{M_l}\mathbb{E}\left[Y_l^{\delta+2}\right]$, επομένως:

$$F\le\frac{\epsilon^{-\delta}}{Var[I]^{1+\delta/2}}\sum_{l=0}^{L}M_l^{-1-\delta}\mathbb{E}\left[Y_l^{\delta+2}\right].$$

Αν τώρα θεωρήσουμε τη μεταβλητή V_l με:

$$V_l=\begin{cases} Var\left[\overline{G}^l-\overline{G}^{l-1}\right], & \text{όταν } l\ge 1\\ Var\left[\overline{G}^0\right], & \text{όταν } l=0\end{cases},$$

βλέπουμε πως το άνω φράγμα της ποσότητας F μπορεί να γραφτεί ως εξής:

$F \leq \frac{\epsilon^{-\delta}}{Var[I]^{1+\delta/2}} \sum_{l=0}^{L} M_l^{-1-\delta/2} M_l^{-\delta/2} V_l^{1+\delta/2} V_l^{-1-\delta/2} \mathbb{E}\left[Y_l^{\delta+2}\right]$ ή ισοδύναμα:

$F \leq \frac{\epsilon^{-\delta}}{Var[I]^{1+\delta/2}} \sum_{l=0}^{L} \left(M_l^{-1} V_l\right)^{1+\delta/2} \left(V_l^{-1-\delta/2} M_l^{-\delta/2}\right) \mathbb{E}\left[Y_l^{\delta+2}\right]. \quad (1.22)$

Από το σημείο αυτό και ύστερα θα κάνουμε εκτεταμένη χρήση της πρότασης 1.4.2. Έτσι βλέπουμε καταρχήν πως από τη σχέση (1.22) λαμβάνουμε ότι:

$$F \leq \frac{\epsilon^{-\delta}}{Var[I]^{1+\delta/2}} \left(\sum_{l=0}^{L} M_l^{-1} V_l\right)^{1+\delta/2} \sum_{l=0}^{L} V_l^{-1-\delta/2} M_l^{-\delta/2} \mathbb{E}\left[Y_l^{\delta+2}\right].$$

Σε προηγούμενη όμως ανάλυσή μας δείξαμε πως για τη μεταβλητή V_l, όταν αυτή είναι ορισμένη όπως εδώ, ισχύει ότι:

$$Var[I] = \sum_{l=0}^{L} M_l^{-1} V_l,$$

επομένως παίρνουμε τελικά πως:

$$F \leq \epsilon^{-\delta} \sum_{l=0}^{L} V_l^{-1-\delta/2} M_l^{-\delta/2} \mathbb{E}\left[Y_l^{\delta+2}\right].$$

Από την υπόθεσή μας όμως για το πλήθος του δείγματος ισχύει ότι:

$$M_l \geq \beta^{-q_2 l} TOL^{-2} H_l^{-1} \left(\sum_{l=0}^{L} H_l\right) \Longrightarrow M_l^{-\delta/2} \leq \beta^{\delta q_2 l/2} TOL^{\delta} H_l^{\delta/2} \left(\sum_{l=0}^{L} H_l\right)^{-\delta/2}.$$

Συνδυάζοντας τα παραπάνω έχουμε πως:

$$\begin{aligned} F &\leq \epsilon^{-\delta} \sum_{l=0}^{L} V_l^{-1-\delta/2} M_l^{-\delta/2} \mathbb{E}\left[Y_l^{\delta+2}\right] \\ &\leq \epsilon^{-\delta} \sum_{l=0}^{L} V_l^{-1-\delta/2} \beta^{\delta q_2 l/2} TOL^{\delta} H_l^{\delta/2} \left(\sum_{l=0}^{L} H_l\right)^{-\delta/2} \mathbb{E}\left[Y_l^{\delta+2}\right] \\ &= \epsilon^{-\delta} TOL^{\delta} \left(\sum_{l=0}^{L} H_l\right)^{-\delta/2} \sum_{l=0}^{L} H_l^{\delta/2} \left(V_l^{-1-\delta/2} \beta^{\delta q_2 l/2} \mathbb{E}\left[Y_l^{\delta+2}\right]\right) \\ &\overset{1.4.2}{\leq} \epsilon^{-\delta} TOL^{\delta} \left(\sum_{l=0}^{L} H_l\right)^{-\delta/2} \left(\sum_{l=0}^{L} H_l\right)^{\delta/2} \sum_{l=0}^{L} V_l^{-1-\delta/2} \beta^{\delta q_2 l/2} \mathbb{E}\left[Y_l^{\delta+2}\right] \\ &= \epsilon^{-\delta} TOL^{\delta} \sum_{l=0}^{L} V_l^{-1-\delta/2} \beta^{\delta q_2 l/2} \mathbb{E}\left[Y_l^{\delta+2}\right]. \quad (1.23) \end{aligned}$$

Παρατηρούμε όμως πως ισχύουν τα ακόλουθα:

- $c_1 \beta^{-q_3 l} \leq \mathbb{E}\left[Y_l^2\right] \Longrightarrow c_1 \beta^{-q_3 l} \leq \mathbb{E}\left[\Psi_l^2\right] \Longrightarrow c_1 \beta^{-q_3 l} \leq \mathbb{E}\left[(G_l - \mathbb{E}[G_l])^2\right]$, ή αλλιώς:

$c_1 \beta^{-q_3 l} \leq Var[G_l] \overset{\text{ορ.}}{\Longrightarrow} c_1 \beta^{-q_3 l} \leq V_l \Longrightarrow V_l^{-1-\delta/2} \leq c_1^{-1-\delta/2} \beta^{(1+\delta/2) q_3 l} \quad (1.24)$

- $\mathbb{E}\left[Y_l^{\delta+2}\right] \leq c_2 \beta^{-\tau l}. \quad (1.25)$

Η ανισοτική σχέση (1.23) με τη βοήθεια των σχέσεων (1.24) και (1.25) αποκτά τελικά την παρακάτω μορφή:

$$F \leq \epsilon^{-\delta} TOL^{\delta} \sum_{l=0}^{L} c_1^{-1-\delta/2} \beta^{(1+\delta/2)q_3 l} \beta^{\delta q_2 l/2} c_2 \beta^{-\tau l} = \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \sum_{l=0}^{L} \beta^{\{(1+\delta/2)q_3+(\delta/2)q_2-\tau\}l}.$$

Αξιοποιώντας τώρα τον ορισμό της ποσότητας p που εμπεριέχεται στα δεδομένα μας καταλήγουμε πως:

$$F \leq \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \sum_{l=0}^{L} \beta^{\,pl} = \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \sum_{l=0}^{L} (\beta^{\,p})^l.$$

Για να ολοκληρώσουμε την απόδειξή μας θα διακρίνουμε τις ακόλουθες περιπτώσεις:

(1) $\mathbf{p < 0}$

Βλέπουμε πως κάτω από αυτή την υπόθεση ισχύει ότι:

$\sum_{l=0}^{L} (\beta^{\,p})^l < \sum_{l=0}^{\infty} (\beta^{\,p})^l \overset{\beta \geq 1}{\underset{p<0}{=}} \frac{1}{1-\beta^{\,p}}$, συνεπώς προκύπτει πως:

$$F \leq \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \sum_{l=0}^{L} (\beta^{\,p})^l < \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \left(\frac{1}{1-\beta^{\,p}}\right).$$

Παρατηρούμε τώρα πως εξ' ορισμού η ποσότητα F είναι μη αρνητική και φράσσεται από έναν όρο που τείνει στο 0 καθώς $TOL \to 0$, συνεπώς $\lim_{TOL \to 0} F = 0$ και λαμβάνουμε τη ζητούμενη σχέση.

(2) $\mathbf{p > 0}$

Χρησιμοποιώντας τον τύπο που μας δίνει πεπερασμένο άθροισμα όρων γεωμετρικής προόδου έχουμε ότι:

$$F \leq \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \sum_{l=0}^{L} (\beta^{\,p})^l = \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \left(\frac{\beta^{\,(L+1)p} - 1}{\beta^{\,p} - 1}\right). \quad (1.26)$$

Κάτω από το πλαίσιο που μελετάμε τώρα δεν μπορούμε να απαλλαγούμε από το L, αλλά μπορούμε να στηριχτούμε στη γνώση που έχουμε για το άνω φράγμα του. Συγκεκριμένα, για την παράμετρο σφάλματος που θεωρήσαμε έχουμε τις ακόλουθες περιπτώσεις:

- $\frac{c \log(TOL^{-1})}{\log \beta} + C \leq 0$

Για τη συνθήκη αυτή με βάση τα δεδομένα μας έχουμε πως:

$L \leq \max\left(0, \frac{c \log(TOL^{-1})}{\log \beta} + C\right) \Longrightarrow L \leq 0$ και εύκολα λαμβάνουμε ότι:

$$F \leq \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \left(\frac{\beta^{\,(L+1)p} - 1}{\beta^{\,p} - 1}\right) \leq \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \left(\frac{\cancel{\beta^{\,p} - 1}}{\cancel{\beta^{\,p} - 1}}\right) = \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2}.$$

Έτσι, από την τελευταία σχέση βλέπουμε πως $\lim_{TOL \to 0} F = 0$ για το λόγο που εξηγήσαμε προηγουμένως.

- $\frac{c \log(TOL^{-1})}{\log \beta} + C > 0$

Για αυτή την περίπτωση τώρα ισχύει πως:

$$L \leq \max\left(0, \frac{c\log(TOL^{-1})}{\log\beta} + C\right) \Longrightarrow L \leq \frac{\log(TOL^{-c})}{\log\beta} + C \Longrightarrow L \leq \log_\beta(TOL^{-c}) + C.$$

Εφαρμόζοντας το συγκεκριμένο φράγμα στη σχέση (1.26) παίρνουμε ότι:

$$F \leq \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \left(\frac{\beta^{\,p\log_\beta(TOL^{-c})+p(C+1)}-1}{\beta^{\,p}-1}\right) = \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \left(\frac{TOL^{-cp}\beta^{\,(C+1)p}-1}{\beta^{\,p}-1}\right).$$

Τελικά λοιπόν δείξαμε ότι:

$$F \leq \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} \left(\frac{TOL^{-cp}\beta^{\,(C+1)p}-1}{\beta^{\,p}-1}\right) < \epsilon^{-\delta} TOL^{\delta} c_2 c_1^{-1-\delta/2} TOL^{-cp} \left(\frac{\beta^{\,(C+1)p}}{\beta^{\,p}-1}\right).$$

Αν στο σημείο αυτό θέσουμε $\widetilde{C} = \epsilon^{-\delta} c_2 c_1^{-1-\delta/2} \left(\frac{\beta^{\,(C+1)p}}{\beta^{\,p}-1}\right)$ παρατηρούμε ότι:

$$F \leq \widetilde{C} TOL^{\delta - cp}.$$

Προφανώς, τώρα για να ισχύει πως $\lim_{TOL\to 0} F = 0$ θα πρέπει να ικανοποιείται ότι:

$$\lim_{TOL\to 0} \widetilde{C} TOL^{\delta - cp} = 0 \ \text{ ή ισοδύναμα } \ \delta - cp > 0 \overset{p>0}{\Longrightarrow} c < \frac{\delta}{p}.$$

Δείξαμε λοιπόν πως για $p > 0$ ισχύει η ζητούμενη ιδιότητα μόνο εφόσον ικανοποιείται η σχέση $c < \frac{\delta}{p}$.

Με βάση τα παραπάνω συμπεραίνουμε τελικά πως η ζητούμενη σχέση ικανοποιείται όταν $p < 0$ ή όταν $p > 0$ με $c < \frac{\delta}{p}$ και η απόδειξη έχει ολοκληρωθεί. $\square$

2 Εμφωλευμένη Διαδικασία Προσομοίωσης

2.1 Πιθανότητα Απώλειας Χαρτοφυλακίου

Έστω πως επιθυμούμε να μετρήσουμε το ρίσκο ενός χαρτοφυλακίου σε κάποιο μελλοντικό χρόνο T, από την οπτική γωνία ενός παρατηρητή στον αρχικό χρόνο $t = 0$. Με τον όρο χαρτοφυλάκιο εννοούμε ένα σύνολο χρεογράφων (μετοχές, ομόλογα κ.λ.π) αξίας S^i, όπου $i = 0, \dots, d$, τα οποία περιέχονται σε ξεχωριστές μεταξύ τους ποσότητες ξ^i, ορισμένες μέσω μιας διαδικασίας που καλείται επενδυτική στρατηγική. Κάτω από το πλαίσιο αυτό μπορούμε να ορίσουμε την ολική αξία του χαρτοφυλακίου στο χρόνο t μέσω της σχέσης:

$$V_t = \sum_{i=0}^{d} \xi_t^i S_t^i,\ \forall\ t \in [0, T].$$

Υποθέτουμε λοιπόν πως στον αρχικό χρόνο παρατήρησης το χαρτοφυλάκιό μας έχει αξία V_0 και για το χρόνο T θεωρούμε την ποσότητα $L = V_0 - V_T$. Από τη σχέση αυτή γίνεται φανερό πως στην περίπτωση που η τελική αξία χαρτοφυλακίου είναι μικρότερη της αρχικής το L λαμβάνει θετική τιμή και θα μπορούσε να ερμηνευθεί ως η απώλεια στην αξία του χαρτοφυλακίου. Επιπλέον, εύκολα αντιλαμβανόμαστε πως η τιμή ενός χρεογράφου θα μπορούσε να επηρεαστεί από πολλούς παράγοντες/σενάρια με αποτέλεσμα να θεωρούμε πως η κάθε τιμή S_t^i αναπαριστά μια τυχαία μεταβλητή από το χώρο των παραπάνω σεναρίων Ω στους θετικούς πραγματικούς αριθμούς. Έτσι για την ποσότητα που ορίσαμε παραπάνω θα ισχύει ότι $L : \Omega \to \mathbb{R}$ ή με άλλα λόγια $L(\omega) = V_0 - V_T(\omega)$, $\forall\ \omega \in \Omega$, όπου εξ' ορισμού ισχύει πως:

$$V_T(\omega) = \sum_{i=0}^{d} \xi_T^i S_T^i(\omega).$$

Για να ποσοτικοποιήσουμε τώρα την αβεβαιότητά μας γύρω από την τυχαία μεταβλητή L στηριζόμαστε σε μέτρα ρίσκου ρ, τα οποία δεν είναι τίποτα άλλο από συναρτησιακά με τιμές στον $\mathbb{R}$, δηλαδή $\rho : L \to \mathbb{R}$. Το βασικό μέτρο ρίσκου με το οποίο θα εργαστούμε είναι η πιθανότητα μιας μεγάλης απώλειας. Συγκεκριμένα, για ένα φράγμα $c \in \mathbb{R}$ ενδιαφερόμαστε για την πιθανότητα η μεταβλητή L να ξεπεράσει τη συγκεκριμένη τιμή, δηλαδή ενδιαφερόμαστε για την ποσότητα $\alpha = \alpha(L) = \mathbb{P}[L \geq c]$.

Αριθμητική εκτίμηση της πιθανότητας απώλειας

Λόγω του γεγονότος πως ο χώρος των ενδεχομένων Ω είναι πολύ μεγάλος ή και άπειρος καθίσταται δύσκολο να εργαστούμε με την κατανομή της μεταβλητής L και θα πρέπει να στηριχτούμε σε μια εμπειρική κατανομή η οποία λαμβάνεται μέσω προσομοίωσης Monte Carlo. Ειδικότερα, από βασικές ιδιότητες της δείκτριας συνάρτησης παίρνουμε ότι $\mathbb{P}[L \geq c] = \mathbb{P}[L(\omega) \geq c] = \mathbb{E}\left[\mathbb{1}_{L(\omega) \geq c}\right]$, άρα με βάση προγενέστερη ανάλυσή μας βλέπουμε πως:

$$\frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{L(\omega_i) \geq c} \xrightarrow[M \to \infty]{\mathbb{P}} \mathbb{E}\left[\mathbb{1}_{L(\omega) \geq c}\right] = \mathbb{P}[L \geq c].$$

Ωστόσο ο εκτιμητής που ορίσαμε για την προσέγγιση της ζητούμενης μέσης τιμής εμφανίζει ένα βασικό πρόβλημα. Ουσιαστικά, αυτό που κάνει ο συγκεκριμένος εκτιμητής είναι να υπολογίζει την απώλεια χαρτοφυλακίου για διάφορα σενάρια και εν συνεχεία να εξετάζει κατά πόσο η τιμή αυτή ξεπερνά το επιθυμητό φράγμα c. Παρ' όλα αυτά, ακόμα και κάτω από ένα πλήρως καθορισμένο σενάριο ω_i είναι δύσκολο να υπολογίσουμε αναλυτικά την τιμή $L(\omega_i)$ εξαιτίας κάποιας πολυπλοκότητας/αβεβαιότητας που πηγάζει από τη φύση των χρεογράφων μας. Συνεπώς, θα χρειαστούμε ακόμα μια αριθμητική μέθοδο Monte Carlo η οποία θα εκτιμά την απώλεια L του χαρτοφυλακίου κάτω από ένα συγκεκριμένο σενάριο ω_i. Για το σκοπό αυτό, κάτω από το σενάριο ω_i, θεωρούμε το τυχαίο δείγμα $\overline{z}_1^i, \dots, \overline{z}_N^i$ από ανεξάρτητες και ισόνομες μεταβλητές απώλειας που προσεγγίζουν την τιμή $L(\omega_i)$ μέσω της σχέσης:

$\overline{L}_i = \frac{1}{N}\sum_{j=1}^{N} \overline{z}_j^i.$

Επιπλέον, οι μεταβλητές προσέγγισης $\overline{z}_j^i$, δοθέντος του σεναρίου ω_i, θα φέρουν μέση τιμή ίση με $L(\omega_i)$ ώστε ο εκτιμητής $\overline{L}_i$ να είναι αμερόληπτος. Συνοπτικά λοιπόν για την εύρεση της ζητούμενης ποσότητας χρησιμοποιήθηκαν δύο σχήματα Monte Carlo: ένα που για συγκεκριμένο σενάριο ω_i εκτιμά την τιμή $L(\omega_i)$ και καλείται εσωτερικό επίπεδο της προσομοίωσης και ένα που προσεγγίζει την πιθανότητα $\mathbb{P}\left[L(\omega) \geq c\right]$, στηριζόμενο στα αποτελέσματα του προηγούμενου σχήματος, το οποίο καλείται εξωτερικό επίπεδο της προσομοίωσης. Ετσι, αν θέλαμε να γράψουμε αναλυτικά τον εκτιμητή που χρησιμοποιούμε για το ζητούμενο υπολογισμό, αυτός θα ήταν ο:

$$I := \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\overline{L}_i \geq c} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\frac{1}{N}\sum_{j=1}^{N} \overline{z}_j^i \geq c}.$$

Η διαδικασία που περιγράψαμε παραπάνω καλείται εμφωλευμένη διαδικασία προσομοίωσης. Τέλος, αξίζει να αναφέρουμε πως η παραπάνω προσομοίωση είναι ομοιόμορφη υπό την έννοια πως για το κάθε εξεταζόμενο σενάριο ω_i χρησιμοποιείται σταθερό πλήθος εσωτερικού δείγματος N.

2.2 Βέλτιστη Επιλογή Παραμέτρων

Όπως είδαμε λεπτομερώς στην προηγούμενη ενότητα, ο εκτιμητής που χρησιμοποιήσαμε για τον προσδιορισμό της ζητούμενης πιθανότητας είναι συνάρτηση δύο παραμέτρων: του αριθμού των σεναρίων ω_i που συμβολίσαμε ως M και του πλήθους των εσωτρικών προσομοιώσεων σε κάθε σενάριο που συμβολίσαμε ως N. Υπό το πλαίσιο αυτό, θα ήταν χρήσιμο να αναζητήσουμε τις βέλτιστες τιμές των παραπάνω παραμέτρων ή με άλλα λόγια τις τιμές εκείνες που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα MSE, μιας και το τελευταίο αποτελεί δείκτη ακρίβειας της μεθόδου μας. Για να δείξουμε την εξάρτηση του εκτιμητή μας από τις παραμέτρους που περιγράψαμε θα χρησιμοποιήσουμε το συμβολισμό $I \equiv \alpha_{M,N}$, ενώ όπως και πριν η ζητούμενη πιθανότητα θα συμβολιστεί ως α. Με βάση τα παραπάνω γίνεται αντιληπτό πως για τον υπολογισμό της ποσότητας $\alpha_{M,N}$ απαιτούνται $M.N$ συνολικά προσομοιώσεις. Έτσι, αν επιβάλλουμε και τη συνθήκη το ολικό υπολογιστικό μας έργο να μην ξεπερνά ένα φράγμα $k > 0$ μπορούμε να ορίσουμε το ακόλουθο πρόβλημα ελαχιστοποίησης υπό περιορισμό:

$$\begin{cases} \min\limits_{M,N} MSE \\ M.N = k \\ M, N \geq 0 \end{cases} \iff \begin{cases} \min\limits_{M,N} \mathbb{E}\left[(\alpha_{M,N} - \alpha)^2\right] \\ M.N = k \\ M, N \geq 0 \end{cases}.$$

Αυτό που θα εξετάσουμε σε ένα πρώτο στάδιο είναι η ασυμπτωτική συμπεριφορά του μέσου τετραγωνικού σφάλματος, δηλαδή η συμπεριφορά του καθώς $M, N \to \infty$. Για να πετύχουμε το παραπάνω βλέπουμε καταρχάς πως ισχύει ότι:

$$\begin{aligned} MSE &= \mathbb{E}\left[(\alpha_{M,N} - \alpha)^2\right] \\ &= \mathbb{E}\left[(\alpha_{M,N} - \mathbb{E}\left[\alpha_{M,N}\right] + \mathbb{E}\left[\alpha_{M,N}\right] - \alpha)^2\right] \\ &= \mathbb{E}\left[(\alpha_{M,N} - \mathbb{E}\left[\alpha_{M,N}\right])^2\right] + (\mathbb{E}\left[\alpha_{M,N}\right] - \alpha)^2 \\ &= Var\left[\alpha_{M,N}\right] + (\mathbb{E}\left[\alpha_{M,N}\right] - \alpha)^2. \quad (2.1) \end{aligned}$$

Αυτό που καταφέραμε λοιπόν είναι να σπάσουμε το MSE στο άθροισμα της διασποράς του εκτιμητή μας με έναν άλλο όρο που δεν είναι τίποτα άλλο από τη μεροληψία του υψωμένη στο τετράγωνο.

Στο σημείο αυτό θα εξετάσουμε ξεχωριστά την ασυμπτωτική συμπεριφορά που εμφανίζει ο κάθε ένας από τους όρους της σχέσης (2.1). Για να ξεκινήσουμε τη μελέτη μας θα πρέπει να στηριχτούμε στην ακόλουθη αρκετά τεχνική υπόθεση:

Υπόθεση 2.2.1

Θεωρούμε $L(\omega)$ την απώλεια ενός χαρτοφυλακίου κάτω από το ενδεχόμενο ω στο μελλοντικό χρόνο T και $\overline{L}(\omega)$ την αριθμητική της προσέγγιση μέσω του σχήματος Monte Carlo για το συγκεκριμένο σενάριο. Με βάση την παραδοχή πως για τη συγκεκριμένη εκτίμηση χρησιμοποιήθηκαν N ισόνομες και ανεξάρτητες τυχαίες μεταβλητές που προσεγγίζουν την τιμή $L(\omega)$ υποθέτουμε ότι:

1) Η από κοινού συνάρτηση κατανομής πιθανότητας $p_N(l,\bar{l})$ των $L, \overline{L}$ και οι μερικές της παράγωγοι $\frac{\partial}{\partial l}p_N(l,\bar{l})$ και $\frac{\partial^2}{\partial l^2}p_N(l,\bar{l})$ υπάρχουν $\forall\ N \in \mathbb{N}$ και για κάθε ζεύγος $(l,\bar{l})$

2) Υπάρχουν $\forall\ N \geq 1$ συναρτήσεις $f_{0,N}(.)$, $f_{1,N}(.)$ και $f_{2,N}(.)$ τέτοιες ώστε:

$p_N(l,\bar{l}) \leq f_{0,N}(\bar{l}), \quad \left|\frac{\partial}{\partial l}p_N(l,\bar{l})\right| \leq f_{1,N}(\bar{l})$ και $\left|\frac{\partial^2}{\partial l^2}p_N(l,\bar{l})\right| \leq f_{2,N}(\bar{l}),$

για κάθε ζεύγος $(l,\bar{l})$. Επιπλέον, θεωρούμε πως ισχύει ότι:

$\sup\limits_N \int_{-\infty}^{+\infty} \left|\bar{l}\right|^r f_{i,N}(\bar{l})d\bar{l} < \infty, \forall\ i = 0,1,2$ και $0 \leq r \leq 4.$

Με βάση την υπόθεση που μόλις παραθέσαμε είμαστε σε θέση να διατυπώσουμε το παρακάτω θεώρημα:

Θεώρημα 2.2.2(Gordy, Juneja)

Έστω πως η υπόθεση 2.2.1 έχει ισχύ και έστω $f(.)$ η συνάρτηση πυκνότητας πιθανότητας της απώλειας χαρτοφυλακίου L. Τότε καθώς $N \to \infty$ για τη μεροληψία του εκτιμητή $\alpha_{M,N}$ ισχύει ασυμπτωτικά ότι:

$\mathbb{E}\left[\alpha_{M,N} - \alpha\right] = \frac{\theta_c}{N} + \mathcal{O}\left(N^{-3/2}\right)$, όπου

$\theta_c = -Y'(c)$ για $Y(c) = \frac{1}{2}f(c)\mathbb{E}\left[\sigma^2(\omega)\,|\,L(\omega) = c\right]$ και $\sigma^2(\omega)$ τη διασπορά του εσωτερικού δείγματος υπό το ενδεχόμενο ω.

Αξίζει στο σημείο αυτό να αναφέρουμε πως τόσο η υπόθεση 2.2.1 όσο και το θεώρημα 2.2.2 εμπεριέχονται στη δημοσίευση [3] ενώ μια απόδειξη του παραπάνω θεωρήματος μπορεί να βρεθεί στη δημοσίευση [4].

Παρατήρηση 2.2.3

Το θεώρημα 2.2.2 μας παρέχει μια ασυμπτωτική ανάλυση του δεύτερου όρου που εμφανίζεται στο μέσο τετραγωνικό σφάλμα. Συγκεκριμένα, εύκολα βλέπουμε πως για $N \to +\infty$ ισχύει ότι:

$$\mathbb{E}\left[\alpha_{M,N} - \alpha\right] = \mathcal{O}(N^{-1}) \Longleftrightarrow \mathbb{E}\left[\alpha_{M,N}\right] - \alpha = \mathcal{O}(N^{-1}) \Longleftrightarrow \left(\mathbb{E}\left[\alpha_{M,N}\right] - \alpha\right)^2 = \mathcal{O}(N^{-2}).$$

Πόρισμα 2.2.4

Κάτω από την υπόθεση 2.2.1, καθώς $N \to +\infty$, για τη διασπορά του εκτιμητή $\alpha_{M,N}$ παίρνουμε ότι:

$Var\left[\alpha_{M,N}\right] = \frac{\alpha(1-\alpha)}{M} + \mathcal{O}(M^{-1}N^{-1})$.

Απόδειξη: Για τη διασπορά του εκτιμητή $\alpha_{M,N}$, αξιοποιώντας το γεγονός πως οι μεταβλητές $\overline{L}_i$ είναι ανεξάρτητες και ισόνομες, παίρνουμε ότι:

$$\begin{aligned}
Var\left[\alpha_{M,N}\right] &= Var\left[\frac{1}{M}\sum_{i=1}^{M}\mathbb{1}_{\overline{L}_i \geq c}\right] \\
&= \frac{1}{M^2}Var\left[\sum_{i=1}^{M}\mathbb{1}_{\overline{L}_i \geq c}\right] \\
&= \frac{1}{M^2}\sum_{i=1}^{M}Var\left[\mathbb{1}_{\overline{L}_i \geq c}\right] \\
&= \frac{1}{M^2}\sum_{i=1}^{M}Var\left[\mathbb{1}_{\overline{L} \geq c}\right] \\
&= \frac{1}{M}Var\left[\mathbb{1}_{\overline{L} \geq c}\right].
\end{aligned}$$

Ως γνωστόν όμως για τη δείκτρια ενός ενδεχομένου A ισχύει η ιδιότητα $Var\left[\mathbb{1}_A\right] = \mathbb{P}[A](1-\mathbb{P}[A])$ με αποτέλεσμα να παίρνουμε ότι:

$$Var\left[\alpha_{M,N}\right] = \frac{1}{M}Var\left[\mathbb{1}_{\overline{L} \geq c}\right] = \frac{1}{M}\mathbb{P}\left[\overline{L} \geq c\right]\left(1 - \mathbb{P}[\overline{L} \geq c]\right). \quad (2.2)$$

Βλέπουμε επιπλέον πως για τη μέση τιμή της εκτιμήτριάς μας ισχύει ότι:

$$\begin{aligned}
\mathbb{E}\left[\alpha_{M,N}\right] &= \mathbb{E}\left[\frac{1}{M}\sum_{i=1}^{M}\mathbb{1}_{\overline{L}_i \geq c}\right] \\
&= \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\left[\mathbb{1}_{\overline{L}_i \geq c}\right] \\
&= \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\left[\mathbb{1}_{\overline{L} \geq c}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\overline{L} \geq c}\right].
\end{aligned}$$

Παρατηρούμε όμως πως $\mathbb{E}\left[\mathbb{1}_{\overline{L} \geq c}\right] = \mathbb{P}\left[\overline{L} \geq c\right]$, οπότε λαμβάνουμε ότι $\mathbb{E}\left[\alpha_{M,N}\right] = \mathbb{P}\left[\overline{L} \geq c\right]$. (2.3)

Συνδυάζοντας τώρα τις σχέσεις (2.2) και (2.3) καταλήγουμε πως:

$Var\left[\alpha_{M,N}\right] = \frac{1}{M}\mathbb{E}\left[\alpha_{M,N}\right]\left(1 - \mathbb{E}\left[\alpha_{M,N}\right]\right)$.

Εφόσον όμως σύμφωνα με τα δεδομένα μας ισχύει η υπόθεση 2.2.1, με βάση το θεώρημα 2.2.2, καθώς $N \to +\infty$ βλέπουμε ότι:

$\mathbb{E}\left[\alpha_{M,N} - \alpha\right] = \frac{\theta_c}{N} + \mathcal{O}\left(N^{-3/2}\right) \iff \mathbb{E}\left[\alpha_{M,N}\right] = \alpha + \frac{\theta_c}{N} + \mathcal{O}\left(N^{-3/2}\right)$, συνεπώς:

$Var\left[\alpha_{M,N}\right] = \frac{1}{M}\mathbb{E}\left[\alpha_{M,N}\right]\left(1 - \mathbb{E}\left[\alpha_{M,N}\right]\right) = \frac{1}{M}\left(\alpha + \frac{\theta_c}{N} + \mathcal{O}\left(N^{-3/2}\right)\right)\left(1 - \alpha - \frac{\theta_c}{N} - \mathcal{O}\left(N^{-3/2}\right)\right)$.

Αν στην τελευταία σχέση επιχειρήσουμε να ομαδοποιήσουμε τους όρους που εμφανίζονται προκύπτει ότι:

$Var\left[\alpha_{M,N}\right] = \frac{1}{M}\left(\alpha(1-\alpha) + (1-2\alpha)\frac{\theta_c}{N} + \mathcal{O}\left(N^{-3/2}\right)\right) = \frac{1}{M}\left(\alpha(1-\alpha) + \mathcal{O}\left(N^{-1}\right)\right)$, οπότε τελικά λαμβάνουμε πως:

$Var\left[\alpha_{M,N}\right] = \frac{1}{M}\alpha(1-\alpha) + \mathcal{O}\left(M^{-1}N^{-1}\right)$, ακριβώς όπως αναμέναμε. □

Παρατήρηση 2.2.5

Τα παραπάνω αποτελέσματα μας βοηθούν να κατανοήσουμε συνολικά την ασυμπτωτική συμπεριφορά του μέσου τετραγωνικού σφάλματος MSE, μιας και γνωρίζουμε πλέον τη συμπεριφορά των δύο όρων που το απαρτίζουν. Συγκεκριμένα, βλέπουμε πως η διασπορά του εκτιμητή $\alpha_{M,N}$ κυριαρχείται από το πλήθος των σεναρίων M και φθίνει ως M^{-1}, σε αντίθεση με το τετράγωνο της μεροληψίας του εκτιμητή που καθορίζεται από το εσωτερικό πλήθος δείγματος N και φθίνει ως N^{-2}. Έτσι λοιπόν παρατηρούμε πως ισχύει ότι:

$$MSE = Var\left[\alpha_{M,N}\right] + \left(\mathbb{E}\left[\alpha_{M,N} - \alpha\right]\right)^2 \approx \frac{1}{M}\alpha(1-\alpha) + \left(\frac{\theta_c}{N}\right)^2 = \frac{1}{M}\alpha(1-\alpha) + \frac{\theta_c^2}{N^2}\ . \quad (2.4)$$

Έχοντας πλέον στα χέρια μας μια ασυμπτωτική μορφή για την ποσότητα MSE επιστρέφουμε στο αρχικό μας πρόβλημα όπου και αναζητούσαμε τις τιμές των M, N που ελαχιστοποιούν τον παραπάνω δείκτη σφάλματος υπό τον περιορισμό $M.N = k$. Προτού όμως απαντήσουμε στο συγκεκριμένο ερώτημα θα μελετήσουμε τη συμπεριφορά του δείκτη αυτού στην περίπτωση που οι δύο μας παράμετροι είναι ίσοι, με σκοπό να εξετάσουμε το ενδεχόμενο κάποια από αυτές να είναι σημαντικότερη της άλλης. Έτσι λοιπόν θεωρούμε πως $M = N$ και αφού $M.N = k$ υποχρεωτικά παίρνουμε ότι $M = N = k^{1/2}$. Τότε με βάση τη μελέτη που προηγήθηκε λαμβάνουμε ασυμπτωτικά ότι:

$$MSE = \frac{1}{M}\alpha(1-\alpha) + \frac{\theta_c^2}{N^2} = \alpha(1-\alpha)k^{-1/2} + \theta_c^2 k^{-1} = \mathcal{O}(k^{-1/2})\ .$$

Αυτό που αντιλαμβανόμαστε από την παραπάνω σχέση είναι πως στην περίπτωση που $M = N$ το μέσο τετραγωνικό σφάλμα κυριαρχείται από τον όρο διασποράς του εκτιμητή. Στην ιδανική όμως περίπτωση θα θέλαμε οι δυο όροι που εμφανίζονται στη σχέση (2.4) να φέρουν την ίδια τάξη μεγέθους ώστε να συνεισφέρουν κατά το δυνατόν ισόποσα στον καθορισμό της συμπεριφοράς του MSE. Για να το επιτύχουμε αυτό θα πρέπει να ισχύει σίγουρα ότι $ord(M) > ord(N)$ και ειδικότερα θα πρέπει να ικανοποιείται η συνθήκη $ord(M) = ord(N)^2$. Έτσι εντελώς διαισθητικά θα θέλαμε να επιλέξουμε τις παραμέτρους μας με τέτοιο τρόπο ώστε $M.N = k$ και η τάξη του N να ισούται με τη ρίζα της τάξης του M. Είναι εύκολο τώρα να δούμε πως για την επιλογή $M = ck^{2/3}$ και $N = \frac{1}{c}k^{1/3}$, με $c \in \mathbb{R}_+$, ικανοποιούνται οι παραπάνω απαιτήσεις και η τάξη του MSE ισούται με $k^{-2/3}$.

Θα προσπαθήσουμε τώρα να υποστηρίξουμε τα παραπάνω λογικά συμπεράσματα αναλυτικά αξιοποιώντας τη μέθοδο του Lagrange για το ακόλουθο πρόβλημα ελαχιστοποίησης υπό περιορισμό:

$$\begin{cases} \min\limits_{M,N} MSE \\ M.N = k \\ M,N \geq 0 \end{cases} \quad \text{ή διαφορετικά} \quad \begin{cases} \min\limits_{M,N} \left[\frac{1}{M}\alpha(1-\alpha) + \frac{\theta_c^2}{N^2}\right] \\ M.N = k \\ M,N \geq 0 \end{cases} .$$

Λύση του προβλήματος

Ορίζουμε τη συνάρτηση του Lagrange μέσα από τη σχέση $L(M,N,\lambda) = \frac{1}{M}\alpha(1-\alpha) + \frac{\theta_c^2}{N^2} + \lambda(MN-k)$ και παίρνουμε ότι:

- $\frac{\partial L}{\partial M} = 0 \Longrightarrow -\frac{1}{M^2}\alpha(1-\alpha) + \lambda N = 0$
- $\frac{\partial L}{\partial N} = 0 \Longrightarrow -\frac{2}{N^3}\theta_c^2 + \lambda M = 0.$

Επιπλέον, ισχύει και ο περιορισμός $MN = k \Longrightarrow N = \frac{k}{M}$. Προχωρώντας σε αντικατάσταση της τελευταίας έκφρασης παίρνουμε ότι:

$$-\frac{2}{N^3}\theta_c^2 + \lambda M = 0 \Longrightarrow -\frac{2}{(k/M)^3}\theta_c^2 + \lambda M = 0 \Longrightarrow -2\theta_c^2\frac{M^3}{k^3} + \lambda M = 0 \overset{M>0}{\Longrightarrow} \lambda = 2\theta_c^2\frac{M^2}{k^3}.$$

Από την πρώτη τώρα σχέση βλέπουμε πως:

$$-\frac{1}{M^2}\alpha(1-\alpha) + \lambda N = 0 \Longrightarrow -\frac{1}{M^2}\alpha(1-\alpha) + 2\theta_c^2\frac{M^2}{k^3}\frac{k}{M} = 0 \Longrightarrow M^3 = \frac{\alpha(1-\alpha)}{2\theta_c^2}k^2, \text{ συνεπώς:}$$

$$M = \left(\frac{\alpha(1-\alpha)}{2\theta_c^2}\right)^{1/3} k^{2/3} \equiv \beta^* k^{2/3}.$$

Τέλος, για την παράμετρο του εσωτερικού δείγματος έχουμε ότι:

$$N = \frac{k}{M} \Longrightarrow \frac{k}{\beta^* k^{2/3}} = \frac{k^{1/3}}{\beta^*}.$$

Παρατήρηση 2.2.6

Στο σημείο αυτό βλέπουμε πως τα αναλυτικά μας αποτελέσματα έρχονται σε πλήρη συμφωνία με τα διαισθητικά, όπου το ρόλο της θεωρητικής σταθεράς παίζει η ποσότητα $\beta^* = \left(\frac{\alpha(1-\alpha)}{2\theta_c^2}\right)^{1/3}$.

2.3 Παραλλαγή και Γενίκευση του Προβλήματός μας

Έως τώρα προσπαθήσαμε να προσδιορίσουμε αριθμητικά την πιθανότητα η απώλεια ενός χαρτοφυλακίου να ξεπεράσει ένα φράγμα $c \in \mathbb{R}$ στηριζόμενοι εξ' ολοκλήρου σε πολλαπλά σενάρια/ενδεχόμενα που μπορούσαν να επηρεάσουν τις τιμές των χρεογράφων μας. Με άλλα λόγια, μέχρι στιγμής η απώλεια ενός χαρτοφυλακίου δεν εξετάστηκε ως συνάρτηση των χρεογράφων ή και άλλων παραγόντων ρίσκου που το απαρτίζουν αλλά μέσα από τα σενάρια που επιδρούν σε αυτά. Αν επιθυμούσαμε να εργαστούμε με ένα διαφορετικό τρόπο θα μπορούσαμε να ορίσουμε μια πολυδιάστατη μεταβλητή Y η οποία δεν κάνει τίποτα άλλο από το να περιέχει τιμές χρεογράφων ή άλλων παραγόντων ρίσκου για ένα μικρό χρονικό ορίζοντα Δt. Κάτω από το συγκεκριμένο πλαίσιο, θα μπορούσαμε να θεωρήσουμε πως η εξεταζόμενη απώλεια δε συναρτάται αποκλειστικά από ενδεχόμενα $\omega \in \Omega$, αλλά και από την τιμή της νέας μεταβλητής Y. Προφανώς τώρα η απώλεια του χαρτοφυλακίου L δεν είναι μια κατάλληλη μεταβλητή για τη μελέτη μας καθώς δεν καθορίζεται πλήρως δίχως τη δέσμευση ως προς κάποια τιμή της Y. Μια ιδέα για να αντιμετωπίσουμε το πρόβλημα αυτό είναι να την αντικαταστήσουμε από τη μεταβλητή $\mathbb{E}[L|Y]$, η οποία για κάθε τιμή y της Y δίνει μια συγκεκριμένη τιμή ακριβώς όπως συνέβαινε για το κάθε σενάριο $\omega \in \Omega$. Η διαφορά στην περίπτωση αυτή είναι πως για κάθε τιμή y της μεταβλητής Y, η ποσότητα $\mathbb{E}[L|Y=y]$ δε δίνει μια συγκεκριμένη απώλεια όπως ίσχυε για τα σενάρια ω, αλλά την αναμενόμενη τιμή της κάτω από αυτή τη δέσμευση. Με βάση τα παραπάνω είναι εύκολο να δούμε πως το ενδιαφέρον μας πλέον δεν εστιάζεται στην τιμή $\alpha = \mathbb{P}[L \geq c]$, αλλά στην τιμή $\eta = \mathbb{P}[\mathbb{E}[L|Y] \geq c] = \mathbb{P}[\mathbb{E}[L-c|Y] \geq 0]$, όπου για $X = L - c$ τελικά αναζητάμε την ποσότητα $\eta = \mathbb{P}[\mathbb{E}[X|Y] \geq 0]$. Το ερώτημα που απομένει τώρα να απαντήσουμε είναι με ποιόν αριθμητικό τρόπο θα μπορούσαμε να υπολογίσουμε τη συγκεκριμένη πιθανότητα. Το πρόβλημα αυτό με τον τρόπο που μελετήθηκε θα μπορούσε να ενταχθεί σε μια γενικότερη κατηγορία προβλημάτων τα οποία φέρουν την ακόλουθη διατύπωση:

Ορισμός του προβλήματος

Έστω X μια μονοδιάστατη τυχαία μεταβλητή, η οποία δεσμεύεται από την τιμή μιας πολυδιάστατης μεταβλητής Y. Ορίζουμε ως $\mathbb{E}^{\,y}$ την αναμενόμενη τιμή της μεταβλητής X δοθέντος πως η Y λαμβάνει την τιμή y, δηλαδή θεωρούμε ότι $\mathbb{E}^{\,y}[X] = \mathbb{E}[X|Y = y]$. Καλούμαστε τότε να υπολογίσουμε την πιθανότητα η δεσμευμένη μέση τιμή της X ως προς Y να είναι μη αρνητική, δηλαδή ζητάμε να προσδιορίσουμε την τιμή $\eta = \mathbb{P}\,[\mathbb{E}[X\,|\,Y] \geq 0]$.

Αριθμητική προσέγγιση

Οι βασικές ιδέες στις οποίες θα στηριχτούμε για να εκτιμήσουμε τη ζητούμενη πιθανότητα είναι ανάλογες με εκείνες που είδαμε στην εμφωλευμένη διαδικασία προσομοίωσης της τιμής $\mathbb{P}[L \geq c]$. Καταρχάς από βασικές ιδιότητες της δείκτριας συνάρτησης βλέπουμε πως ισχύει ότι:

$$\eta = \mathbb{P}\,[\mathbb{E}[X\,|\,Y] \geq 0] = \mathbb{E}\left[\mathbb{1}_{\mathbb{E}[X\,|\,Y]\geq 0}\right].$$

Η παραπάνω έκφραση μπορεί να απλοποιηθεί περαιτέρω εάν λάβουμε υπόψη μας τον ορισμό της δείκτριας συνάρτησης και ειδικότερα πως:

$$\mathbb{1}_{\mathbb{E}[X\,|\,Y]\geq 0} = \begin{cases} 1, & \text{αν } \ \mathbb{E}[X\,|\,Y] \geq 0 \\ 0, & \text{διαφορετικά} \end{cases}.$$

Συγκεκριμένα, παρατηρούμε πως η δείκτρια του εξεταζόμενου ενδεχομένου συμπεριφέρεται σα μια συνάρτηση Heaviside που λαμβάνει τιμές της μορφής $\mathbb{E}[X\,|\,Y]$. Με άλλα λόγια, εύκολα συμπεραίνουμε πως $\mathbb{1}_{\mathbb{E}[X\,|\,Y]\geq 0} = H(\mathbb{E}[X\,|\,Y])$ και η ποσότητα που καλούμαστε να προσδιορίσουμε μπορεί να γραφτεί ως εξής:

$$\eta = \mathbb{E}\left[\mathbb{1}_{\mathbb{E}[X\,|\,Y]\geq 0}\right] = \mathbb{E}[H(\mathbb{E}[X\,|\,Y])].$$

Στο σημείο αυτό, θα προσπαθήσουμε να εκτιμήσουμε την παραπάνω ποσότητα αριθμητικά κάνοντας χρήση της μεθόδου Monte Carlo. Έτσι, σε ένα πρώτο στάδιο, για ένα ανεξάρτητο και ισόνομο δείγμα $y_1, y_2, \ldots, y_M$ από το νόμο των μεγάλων αριθμών λαμβάνουμε ότι:

$$\eta = \mathbb{E}[H(\mathbb{E}[X\,|\,Y])] \approx \frac{1}{M}\sum_{i=1}^{M} H(\mathbb{E}[X|Y = y_i]) \overset{\text{ορ.}}{=} \frac{1}{M}\sum_{i=1}^{M} H(\mathbb{E}^{\,y_i}[X]).$$

Βέβαια, σε πολλές περιπτώσεις είναι δύσκολος ο αναλυτικός υπολογισμός της ποσότητας $\mathbb{E}^{\,y_i}[X]$, γεγονός που μας ωθεί να καταφύγουμε και σε ένα εσωτερικό σχήμα Monte Carlo το οποίο θα υπολογίζει τη μέση τιμή της μεταβλητής X ως προς την εκάστοτε δέσμευση y_i, $i = 1, \ldots, M$. Συνεπώς, αν ορίσουμε για την τιμή y_i της μεταβλητής Y ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $x_1^{y_i}, x_2^{y_i}, \ldots, x_N^{y_i}$ με $\mathbb{E}\left[x_j^{y_i}\right] = \mathbb{E}^{\,y_i}[X]$ βλέπουμε ότι:

$$\overline{\mathbb{E}}^{\,y_i}[X] = \frac{1}{N}\sum_{j=1}^{N} x_j^{y_i} \ \xrightarrow[N\to\infty]{\mathbb{P}} \ \mathbb{E}^{\,y_i}[X].$$

Τελικά λοιπον, για τον προσδιορισμό της ποσότητας η στηριζόμαστε στον εκτιμητή I ο οποίος δίνεται από τον ακόλουθο τύπο:

$$I := \frac{1}{M}\sum_{i=1}^{M} H\left(\overline{\mathbb{E}}^{\,y_i}[X]\right) = \frac{1}{M}\sum_{i=1}^{M} H\left(\frac{1}{N}\sum_{j=1}^{N} x_j^{y_i}\right).$$

Όπως έχουμε ήδη εξηγήσει, στον τύπο αυτό δεν κάνουμε κάτι διαφορετικό από το να εφαρμόζουμε την απλή μέθοδο Monte Carlo σε δύο στάδια, ένα εξωτερικό και ένα εσωτερικό. Χρησιμοποιώντας παρόμοια επιχειρήματα και θεωρήματα με εκείνα που είδαμε στο πρόβλημα απώλειας χαρτοφυλακίου, για το μέσο τετραγωνικό σφάλμα του εκτιμητή που εξετάζουμε μπορεί να δειχθεί πως ισχύει ότι:

$$MSE = \mathbb{E}\left[\left(\frac{1}{M}\sum_{i=1}^{M} H\left(\overline{\mathbb{E}}^{\,y_i}[X]\right) - \mathbb{P}\,[\mathbb{E}[X\,|\,Y] \geq 0]\right)^2\right] = \mathcal{O}(M^{-1} + N^{-2}).$$

Συνεπώς, αν επιθυμούσαμε να εξασφαλίσουμε ένα μέσο τετραγωνικό σφάλμα τάξης ϵ^2, όπου $\epsilon > 0$ θα επιβάλλαμε τις συνθήκες $M = \mathcal{O}(\epsilon^{-2})$ και $N = \mathcal{O}(\epsilon^{-1})$. Επιπλέον, υπενθυμίζουμε πως η πολυπλοκότητα της μεθόδου μας δίνεται μέσω της σχέσης $\mathcal{C} = M.N$, με αποτέλεσμα στο πλαίσιο που μελετάμε να παίρνουμε πως $\mathcal{C} = \mathcal{O}(\epsilon^{-2})\mathcal{O}(\epsilon^{-1}) = \mathcal{O}(\epsilon^{-3})$.

Προσαρμοσμένη επιλογή πλήθους προσομοιώσεων

Το ερώτημα που δημιουργείται στο σημείο αυτό είναι εάν υπάρχει η δυνατότητα μείωσης του κόστους της μεθόδου μέσα από κατάλληλη επιλογή παραμέτρων. Για να απαντήσουμε στο συγκεκριμένο ερώτημα θα πρέπει να εξετάσουμε προσεκτικά τον εκτιμητή που χρησιμοποιούμε. Έως τώρα στηριζόμασταν σε μια εμφωλευμένη διαδικασία προσομοίωσης όπου για κάθε εξωτερική τιμή y_i της μεταβλητής Y δημιουργούσαμε ένα δείγμα N ανεξάρτητων και ισόνομων τυχαίων μεταβλητών που δεσμεύονταν ως προς την τιμή αυτή. Με άλλα λόγια, διατηρούσαμε ένα σταθερό πλήθος εσωτερικών προσομοιώσεων για κάθε εξωτερική τιμή y_i. Αυτό που καλούμαστε τώρα να εξετάσουμε είναι αν υπάρχει κάποιος τρόπος να προσαρμόσουμε κατάλληλα το πλήθος των εσωτερικών προσομοιώσεων ανάλογα με τις υπάρχουσες κάθε φορά ανάγκες. Επειδή η προηγούμενη πρόταση με μια πρώτη ματιά φαίνεται κάπως ασαφής θα προσπαθήσουμε να την κατανοήσουμε μέσα από ένα παράδειγμα.

Έστω y_1 και y_2 δυο τιμές της μεταβλητής Y τέτοιες ώστε $\mathbb{E}[X|Y = y_1] = c_1$, όπου $c_1 >> 0$ και $\mathbb{E}[X|Y = y_2] = c_2$, όπου $c_2 \in [-\epsilon, \epsilon]$ με $\epsilon << 1$. Παρά το γεγονός πως οι εσωτρικοί εκτιμητές $\overline{\mathbb{E}}^{\,y_1}[X]$ και $\overline{\mathbb{E}}^{\,y_2}[X]$ εμφανίζουν κάποια αβεβαιότητα μπορούμε να πούμε πως είναι αρκετά πιθανό για ένα ικανοποιητικό δείγμα να βρίσκονται κοντά στη μέση τιμή τους, γεγονός που μας οδηγεί σε σημαντικά συμπεράσματα. Συγκεκριμένα, για την πρώτη τιμή y_1 είναι απόλυτα λογικό πως ακόμα και για ένα πολύ μικρό δείγμα θα πάρουμε ότι $\overline{\mathbb{E}}^{\,y_1}[X] > 0$ και κατά συνέπεια από την ποσότητα $H\left(\overline{\mathbb{E}}^{\,y_1}[X]\right)$ λαμβάνουμε το σωστό αποτέλεσμα. Αντίθετα, για τη δεύτερη τιμή y_2 φαίνεται πως θα χρειαστούμε ένα κατά πολύ μεγαλύτερο δείγμα για να εξασφαλίσουμε πως η εκτιμήτρια $\overline{\mathbb{E}}^{\,y_2}[X]$ φέρει το σωστό πρόσημο και άρα η ποσότητα $H\left(\overline{\mathbb{E}}^{\,y_2}[X]\right)$ δίνει τη σωστή τιμή. Μέσα από το παράδειγμα αυτό γίνεται αντιληπτό πως για να αποφανθούμε σχετικά με το πλήθος του δείγματος N κάθε φορά αυτό που οφείλουμε να ελέγξουμε είναι κατά πόσο η τιμή $|\mathbb{E}[X|Y = y_i]|$ βρίσκεται κοντά ή μακριά από το σημείο 0. Βέβαια, από αυτά που έχουμε δει έως τώρα εύκολα κατανοούμε πως η παραπάνω τιμή δε μας είναι a-priori γνωστή, γεγονός που μας ωθεί να καταφύγουμε σε κάποιον άλλο τρόπο καθορισμού του εσωτερικού δείγματος N σε κάθε επανάληψη. Αυτό που θα κάνουμε είναι δοθέντος ενός εσωτερικού εκτιμητή $\overline{\mathbb{E}}^{\,y_i}_{N_i}[X] > 0$ να προσδιορίσουμε ένα άνω φράγμα για την πιθανότητα ο εκτιμητής να αλλάξει πρόσημο αυξάνοντας το πλήθος του δείγματος κατά μία μεταβλητή. Αξίζει να αναφέρουμε πως στον εκτιμητή $\overline{\mathbb{E}}^{\,y_i}[X]$ χρησιμοποιήσαμε το δείκτη N_i ώστε να δείξουμε το πλήθος του δείγματος από το οποίο προέρχεται.

Εκτίμηση ζητούμενου άνω φράγματος

Βλέπουμε καταρχάς πως για τον εκτιμητή που παίρνουμε με βάση το νέο δείγμα ισχύει ότι:

$$
\begin{aligned}
\overline{\mathbb{E}}^{\,y_i}_{N_i+1}[X] &= \frac{1}{N_i+1}\sum_{j=1}^{N_i+1} x_j^{y_i} \\
&= \frac{1}{N_i+1}x^{y_i}_{N_i+1} + \frac{1}{N_i+1}\sum_{j=1}^{N_i} x_j^{y_i} \\
&= \frac{1}{N_i+1}x^{y_i}_{N_i+1} + \frac{N_i}{N_i+1}\left(\frac{1}{N_i}\sum_{j=1}^{N_i} x_j^{y_i}\right) \\
&= \frac{1}{N_i+1}x^{y_i}_{N_i+1} + \frac{N_i}{N_i+1}\overline{\mathbb{E}}^{\,y_i}_{N_i}[X].
\end{aligned}
$$

Επιπλέον, θεωρούμε $\mu_i = \mathbb{E}[X|Y = y_i] > 0$ και $\sigma_i^2 = Var[X|Y = y_i]$ και αξιοποιώντας το αποτέλεσμα που μόλις πήραμε λαμβάνουμε ότι:

$$
\begin{aligned}
\mathbb{P}\left[\overline{\mathbb{E}}\,^{y_i}_{N_i+1}[X] \leq 0 \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] &= \mathbb{P}\left[\tfrac{1}{N_i+1}x^{y_i}_{N_i+1} + \tfrac{N_i}{N_i+1}\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] \leq 0 \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] \\
&= \mathbb{P}\left[x^{y_i}_{N_i+1} + N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] \leq 0 \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] \\
&= \mathbb{P}\left[N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] \leq -x^{y_i}_{N_i+1} \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] \\
&= \mathbb{P}\left[\mu_i + N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] \leq \mu_i - x^{y_i}_{N_i+1} \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] \\
&\leq \mathbb{P}\left[\mu_i + N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] \leq \left|\mu_i - x^{y_i}_{N_i+1}\right| \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] \\
&= \mathbb{P}\left[\mu_i + N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] \leq \left|x^{y_i}_{N_i+1} - \mu_i\right| \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right]. \quad (2.5)
\end{aligned}
$$

Παρατηρούμε όμως ότι $\mu_i = \mathbb{E}[X|Y = y_i] = \mathbb{E}\left[x^{y_i}_{N_i+1}\right]$ και εφόσον υποθέσαμε ότι $\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] > 0$ από τη σχέση (2.5) και την ανισότητα του Chebysev βλέπουμε ότι:

$$\mathbb{P}\left[\overline{\mathbb{E}}\,^{y_i}_{N_i+1}[X] \leq 0 \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] \leq \frac{Var\left[x^{y_i}_{N_i+1}\right]}{\left(\mu_i+N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right)^2} = \frac{Var[X|Y=y_i]}{\left(\mu_i+N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right)^2} = \frac{\sigma_i^2}{\left(\mu_i+N_i\overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right)^2}\ .$$

Τέλος, αν θεωρήσουμε πως το εσωτερικό πλήθος δείγματος είναι αρκετά μεγάλο με αποτέλεσμα να ισχύει ότι:

$$\overline{\mathbb{E}}\,^{y_i}_{N_i}[X] \approx \mathbb{E}[X|Y = y_i] = \mu_i \;\; \text{και} \;\; \frac{1}{N_i} \approx \frac{1}{N_i+1}$$

καταλήγουμε πως:

$$\mathbb{P}\left[\overline{\mathbb{E}}\,^{y_i}_{N_i+1}[X] \leq 0 \,\middle|\, \overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right] \leq \frac{\sigma_i^2}{N_i^2\mu_i^2}\ .$$

Παρατήρηση 2.3.1

Με βάση το παραπάνω αποτέλεσμα, αν επιθυμούσαμε η προσθήκη μιας επιπλέον μεταβλητής να μην αλλάζει πρόσημο στον εκτιμητή μας με πιθανότητα $1-\epsilon$, $\epsilon << 1$ θα αρκούσε να επιλέξουμε το εσωτερικό πλήθος N_i με τέτοιο τρόπο ώστε να πληροί τη συνθήκη:

$$\frac{\sigma_i^2}{N_i^2\mu_i^2} \leq \epsilon \Longrightarrow N_i^2 \geq \frac{\sigma_i^2}{\epsilon\mu_i^2} \Longrightarrow N_i \geq \frac{\sigma_i}{\epsilon^{1/2}|\mu_i|} = \epsilon^{-1/2}\frac{\sigma_i}{|\mu_i|}.$$

Αν τώρα θέσουμε $d_i \equiv |\mu_i|$ λαμβάνουμε τελικά ότι $N_i \geq \epsilon^{-1/2}\frac{\sigma_i}{d_i}$. Είναι εύκολο στο σημείο αυτό να δούμε πως στην περίπτωση που το πλήθος δείγματος N_i ξεπερνά το συγκεκριμένο φράγμα ισχύει με πιθανότητα $1-\epsilon$ ότι:

$$H\left(\overline{\mathbb{E}}\,^{y_i}_{N_i}[X]\right) = H\left(\overline{\mathbb{E}}\,^{y_i}_{N_i+1}[X]\right) = H\left(\mathbb{E}[X|Y = y_i]\right),$$

με αποτέλεσμα να παίρνουμε τη σωστή εκτίμηση με την παραπάνω υψηλή πιθανότητα. Προφανώς, επιθυμούμε να ελαχιστοποιήσουμε το κόστος που δαπανούμε και για το λόγο αυτό από όλες τις τιμές N_i που ικανοποιούν τη συνθήκη που ορίσαμε θα επιλέγουμε κάθε φορά τη μικρότερη δυνατή ή με άλλα λόγια θα επιλέγουμε $N_i = \epsilon^{-1/2}\frac{\sigma_i}{d_i}$. Ωστόσο, υπάρχει ένα ακόμα λεπτό σημείο στο οποίο αξίζει να δοθεί ιδιαίτερη προσοχή. Υπενθυμίζουμε πως στην περίπτωση της ομοιόμορφης διαμέρισης για να επιτύχουμε τάξη του MSE ίση με ϵ^2 έπρεπε να πάρουμε εσωτερικό δείγμα τάξης $\mathcal{O}(\epsilon^{-1})$ σε κάθε επανάληψη. Έτσι, αυτό που πρέπει να απαιτήσουμε είναι το φράγμα που έχουμε προσδιορίσει να μην ξεπερνά σε τάξη την παλιά τιμή πλήθους δείγματος N. Με πιο απλά λόγια, θα θέλαμε για την κάθε εξωτερική επανάληψη που τρέχουμε να δαπανούμε μικρότερο κόστος από εκείνο που καταναλώναμε έως τώρα. Βέβαια, κάτι τέτοιο δεν είναι πάντα εφικτό μιας και στην περίπτωση που $\delta_i := d_i/\sigma_i = o(\epsilon^{1/2})$, καθώς $\epsilon \to 0$, βλέπουμε ότι:

$$\delta_i \leq \eta\epsilon^{1/2} \ , \forall \ \eta > 0 \Longrightarrow \frac{1}{\delta_i} \geq \frac{1}{\eta}\epsilon^{-1/2} \ , \forall \ \eta > 0 \Longrightarrow \epsilon^{-1/2}\frac{\sigma_i}{d_i} \geq \frac{1}{\eta}\epsilon^{-1} \ , \forall \ \eta > 0.$$

Με βάση λοιπόν τις πράξεις αυτές, οδηγούμαστε στο συμπέρασμα πως για τις τιμές y_i τέτοιες ώστε $\frac{d_i}{\sigma_i} = o(\epsilon^{1/2})$ θα χρησιμοποιούμε την παλιά τιμή εσωτερικού δείγματος τάξης $\mathcal{O}(\epsilon^{-1})$. Τέλος, δε θα πρέπει να ξεχνάμε πως οι ποσότητες N_i είναι θετικοί ακέραιοι αριθμοί. Αν τώρα συμπεριλάβουμε όλες τις παραπάνω παρατηρήσεις επιλέγουμε την τιμή N_i να δίνεται μέσω του τύπου:

$$N_i = \left\lceil \min\left(\mathcal{O}(\epsilon^{-1}), \epsilon^{-1/2}\frac{\sigma_i}{d_i}\right)\right\rceil. \quad (2.6)$$

2.4 Κόστος Προσαρμοσμένης Προσομοίωσης

Με βάση αυτά που είδαμε στην προηγούμενη ενότητα, εύκολα αντιλαμβανόμαστε πως για να προσδιορίσουμε τελικά τη ζητούμενη τιμή η, εφαρμόζουμε ένα διπλό σχήμα Monte Carlo, όπου για κάθε εξωτερική τιμή y_i της τυχαίας μεταβλητής Y, στηριζόμαστε σε N_i το πλήθος εσωτερικό δείγμα προσομοίωσης, το οποίο υπολογίζεται με βάση τη σχέση (2.6). Συνεπώς, θα μπορούσαμε να πάρουμε πως για τη μέθοδο που έχουμε πλέον στα χέρια μας ισχύει ότι:

$$\mathcal{C} = \sum_{i=1}^{M} C_i = \sum_{i=1}^{M} N_i.$$

Έτσι, βλέπουμε πως το ολικό υπολογιστικό κόστος της μεθόδου μας διαφοροποιείται ανάλογα με τις τιμές y_i ως προς τις οποίες έχουμε δεσμεύσει, με αποτέλεσμα να καθίσταται αδύνατος ο υπολογισμός του μέσω του παραπάνω τύπου. Για να αντιμετωπίσουμε το πρόβλημα αυτό, θα προσπαθήσουμε να εξετάσουμε την αριθμητική πολυπλοκότητα της μεθόδου μας μέσα από υποθέσεις για την κατανομή της μεταβλητής $\delta = d/\sigma$, άρα και της μεταβλητής N, δίχως να εστιάζουμε στο δείγμα N_i της τελευταίας. Ειδικότερα, έχουμε το ακόλουθο θεώρημα:

Θεώρημα 2.4.1

Έστω X μια μονοδιάστατη τυχαία μεταβλητή, η οποία δεσμεύεται από την τιμή μιας πολυδιάστατης μεταβλητής Y. Έστω επιπλέον I η αριθμητική εκτίμηση της τιμής $\eta = \mathbb{P}\left[\mathbb{E}[X|Y] \geq 0\right]$ και έστω πως επιθυμούμε η μέθοδός μας να φέρει ένα μέσο τετραγωνικό σφάλμα της τάξης του ϵ^2. Ορίζουμε τότε, δοθέντος της μεταβλητής Y, το πλήθος του εσωτερικού δείγματος μέσω της σχέσης:

$$N = \left\lceil \min\left(\mathcal{O}(\epsilon^{-1}), \epsilon^{-1/2}\frac{\sigma}{d}\right)\right\rceil, \text{ όπου } d = \mathbb{E}[X|Y] \text{ και } \sigma^2 = Var[X|Y].$$

Τέλος, υποθέτουμε πως για το εξωτερικό πλήθος δείγματος του αριθμητικού μας σχήματος ισχύει ότι $M = \mathcal{O}(\epsilon^{-2})$ και πως η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής $\delta := d/\sigma$ είναι τοπικά φραγμένη κοντά στο 0. Τότε, για το κόστος της αριθμητικής μας μεθόδου λαμβάνουμε προσεγγιστικά ότι:

$$\mathcal{C} = \mathcal{O}\left(\epsilon^{-5/2}\log\left(\epsilon^{-1/2}\right)\right).$$

Απόδειξη: Έστω ρ η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής δ. Τότε σύμφωνα με την υπόθεσή μας, θα υπάρχουν $\rho_0 > 0$ και $\delta_0 > 0$ τέτοια ώστε $\rho(\delta) \leq \rho_0, \forall \ \delta \in [0, \delta_0]$. Επιπλέον παρατηρούμε πως:

$$N = \lceil \min\left(\mathcal{O}(\epsilon^{-1}), \epsilon^{-1/2}\tfrac{\sigma}{d}\right)\rceil = \left\lceil \min\left(\mathcal{O}(\epsilon^{-1}), \epsilon^{-1/2}\tfrac{1}{\delta}\right)\right\rceil \equiv a(\delta),$$

απ' όπου παίρνουμε εύκολα ότι η συνάρτηση a είναι θετική και μη-αύξουσα. Τότε για την αναμενόμενη τιμή της μεταβλητής N βλέπουμε πως ισχύει ότι:

$$\begin{aligned}
\mathbb{E}[N] = \mathbb{E}[a(\delta)] &= \int_0^{\infty} a(\delta)\rho(\delta)d\delta \\
&= \int_0^{\delta_0} a(\delta)\rho(\delta)d\delta + \int_{\delta_0}^{\infty} a(\delta)\rho(\delta)d\delta \\
&\leq \int_0^{\delta_0} a(\delta)\rho_0 d\delta + \int_{\delta_0}^{\infty} a(\delta)\rho(\delta)d\delta \\
&= \rho_0 \int_0^{\delta_0} a(\delta)d\delta + \int_{\delta_0}^{\infty} a(\delta)\rho(\delta)d\delta \\
&\leq \rho_0 \int_0^{\delta_0} a(\delta)d\delta + a(\delta_0)\int_{\delta_0}^{\infty} \rho(\delta)d\delta \\
&\leq \rho_0 \int_0^{\delta_0} a(\delta)d\delta + a(\delta_0)\int_{0}^{\infty} \rho(\delta)d\delta \\
&= \rho_0 \int_0^{\delta_0} a(\delta)d\delta + a(\delta_0) \\
&\overset{\text{ορ.}}{=} \rho_0 \int_0^{\delta_0} \left\lceil \min\left(\mathcal{O}(\epsilon^{-1}), \epsilon^{-1/2}\tfrac{1}{\delta}\right)\right\rceil d\delta + \left\lceil \min\left(\mathcal{O}(\epsilon^{-1}), \epsilon^{-1/2}\tfrac{1}{\delta_0}\right)\right\rceil. \quad (2.7)
\end{aligned}$$

Εξ' ορισμού όμως για τη συνάρτηση $f(\epsilon) = \mathcal{O}(\epsilon^{-1})$ θα $\exists$ σταθερά $c_1 > 0$ ώστε καθώς $\epsilon \to 0$ να παίρνουμε ότι $f(\epsilon) \leq c_1\epsilon^{-1}$. Έτσι με βάση τη σχέση (2.7) μπορούμε τελικά να δούμε ότι:

$$\mathbb{E}[N] \leq \rho_0 \int_0^{\delta_0} \left\lceil \min\left(c_1\epsilon^{-1}, \epsilon^{-1/2}\tfrac{1}{\delta}\right)\right\rceil d\delta + \left\lceil \min\left(c_1\epsilon^{-1}, \epsilon^{-1/2}\tfrac{1}{\delta_0}\right)\right\rceil.$$

Επιπλέον, θεωρούμε την τιμή r η οποία ορίζεται μέσω της σχέσης $r = \max\left\{1/c_1\epsilon^{1/2}, \delta_0\right\}$ και για την αναμενόμενη τιμή της μεταβλητής N έχουμε ότι:

$$\begin{aligned}
\mathbb{E}[N] &\leq \rho_0 \int_0^{r} \left\lceil \min\left(c_1\epsilon^{-1}, \epsilon^{-1/2}\tfrac{1}{\delta}\right)\right\rceil d\delta + \left\lceil \min\left(c_1\epsilon^{-1}, \epsilon^{-1/2}\tfrac{1}{\delta_0}\right)\right\rceil \\
&\leq \rho_0 \int_0^{\epsilon^{1/2}/c_1} \left\lceil \min\left(c_1\epsilon^{-1}, \epsilon^{-1/2}\tfrac{1}{\delta}\right)\right\rceil d\delta + \rho_0 \int_{\epsilon^{1/2}/c_1}^{r} \left\lceil \min\left(c_1\epsilon^{-1}, \epsilon^{-1/2}\tfrac{1}{\delta}\right)\right\rceil d\delta + \left\lceil \epsilon^{-1/2}\tfrac{1}{\delta_0}\right\rceil \\
&= \rho_0 \int_0^{\epsilon^{1/2}/c_1} \lceil c_1\epsilon^{-1}\rceil d\delta + \rho_0 \int_{\epsilon^{1/2}/c_1}^{r} \left\lceil \epsilon^{-1/2}\tfrac{1}{\delta}\right\rceil d\delta + \left\lceil \epsilon^{-1/2}\tfrac{1}{\delta_0}\right\rceil \\
&= \rho_0 \lceil c_1\epsilon^{-1}\rceil \int_0^{\epsilon^{1/2}/c_1} d\delta + \rho_0 \int_{\epsilon^{1/2}/c_1}^{r} \left\lceil \epsilon^{-1/2}\tfrac{1}{\delta}\right\rceil d\delta + \left\lceil \epsilon^{-1/2}\tfrac{1}{\delta_0}\right\rceil \\
&= \rho_0 \lceil c_1\epsilon^{-1}\rceil \frac{1}{c_1}\epsilon^{1/2} + \rho_0 \int_{\epsilon^{1/2}/c_1}^{r} \left\lceil \epsilon^{-1/2}\tfrac{1}{\delta}\right\rceil d\delta + \left\lceil \epsilon^{-1/2}\tfrac{1}{\delta_0}\right\rceil.
\end{aligned}$$

Από βασική τώρα ιδιότητα του ακέραιου μέρους η τελευταία σχέση φράσσεται ως εξής:

$\mathbb{E}[N] \le \rho_0(1+c_1\epsilon^{-1})\frac{1}{c_1}\epsilon^{1/2} + \rho_0\int_{\epsilon^{1/2}/c_1}^{r}\left(\epsilon^{-1/2}\frac{1}{\delta}+1\right)d\delta + \left(\epsilon^{-1/2}\frac{1}{\delta_0}+1\right).$

Επιλύοντας στην παραπάνω σχέση το ολοκλήρωμα που εμφανίζεται και αξιοποιώντας το γεγονός πως καθώς $\epsilon \to 0$ ισχύει ότι $r \to \delta_0$ οδηγούμαστε στο αποτέλεσμα:

$\mathbb{E}[N] = \mathcal{O}\left(\epsilon^{-1/2}\log\left(\epsilon^{-1/2}\right)\right)$. (2.8)

Τέλος, για την πολυπλοκότητα της αριθμητικής μας μεθόδου λαμβάνουμε ότι:

$\mathcal{C} = \sum_{i=1}^{M} N_i = \frac{N_1+...+N_M}{M}.M \approx \mathbb{E}[N].M \overset{(2.8)}{=} \mathcal{O}(\epsilon^{-2})\mathcal{O}\left(\epsilon^{-1/2}\log\left(\epsilon^{-1/2}\right)\right) = \mathcal{O}\left(\epsilon^{-5/2}\log\left(\epsilon^{-1/2}\right)\right)$

και η απόδειξη έχει ολοκληρωθεί. □

Παρατήρηση 2.4.2

Υπενθυμίζουμε στο σημείο αυτό, πως για το πρόβλημά μας οδηγηθήκαμε στην επιλογή προσαρμοσμένου πλήθους προσομοιώσεων, με σκοπό τη μείωση του ολικού κόστους του αρχικού αριθμητικού μας σχήματος. Έχοντας λοιπόν λάβει μια απάντηση για την πολυπλοκότητα της προσαρμοσμένης εμφωλευμένης προσομοίωσης, θα εξετάσουμε τώρα κατά πόσο πράγματι εξασφαλίσαμε τη ζητούμενη μείωση. Παρά το γεγονός πως στη μία περίπτωση το κόστος είναι πλήρως καθορισμένο και στην άλλη προσεγγιστικό, με αποτέλεσμα να μην είναι δυνατή μια απευθείας σύγκριση, εν τέλει μπορούμε να καταλήξουμε σε σημαντικά συμπεράσματα. Συγκεκριμένα, αξιοποιώντας την ανισότητα $\log(x) < x, \forall\, x > 0$ βλέπουμε πως ισχύει ότι:

$\epsilon^{-5/2}\log\left(\epsilon^{-1/2}\right) < \epsilon^{-5/2}\epsilon^{-1/2} = \epsilon^{-3} \Longrightarrow \mathcal{O}\left(\epsilon^{-5/2}\log\left(\epsilon^{-1/2}\right)\right) = \mathcal{O}\left(\epsilon^{-3}\right),$

απ΄ όπου και συμπεραίνουμε πως κάτω από το πλαίσιο στο οποίο εργαστήκαμε η ολική αριθμητική πολυπλοκότητα έχει πράγματι μειωθεί.

2.5 Μια Ακόμα Επιλογή Εσωτερικού Πλήθους Δείγματος

Με βάση την ανάλυση στην οποία προχωρήσαμε στην αμέσως προηγούμενη ενότητα, είδαμε πως έχουμε τη δυνατότητα να μειώσουμε το ολικό υπολογιστικό κόστος της μεθόδου μας επιλέγοντας το πλήθος του εσωτερικού δείγματος μέσω της σχέσης:

$N_i = \left\lceil \min\left(\mathcal{O}(\epsilon^{-1}), \epsilon^{-1/2}\frac{\sigma_i}{d_i}\right)\right\rceil.$

Ουσιαστικά, για την εξαγωγή του συγκεκριμένου αποτελέσματος στηριχτήκαμε στην εύρεση ενός φράγματος για την πιθανότητα η προσθήκη μιας επιπλέον μεταβλητής διακριτοποίησης να μεταβάλλει το πρόσημο του εσωτερικού εκτιμητή μας. Αξίζει ωστόσο να αναφέρουμε πως υπάρχουν και άλλοι τρόποι επιλογής του βέλτιστου δείγματος N_i, $i = 1, \dots, M$, όπως είναι ένας που απορρέει από το κεντρικό οριακό θεώρημα. Συγκεκριμένα, όπως εξετάσαμε στο πρώτο κεφάλαιο, για τον εκτιμητή Monte Carlo της ποσότητας $\mathbb{E}[X|Y]$, τον οποίο συμβολίζουμε με $\overline{\mathbb{E}}[X|Y]$, θα ισχύει ότι:

$\frac{\overline{\mathbb{E}}[X|Y]-\mathbb{E}[X|Y]}{\sqrt{\frac{Var[X|Y]}{N}}} \underset{N\to\infty}{\longrightarrow} N(0,1),$

με αποτέλεσμα για υψηλές τιμές της μεταβλητής N να παίρνουμε πως:

$\mathbb{P}\left(-c \le \frac{\overline{\mathbb{E}}[X|Y]-\mathbb{E}[X|Y]}{\sqrt{\frac{Var[X|Y]}{N}}} \le c\right) \approx \Phi(c) - \Phi(-c) = \Phi(c) - 1 + \Phi(c) = 2\Phi(c) - 1,$

όπου με Φ για ακόμα μια φορά συμβολίζουμε τη συνάρτηση κατανομής πιθανότητας μιας τυπικής κανονικής κατανομής. Έτσι, αν επιλέξουμε μια σταθερά c αρκετά μεγάλη τέτοια ώστε $\Phi(c) \approx 1$ με μεγάλη πιθανότητα λαμβάνουμε ότι:

$$-c \leq \frac{\overline{\mathbb{E}}[X|Y]-\mathbb{E}[X|Y]}{\sqrt{\frac{Var[X|Y]}{N}}} \leq c \Longrightarrow -c\sqrt{\frac{Var[X|Y]}{N}} \leq \overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y] \leq c\sqrt{\frac{Var[X|Y]}{N}}$$ ή αλλιώς:

$$\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| \leq c\sqrt{\frac{Var[X|Y]}{N}} . \quad (2.9)$$

Θεωρώντας τώρα πως η τιμή της μεταβλητής Y ως προς την οποία έχουμε δεσμεύσει είναι η y_i για κάποιο $i = 1, \ldots, M$ και ορίζοντας τις ποσότητες $d_i = |\mathbb{E}[X|Y = y_i]|$ και $\sigma_i^2 = Var[X|Y = y_i]$, με βάση τη σχέση (2.9) παίρνουμε ότι:

$$\left|\overline{\mathbb{E}}[X|Y = y_i] - \mathbb{E}[X|Y = y_i]\right| \leq c_i\sqrt{\frac{\sigma_i^2}{N_i}} ,$$

όπου c_i είναι η αντίστοιχη σταθερά εμπιστοσύνης, εκείνης που είδαμε παραπάνω. Όπως έχουμε όμως ήδη εξηγήσει, για να λάβει ο εσωτερικός εκτιμητής $H\left(\overline{\mathbb{E}}[X|Y = y_i]\right)$ τη σωστή τιμή αρκεί η ποσότητα $\overline{\mathbb{E}}[X|Y = y_i]$ να είναι ομόσημη με την $\mathbb{E}[X|Y = y_i]$ ή διαφορετικά αρκεί να ισχύει πως:

$$\left|\overline{\mathbb{E}}[X|Y = y_i] - \mathbb{E}[X|Y = y_i]\right| \leq |\mathbb{E}[X|Y = y_i]| = d_i .$$

Έτσι τελικά, για να οδηγηθούμε στο επιθυμητό αποτέλεσμα αρκεί να ικανοποιείται η ακόλουθη σχέση:

$$c_i\sqrt{\frac{\sigma_i^2}{N_i}} \leq d_i \Longrightarrow \sqrt{N_i} \geq \frac{c_i}{d_i}\sqrt{\sigma_i^2} \Longrightarrow N_i \geq c_i^2\frac{\sigma_i^2}{d_i^2} .$$

Επιπλέον, κάνοντας χρήση παρόμοιων επιχειρημάτων με αυτά που αναπτύξαμε στην προηγούμενη διαδικασία επιλογής του εσωτερικού πλήθους δείγματος, θα επιθυμούσαμε η τιμή N_i να είναι φυσικός αριθμός και να έχει τάξη μικρότερη ή ίση του ϵ^{-1}, όπου ϵ είναι η σταθερά σφάλματος της μεθόδου μας. Επομένως, καταλήγουμε πως η εκάστοτε τιμή N_i θα δίνεται με τη βοήθεια του τύπου:

$$N_i = \left\lceil \min\left(\mathcal{O}(\epsilon^{-1}), c_i^2\frac{\sigma_i^2}{d_i^2}\right)\right\rceil .$$

Αν τώρα συγκρίνουμε τις δυο εκτιμήσεις τις οποίες λάβαμε παρατηρούμε πως στη δεύτερη έχει μεταβληθεί η δύναμη στην οποία είναι υψωμένη η μεταβλητή $\delta_i = d_i/\sigma_i$ από -1 σε -2 και έχουμε εισαγάγει μια σταθερά εμπιστοσύνης c_i.

Παρατήρηση 2.5.1

Αν εξετάσουμε προσεκτικά τις εκφράσεις που πήραμε για το προσαρμοσμένο πλήθος δείγματος, τόσο στη συγκεκριμένη ενότητα, όσο και στη 2.3, μπορούμε να δούμε πως αυτές εξαρτώνται άμεσα από το εκάστοτε κλάσμα d_i/σ_i. Έτσι, παρά το γεγονός πως η παραπάνω ποσότητα δοθέντος της μεταβλητής Y προσδιορίζεται πλήρως, εν τέλει δε μας είναι γνωστή, όπως φαίνεται καθαρά από την ανάγκη χρήσης του εσωτερικού σχήματος Monte Carlo. Συνεπώς, καλούμαστε για την κάθε τιμή της μεταβλητής Y, να προσεγγίσουμε με κάποιον κατάλληλο αριθμητικό τρόπο την αντίστοιχη τιμή d/σ και εν συνεχεία να αντικαταστήσουμε την εκτίμηση αυτή στις σχέσεις που έχουμε οδηγηθεί. Αξίζει να αναφέρουμε, πως ο συγκεκριμένος τρόπος εργασίας θα καλυφθεί λεπτομερώς στο τέταρτο κεφάλαιο της εργασίας αυτής. Προφανώς οι ιδέες που θα αναπτυχθούν εκεί, μεταφέρονται με φυσιολογικό τρόπο και στο πλαίσιο το οποίο μελετάμε.

3 Εφαρμογή Multilevel Monte Carlo στην Εμφωλευμένη Προσομοίωση

3.1 Εισαγωγή Σχήματος Multilevel Monte Carlo

Μέχρι τώρα, στην προσπάθειά μας να περιορίσουμε το ολικό υπολογιστικό κόστος στην αριθμητική εκτίμηση της τιμής $\eta = \mathbb{P}\left[\mathbb{E}[X|Y] \geq 0\right]$, οδηγηθήκαμε σε μια έκφραση για το πλήθος του εσωτερικού δείγματος, το οποίο διαφοροποιείται για την κάθε τιμή y της μεταβλητής Y που εξετάζουμε. Με άλλα λόγια, για να μειώσουμε το αριθμητικό κόστος της μεθόδου μας εστιάσαμε αποκλειστικά στο εμφωλευμένο δείγμα N και όχι στο εξωτερικό, για το οποίο θεωρήσαμε απλώς ότι $M = \mathcal{O}(\epsilon^{-2})$. Το αποτέλεσμα της λογικής αυτής, όπως είδαμε, είναι για τις τιμές y_i που πληρούν τη συνθήκη $d_i/\sigma_i = o(\epsilon^{1/2})$ να λαμβάνουμε ένα κόστος διακριτοποίησης της τάξης του $\mathcal{O}(\epsilon^{-1})$, το οποίο σε συνδυασμό με την τάξη του M ενδέχεται να δώσει μια υψηλή ολική υπολογιστική πολυπλοκότητα. Αυτό που θα θέλαμε ιδανικά θα ήταν να χρησιμοποιούμε πολύ λίγες τιμές y_i όταν το εσωτερικό κόστος προσομοίωσης είναι υψηλό και πολύ περισσότερες όταν αυτό κυμαίνεται σε σχετικά χαμηλές τιμές. Με βάση το παραπάνω σκεπτικό θα μελετήσουμε κατά πόσο η εφαρμογή μιας μεθόδου Multilevel Monte Carlo, με μεταβλητή διακριτοποίησης τη Heaviside συνάρτηση της προσέγγισης που προκύπτει από το εσωτερικό σχήμα Monte Carlo, θα μπορούσε να οδηγήσει σε μια ικανοποιητική αριθμητική πολυπλοκότητα.

Έτσι, θεωρούμε την πεπερασμένη ακολουθία επιπέδων $l = 0, 1, \ldots, L$, M_l το πλήθος των τιμών y_i στο l-οστό επίπεδο και N_l το πλήθος του εσωτερικού δείγματος στο επίπεδο αυτό. Για να είμαστε συνεπείς με προγενέστερη ανάλυσή μας, όπου θεωρήσαμε πως το κόστος της μεταβλητής διακριτοποίησης αυξάνεται καθώς μεταβαίνουμε από χαμηλότερα επίπεδα σε πιο υψηλά, θα δεχτούμε πως η ακολουθία N_l είναι αύξουσα. Για να προχωρήσουμε την ανάλυσή μας, στο $l-$οστό επίπεδο μελέτης ορίζουμε $\{y_j\}_{j=1}^{M_l}$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την κατανομή της Y. Τότε, για την κάθε τιμή y_j δημιουργούμε ένα ανεξάρτητο και ισόνομο δείγμα της μεταβλητής X δοθέντος $Y = y_j$ πλήθους N_l, με $l = 0, 1, \ldots, L$, τις οποίες συμβολίζουμε ως $\{x_i^{l,y_j}\}_{i=\{1,\ldots,N_l\}}$. Αξίζει στο σημείο αυτό να αναφέρουμε πως ο παραπάνω συμβολισμός χρησιμοποιήθηκε ώστε να φαίνεται καθαρά η τιμή της μεταβλητής Y ως προς την οποία δεσμεύουμε αλλά και το επίπεδο στο οποίο βρισκόμαστε. Τέλος, ορίζουμε τη μεταβλητή μας διακριτοποίησης μέσω της σχέσης:

$$H\left(\overline{\mathbb{E}}^{l,y_j}[X]\right) := H\left(\frac{1}{N_l}\sum_{i=1}^{N_l} x_i^{l,y_j}\right),$$

όπου $j = 1, \ldots, M_l$ και $l = 0, \ldots, L$. Βλέπουμε όμως ότι το $\{\overline{\mathbb{E}}^{l,y_j}[X]\}_{j=\{1,\ldots,M_l\}}$ είναι ανεξάρτητο και ισόνομο δείγμα της μεταβλητής $\overline{\mathbb{E}}^{l}[X|Y]$, συνεπώς από το νόμο των μεγάλων αριθμών παίρνουμε ότι:

$$\frac{1}{M_l}\sum_{j=1}^{M_l} H\left(\overline{\mathbb{E}}^{l,y_j}[X]\right) \xrightarrow[M_l\to\infty]{\mathbb{P}} \mathbb{E}\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)\right] \equiv \mathbb{E}\left[\overline{G}^{l}\right].$$

Προφανώς, στα υψηλά επίπεδα μελέτης ισχύει ότι $\overline{\mathbb{E}}^{l}[X|Y] \approx \mathbb{E}[X|Y]$, απ' όπου και καταλήγουμε πως:

$$\frac{1}{M_l}\sum_{j=1}^{M_l} H\left(\overline{\mathbb{E}}^{l,y_j}[X]\right) \approx \mathbb{E}[H(\mathbb{E}[X|Y])].$$

Με βάση τα παραπάνω, αλλά και αυτά που μελετήσαμε στο πρώτο κεφάλαιο, ορίζουμε τον εκτιμητή Multilievel Monte Carlo της ποσότητας $\eta = \mathbb{E}[H(\mathbb{E}[X|Y])] \equiv \mathbb{E}[G]$ μέσω της σχέσης:

$$\widetilde{I} = \frac{1}{M_0}\sum_{j=1}^{M_0} \overline{G}_j^{0} + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{j=1}^{M_l}\left(\overline{G}_j^{l} - \overline{G}_j^{l-1}\right) \text{ και εξ' ορισμού προκύπτει ότι:}$$

$$\widetilde{I} = \frac{1}{M_0}\sum_{j=1}^{M_0} H\left(\overline{\mathbb{E}}^{0,y_j}[X]\right) + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{j=1}^{M_l}\left(H\left(\overline{\mathbb{E}}^{l,y_j}[X]\right) - H\left(\overline{\mathbb{E}}^{l-1,y_j}[X]\right)\right),$$

όπου όπως ήδη αναφέραμε:

$\overline{\mathbb{E}}^{l,y_j}[X] = \frac{1}{N_l}\sum\limits_{i=1}^{N_l} x_i^{l,y_j}$, $\forall\ l = 0, 1, \ldots, L$ και $\forall\ j = 1, \ldots, M_l$.

3.2 Μελέτη Ολικού Υπολογιστικού Κόστους

Στην ενότητα αυτή θα επιχειρήσουμε να μελετήσουμε το ολικό υπολογιστικό κόστος της νέας μας μεθόδου στηριζόμενοι στο θεώρημα ανάλυσης πολυπλοκότητας που εξετάσαμε στο πρώτο κεφάλαιο αυτής της εργασίας. Υπενθυμίζουμε πως το θεώρημα αυτό έχει τρεις βασικές προϋποθέσεις. Η πρώτη προϋπόθεση αφορά το σφάλμα διακριτοποίησης στο κάθε επίπεδο, το οποίο θα πρέπει να φθίνει με ταχύ ρυθμό καθώς μεταβαίνουμε σε υψηλότερα επίπεδα l. Η δεύτερη προϋπόθεση σχετίζεται με τη διασπορά της διαφοράς δύο διαδοχικών μεταβλητών διακριτοποίησης η οποία θα πρέπει να κατευθύνεται ικανοποιητικά γρήγορα προς την τιμή 0 καθώς $l \to L$. Τέλος, θα πρέπει το κόστος παραγωγής της εκάστοτε μεταβλητής διακριτοποίησης να αυξάνεται καθώς μετακινούμαστε σε υψηλότερα επίπεδα μελέτης, το οποίο στην περίπτωσή μας μεταφράζεται ως η μεταβλητή N_l είναι αύξουσα. Καταρχάς, αξιοποιώντας για μια ακόμα φορά το θεώρημα 2.2.2 και στηριζόμενοι σε ανάλογες υποθέσεις με αυτές που αναφέρονται εκεί μπορούμε να πάρουμε πως ισχύει ότι:

$$\left|\mathbb{E}\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right) - H(\mathbb{E}[X|Y])\right]\right| = \mathcal{O}\left(N_l^{-1}\right) \overset{\text{συμ.}}{\Longrightarrow} \left|\mathbb{E}\left[\overline{G}^{l} - G\right]\right| = \mathcal{O}\left(N_l^{-1}\right),$$

ή αλλιώς καθώς $N_l \to \infty$ θα $\exists\ c_1 > 0$ ώστε:

$$\left|\mathbb{E}\left[\overline{G}^{l} - G\right]\right| \le c_1 N_l^{-1}\ . \quad (3.1)$$

Επιπλέον, εύκολα αντιλαμβανόμαστε πως το κόστος της εκάστοτε μεταβλητής διακριτοποίησης είναι τάξης N_l, δηλαδή έχουμε ότι:

$C_l = \mathcal{O}(N_l)$, συνεπώς για μεγάλες τιμές του πλήθους N_l θα $\exists\ c_3 > 0$ ώστε $C_l \le c_3 N_l$. (3.2)

Για να ελέγξουμε την ισχύ της δεύτερης συνθήκης θα στηριχτούμε σε μια σειρά από λήμματα τα οποία και θα αποδείξουμε αναλυτικά στο κομμάτι της εργασίας που ακολουθεί.

Λήμμα 3.2.1

Έστω δ μια μη αρνητική και μονοδιάστατη τυχαία μεταβλητή, η οποία έχει συνάρτηση πυκνότητας πιθανότητας ρ. Υποθέτουμε επιπλέον πως η παραπάνω συνάρτηση πυκνότητας πιθανότητας είναι τοπικά φραγμένη κοντά στο 0, δηλαδή $\exists$ σταθερές $\rho_0 > 0$ και $\delta_0 > 0$ ώστε να ικανοποιείται η ανισοτική σχέση:

$\rho(\delta) \le \rho_0$, $\forall\ \delta \in [0, \delta_0]$. (3.3)

Τότε, για κάθε θετική και μη αύξουσα συνάρτηση α ισχύει ότι:

$\int_0^\infty \alpha(\delta)\rho(\delta)d\delta \le \rho_0 \int_0^\infty \alpha(\delta)d\delta + \alpha(\delta_0)$.

Απόδειξη: Έστω α μια θετική και μη αύξουσα πραγματική συνάρτηση. Τότε για το άνω φράγμα του ζητούμενου ολοκληρώματος λαμβάνουμε ότι:

$$
\begin{aligned}
\int_0^\infty \alpha(\delta)\rho(\delta)d\delta &= \int_0^{\delta_0} \alpha(\delta)\rho(\delta)d\delta + \int_{\delta_0}^\infty \alpha(\delta)\rho(\delta)d\delta \\
&\overset{(3.3)}{\leq} \int_0^{\delta_0} \alpha(\delta)\rho_0 d\delta + \int_{\delta_0}^\infty \alpha(\delta)\rho(\delta)d\delta \\
&\overset{\alpha>0}{\leq} \rho_0 \int_0^\infty \alpha(\delta)d\delta + \int_{\delta_0}^\infty \alpha(\delta)\rho(\delta)d\delta \\
&\leq \rho_0 \int_0^\infty \alpha(\delta)d\delta + \int_{\delta_0}^\infty \alpha(\delta_0)\rho(\delta)d\delta\ .
\end{aligned}
$$

Αξίζει να αναφέρουμε πως για την εξαγωγή της τελευταίας σχέσης στηριχτήκαμε στο δεδομένο πως η συνάρτηση α είναι μη αύξουσα. Τελικά λοιπόν βλέπουμε πως ισχύει ότι:

$\int_0^\infty \alpha(\delta)\rho(\delta)d\delta \leq \rho_0 \int_0^\infty \alpha(\delta)d\delta + \alpha(\delta_0) \int_{\delta_0}^\infty \rho(\delta)d\delta \leq \rho_0 \int_0^\infty \alpha(\delta)d\delta + \alpha(\delta_0) \int_0^\infty \rho(\delta)d\delta,$

όπου λόγω του γεγονότος πως η ρ είναι συνάρτηση πυκνότητας πιθανότητας προκύπτει ότι:

$\int_0^\infty \alpha(\delta)\rho(\delta)d\delta \leq \rho_0 \int_0^\infty \alpha(\delta)d\delta + \alpha(\delta_0)$

και η απόδειξη έχει ολοκληρωθεί. □

Λήμμα 3.2.2

Έστω $b > 0$ μια σταθερά και μια τιμή $q > 1$. Τότε ισχύει η ακόλουθη ταυτότητα:

$\int_0^\infty \min(1, bx^{-q})\, dx = \frac{qb^{1/q}}{q-1}.$

Απόδειξη: Σε ένα πρώτο στάδιο, θα προσδιορίσουμε την τιμή της μεταβλητής x για την οποία ισχύει ότι $\min(1, bx^{-q}) = 1$. Έτσι, για $x > 0$ εύκολα παίρνουμε πως:

$1 \leq bx^{-q} \implies x^q \leq b \implies x \leq b^{1/q}.$

Με βάση το τελευταίο αποτέλεσμα έχουμε ότι:

$$
\begin{aligned}
\int_0^\infty \min\left(1, bx^{-q}\right) dx &= \int_0^{b^{1/q}} \min\left(1, bx^{-q}\right) dx + \int_{b^{1/q}}^\infty \min\left(1, bx^{-q}\right) dx \\
&= \int_0^{b^{1/q}} 1\, dx + \int_{b^{1/q}}^\infty bx^{-q}\, dx \\
&= b^{1/q} + \left.\frac{bx^{1-q}}{1-q}\right|_{b^{1/q}}^\infty \\
&= b^{1/q} + \frac{b^{1/q}}{q-1} \\
&= (q - \cancel{1} + \cancel{1})\frac{b^{1/q}}{q-1} \\
&= \frac{qb^{1/q}}{q-1},
\end{aligned}
$$

συνεπώς οδηγηθήκαμε στο επιθυμητό αποτέλεσμα. □

Λήμμα 3.2.3

Έστω X μια μονοδιάστατη τυχαία μεταβλητή, η οποία δεσμεύεται από την τιμή μιας πολυδιάστατης μεταβλητής Y. Θεωρούμε επιπλέον τις ποσότητες $d = |\mathbb{E}[X|Y]|$, $\sigma^2 = Var[X|Y]$ και ορίζουμε τη μη αρνητική τυχαία μεταβλητή $\delta = d/\sigma$, η οποία υποθέτουμε πως έχει συνάρτηση πυκνότητας πιθανότητας ρ. Τέλος, δεχόμαστε πως η παραπάνω συνάρτηση πυκνότητας πιθανότητας είναι τοπικά φραγμένη κοντά στο 0. Τότε, για τη διασπορά του εσωτερικού εκτιμητή της ποσότητας $\eta = \mathbb{E}[H(\mathbb{E}[X|Y])]$ θα ισχύει ότι:

$$Var\left[H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \leq \mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2\right] = \mathcal{O}(N^{-1/2}),$$

όπου N είναι το πλήθος του εσωτερικού δείγματος.

Απόδειξη: Καταρχάς, για οποιαδήποτε τυχαία μεταβλητή X ισχύει πως:

$$Var[X] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2 \Longrightarrow Var[X] \leq \mathbb{E}\left[X^2\right],$$

με αποτέλεσμα για την $A := H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)$ να λαμβάνουμε ότι:

$$Var\left[H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \leq \mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2\right] . \quad (3.4)$$

Παρατηρούμε τώρα πως η συνάρτηση Heaviside λαμβάνει αποκλειστικά τις τιμές 0 και 1, συνεπώς το ίδιο θα ισχύει και για τη μεταβλητή $\left|H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right|$. Έτσι, εξ' ορισμού παίρνουμε ότι:

$$\begin{aligned}
\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2 |Y\right] &= \mathbb{E}\left[\left|H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right|^2 |Y\right] \\
&= \mathbb{P}\left[\left|H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right| = 1|Y\right].1^2 + 0 \\
&= \mathbb{P}\left[\left|H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right| = 1|Y\right] . \quad (3.5)
\end{aligned}$$

Για να απλοποιήσουμε την παραπάνω έκφραση αρκεί να κατανοήσουμε τι ακριβώς μας δηλώνει. Συγκεκριμένα, παρατηρούμε πως για τα ενδεχόμενα $\omega \in \Omega$ για τα οποία $\left|H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right| = 1$ θα πρέπει υποχρεωτικά η μία από τις ποσότητες $\overline{\mathbb{E}}[X|Y]$, $\mathbb{E}[X|Y]$ να είναι θετική και η άλλη αρνητική. Επιπλέον, γνωρίζουμε πως για τη διαφορά δύο ετερόσημων αριθμών ισχύει ότι:

$$|d_1 - d_2| \geq \max(|d_1|,\ |d_2|),$$

με αποτέλεσμα για τα $\omega \in \Omega$ που ικανοποιούν την παραπάνω συνθήκη να καταλήγουμε πως:

$$\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| \geq |\mathbb{E}[X|Y]| \Longrightarrow \left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| \geq d .$$

Με βάση την παραπάνω παρατήρηση αλλά και τη σχέση (3.5) παίρνουμε τελικά ότι:

$$\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2 |Y\right] \leq \mathbb{P}\left[\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| \geq d|Y\right] .$$

Αξιοποιώντας τώρα την ανισότητα του Chebysev για την τυχαία μεταβλητή $\overline{\mathbb{E}}[X|Y]$, η οποία δοθέντος Y έχει μέση τιμή ίση με $\mathbb{E}[X|Y]$, βλέπουμε ότι:

$$\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2 |Y\right] \leq \mathbb{P}\left[\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| \geq d|Y\right] \leq d^{-2} Var\left[\overline{\mathbb{E}}[X|Y]\,|Y\right].$$

Προφανώς όμως η πιθανότητα που εμφανίζεται στην παραπάνω σχέση φράσσεται και από την τιμή

1, συνεπώς ισχύει ότι:

$\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right)^2|Y\right]\leq \min\left(1, d^{-2}Var\left[\overline{\mathbb{E}}[X|Y]|Y\right]\right)$.

Επιπλέον, παρατηρούμε ότι:

$$\begin{aligned}
d^{-2}Var\left[\overline{\mathbb{E}}[X|Y]|Y\right] &= d^{-2}Var\left[\frac{1}{N}\sum_{j=1}^{N}x_j^Y|Y\right] \\
&= d^{-2}\frac{1}{N^2}Var\left[\sum_{j=1}^{N}x_j^Y|Y\right] \\
&= \frac{d^{-2}}{N^2}\sum_{j=1}^{N}Var\left[x_j^Y|Y\right] \\
&= \frac{d^{-2}}{N^2}\sum_{j=1}^{N}Var\left[x^Y|Y\right] \\
&= \frac{d^{-2}}{N}Var\left[X|Y\right] \\
&= \frac{d^{-2}}{N}\sigma^2 \\
&= \delta^{-2}N^{-1} \ ,
\end{aligned}$$

οπότε προκύπτει τελικά πως:

$\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right)^2|Y\right]\leq \min\left(1, d^{-2}Var\left[\overline{\mathbb{E}}[X|Y]|Y\right]\right)=\min\left(1,\delta^{-2}N^{-1}\right)$.

Εφαρμόζοντας μέση τιμή στην παραπάνω σχέση έχουμε ότι:

$\mathbb{E}\left[\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right)^2|Y\right]\right]\leq \mathbb{E}\left[\min\left(1,\delta^{-2}N^{-1}\right)\right]$ ή ισοδύναμα:

$\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right)^2\right]\leq \mathbb{E}\left[\min\left(1,\delta^{-2}N^{-1}\right)\right]$. (3.6)

Βλέπουμε όμως από τα δεδομένα μας πως η τυχαία μεταβλητή δ έχει συνάρτηση πυκνότητας πιθανότητας ρ, συνεπώς για τη συνάρτηση $\alpha(\delta) := \min\left(1,\delta^{-2}N^{-1}\right)$ θα ισχύει ότι:

$\mathbb{E}\left[\min\left(1,\delta^{-2}N^{-1}\right)\right]=\mathbb{E}[\alpha(\delta)]=\int_0^\infty \alpha(\delta)\rho(\delta)d\delta$.

Αξιοποιώντας για μια ακόμα φορά τα δεδομένα μας παρατηρούμε πως η συνάρτηση πυκνότητας πιθανότητας ρ είναι φραγμένη κοντά στο 0, δηλαδή $\exists$ σταθερές $\rho_0 > 0$ και $\delta_0 > 0$ ώστε $\rho(\delta) \leq \rho_0$ $\forall\ \delta \in [0,\delta_0]$. Επιπλέον, εύκολα γίνεται αντιληπτό πως η συνάρτηση α με τον τρόπο που ορίστηκε είναι θετική και μη αύξουσα. Με βάση λοιπόν το λήμμα 3.2.1 λαμβάνουμε ότι:

$\mathbb{E}\left[\min\left(1,\delta^{-2}N^{-1}\right)\right]=\int_0^\infty \alpha(\delta)\rho(\delta)d\delta\leq\rho_0\int_0^\infty\alpha(\delta)d\delta+\alpha(\delta_0)$ ή με άλλα λόγια

$\mathbb{E}\left[\min\left(1,\delta^{-2}N^{-1}\right)\right]\leq\rho_0\int_0^\infty\min\left(1,\delta^{-2}N^{-1}\right)d\delta+\min\left(1,\delta_0^{-2}N^{-1}\right)\leq\rho_0\int_0^\infty\min\left(1,\delta^{-2}N^{-1}\right)d\delta$
$+\ \delta_0^{-2}N^{-1}$. (3.7)

Στο σημείο αυτό όμως μπορούμε να δούμε πως το ολοκλήρωμα της σχέσης (3.7) υπολογίζεται με βάση το λήμμα 3.2.2, γεγονός που μας οδηγεί στο αποτέλεσμα:

$$\mathbb{E}\left[\min\left(1,\delta^{-2}N^{-1}\right)\right] \leq 2\rho_0 N^{-1/2} + \delta_0^{-2}N^{-1} \overset{N\to\infty}{\leq} \left(2\rho_0 + \delta_0^{-2}\right) N^{-1/2} \equiv cN^{-1/2}\ . \quad (3.8)$$

Τέλος, με τη βοήθεια των σχέσεων (3.4),(3.6) και (3.8) καταλήγουμε πως καθώς $N \to \infty$ ισχύει ότι:

$$Var\left[H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \leq cN^{-1/2} \Longrightarrow Var\left[H\left(\overline{\mathbb{E}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] = \mathcal{O}\left(N^{-1/2}\right)$$

και η απόδειξη έχει ολοκληρωθεί. □

Λήμμα 3.2.4

Για οποιεσδήποτε δύο τυχαίες μεταβλητές V, W ικανοποιείται η παρακάτω ανισοτική σχέση:

$Var[V-W] \leq 2Var[V] + 2Var[W]$.

Απόδειξη: Σε ένα πρώτο στάδιο, ισχυριζόμαστε πως για τις τυχούσες μεταβλητές V, W ισχύει ότι:

$|Cov(V,W)| \leq \sqrt{Var[V]}\sqrt{Var[W]}$.

Πράγματι, για κάθε πραγματικό αριθμό λ έχουμε πως:

$$\begin{aligned} Var[V-\lambda W] \geq 0 &\Longrightarrow Cov(V-\lambda W, V-\lambda W) \geq 0 \\ &\Longrightarrow \lambda^2 Cov(W,W) - 2\lambda Cov(V,W) + Cov(V,V) \geq 0 \\ &\Longrightarrow Var[W]\lambda^2 - 2Cov(V,W)\lambda + Var[V] \geq 0,\ \forall\ \lambda \in \mathbb{R}. \end{aligned}$$

Επειδή όμως ο μεγιστοβάθμιος όρος του τριωνύμου που εμφανίζεται είναι θετικός, για να έχει ισχύ η παραπάνω σχέση θα πρέπει να ικανοποιείται η συνθήκη:

$$\Delta \leq 0 \Longrightarrow 4\left(Cov(V,W)\right)^2 - 4Var[W]Var[V] \leq 0 \Longrightarrow |Cov(V,W)| \leq \sqrt{Var[V]}\sqrt{Var[W]}. \quad (3.9)$$

Για τη διασπορά τώρα της διαφοράς των δύο τυχαίων μεταβλητών παίρνουμε εξ' ορισμού ότι:

$$Var[V-W] = Var[V] - 2Cov(V,W) + Var[W] \overset{(3.9)}{\leq} Var[V] + Var[W] + 2\sqrt{Var[V]}\sqrt{Var[W]}.$$

Ισχύει όμως και η προφανής ανισότητα:

$\left(\sqrt{Var[V]} - \sqrt{Var[W]}\right)^2 \geq 0 \Longrightarrow 2\sqrt{Var[V]}\sqrt{Var[W]} \leq Var[V] + Var[W]$, συνεπώς:

$Var[V-W] \leq Var[V] + Var[W] + Var[V] + Var[W] = 2Var[V] + 2Var[W]$. □

Θεώρημα 3.2.5

Έστω X μια μονοδιάστατη τυχαία μεταβλητή, η οποία δεσμεύεται από την τιμή μιας πολυδιάστατης μεταβλητής Y. Έστω επιπλέον $\widetilde{I}$ ο Multilevel Monte Carlo εκτιμητής της ποσότητας $\eta = \mathbb{E}[H(\mathbb{E}[X|Y])]$. Τέλος, υποθέτουμε πως στο κάθε επίπεδο l στηριζόμαστε σε ντετερμινιστικό πλήθος N_l, το οποίο αυξάνει καθώς μεταβαίνουμε σε υψηλότερα επίπεδα μελέτης. Τότε, για τη διασπορά της διαφοράς δυο διαδοχικών μεταβλητών διακριτοποίησης λαμβάνουμε ότι:

$Var\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)\right]=\mathcal{O}\left(N_{l-1}^{-1/2}\right).$

<u>**Απόδειξη**</u>: Σε ένα πρώτο στάδιο, ορίζουμε τις τυχαίες μεταβλητές:

$V:=H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)$ και $W:=H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)$

καθώς και τη μεταβλητή A η οποία δίνεται με τη βοήθεια του τύπου:

$A:=H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right).$

Τότε, για τη διασπορά της τελευταίας βλέπουμε ότι:

$$\begin{aligned}
Var\left[A\right]&=Var\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)\right]\\
&=Var\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)+H\left(\mathbb{E}[X|Y]\right)-H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)\right]\\
&=Var\left[\left(H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right)-\left(H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right)\right]\\
&\overset{\text{ορ.}}{=}Var\left[V-W\right]\\
&\overset{3.2.4}{\leq}2Var[V]+2Var[W]\\
&=2Var\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right]+2Var\left[H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right].
\end{aligned}$$

Εφόσον τώρα από την υπόθεσή μας το εσωτερικό πλήθος δείγατος στο κάθε επίπεδο είναι ντετερμινιστικό αξιοποιώντας το λήμμα 3.2.3 και το τελευταίο μας αποτέλεσμα παίρνουμε πως:

$Var\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)\right]=\mathcal{O}\left(N_{l}^{-1/2}\right)+\mathcal{O}\left(N_{l-1}^{-1/2}\right),$

ενώ λόγω του γεγονότος πως η N_l είναι αύξουσα ακολουθία λαμβάνουμε τελικά ότι:

$Var\left[H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)-H\left(\overline{\mathbb{E}}^{l-1}[X|Y]\right)\right]=\mathcal{O}\left(N_{l-1}^{-1/2}\right).\quad(3.10)$ □

<u>Παρατήρηση 3.2.6</u>

Με βάση τα παραπάνω, είμαστε πλέον σε θέση να μελετήσουμε την πολυπλοκότητα της μεθόδου μας για μια ντετερμινιστική και αύξουσα ακολουθία N_l, όπου $l=0,1,\ldots,L$. Συγκεκριμένα, αν θεωρήσουμε πως $N_l=N_0 2^l$ για τη μεταβλητή διακριτοποίησης $\overline{G}^l=H\left(\overline{\mathbb{E}}^{l}[X|Y]\right)$ και για την εξεταζόμενη μεταβλητή $G=H\left(\mathbb{E}[X|Y]\right)$ από τις σχέσεις (3.1), (3.2) και (3.10) έχουμε ότι:

- $\left|\mathbb{E}\left[\overline{G}^l-G\right]\right|\leq c_1N_l^{-1}\Longrightarrow\left|\mathbb{E}\left[\overline{G}^l-G\right]\right|\leq\widetilde{c}_1 2^{-1.l}$, όπου $\widetilde{c}_1=N_0^{-1}c_1$
- $Var\left[\overline{G}^l-\overline{G}^{l-1}\right]=\mathcal{O}\left(N_{l-1}^{-1/2}\right)\Longrightarrow Var\left[\overline{G}^l-\overline{G}^{l-1}\right]\leq c_2N_{l-1}^{-1/2}$, συνπώς προκύπτει ότι:

$Var\left[\overline{G}^l-\overline{G}^{l-1}\right]\leq c_2\left(N_0 2^{l-1}\right)^{-1/2}=\sqrt{2}c_2N_0^{-1/2}2^{-\frac{1}{2}l}\equiv\widetilde{c}_2 2^{-\frac{1}{2}l}$, όπου:

$\widetilde{c}_2=\sqrt{2}c_2N_0^{-1/2}$

- $C_l\leq c_3N_l\Longrightarrow C_l\leq c_3N_0 2^{1.l}\Longrightarrow C_l\leq\widetilde{c}_3 2^{1.l}$, όπου $\widetilde{c}_3=N_0c_3$.

Έτσι, παρατηρούμε πως για τις σταθερές $\widetilde{c}_1$, $\widetilde{c}_2$ και $\widetilde{c}_3$ και για τις τιμές $a = \gamma = 1$, $\beta = 1/2$ (προφανώς τότε $a \geq \min(\beta, \gamma)/2$) ικανοποιούνται οι υποθέσεις του θεωρήματος 1.3.1 από το οποίο λαμβάνουμε ότι:

$$\mathcal{C} = \mathcal{O}\left(\epsilon^{-2-(\gamma-\beta)/a}\right) \Longrightarrow \mathcal{C} = \mathcal{O}\left(\epsilon^{-2-1/2}\right) \Longrightarrow \mathcal{C} = \mathcal{O}\left(\epsilon^{-5/2}\right), \text{ μιας και } \beta = 1/2 < 1 = \gamma.$$

3.3 Προσαρμοσμένο Δείγμα στη Μέθοδο Multilevel Monte Carlo Μαθηματικό Υπόβαθρο

Μέχρι τώρα, εξετάσαμε το πρόβλημα αριθμητικής εκτίμησης της ποσότητας $\eta = \mathbb{E}[H(\mathbb{E}[X|Y])]$ μέσα από δύο διαφορετικές προσεγγίσεις. Στην πρώτη προσέγγιση επιλέγαμε το εσωτερικό πλήθος δείγματος N_i, $i = 1\ldots, M$, με βάση τις υπάρχουσες κάθε φορά ανάγκες ή με άλλα λόγια με βάση την τάξη της ποσότητας $\delta_i = d_i/\sigma_i$. Όπως είδαμε αναλυτικά, μέσα από την παραπάνω διαδικασία οδηγούμαστε σε ένα προσεγγιστικό αριθμητικό κόστος της μεθόδου της τάξης του $\epsilon^{-5/2}\log\left(\epsilon^{-1/2}\right)$, όπου ϵ είναι η ακρίβεια που επιθυμούμε να επιτύχουμε. Στη δεύτερη προσέγγιση, εφαρμόζουμε ένα αριθμητικό σχήμα Multilevel Monte Carlo και θεωρούμε πως το πλήθος του εσωτερικού δείγματος είναι μια αύξουσα και ντετερμινιστική ακολουθία. Το αποτέλεσμα της λογικής αυτής είναι για μια επιλογή $N_l = \mathcal{O}\left(2^l\right)$ να λαμβάνουμε μια αριθμητική πολυπλοκότητα της τάξης του $\epsilon^{-5/2}$. Αυτό που θα επιχειρήσουμε να κάνουμε στη συγκεκριμένη ενότητα, είναι να παραθέσουμε τα κατάλληλα μαθηματικά εργαλεία, στα οποία και θα στηριχτούμε με σκοπό να συνδυάσουμε τις δύο παραπάνω προσεγγίσεις και να οδηγηθούμε στη μείωση της ολικής πολυπλοκότητας. Προτού ξεκινήσουμε τη μελέτη μας θα αναφέρουμε την ακόλουθη υπόθεση:

Υπόθεση 3.3.1

Έστω X μια μονοδιάστατη τυχαία μεταβλητή, η οποία δεσμεύεται από την τιμή μιας πολυδιάστατης μεταβλητής Y. Θεωρούμε τότε πως $\exists$ τιμή $2 < q < \infty$ ώστε η κανονικοποιημένη και κεντραρισμένη q-οστή ροπή της μεταβλητής X, δοθέντος Y, να είναι ομοιόμορφα φραγμένη για όλες τις τιμές της μεταβλητής Y, δηλαδή:

$\kappa_q = \sup\limits_y \mathbb{E}\left[(Var[X|Y])^{-q}|X(Y) - \mathbb{E}[X|Y]|^q|Y = y\right] < +\infty$ ή εισάγοντας κατάλληλο συμβολισμό:

$$\sup_y \mathbb{E}\left[\sigma^{-q}|X(Y) - \mathbb{E}[X|Y]|^q|Y = y\right] < +\infty.$$

Προχωράμε τώρα στη διατύπωση του βασικού αποτελέσματος της συγκεκριμένης ενότητας:

Λήμμα 3.3.2

Έστω $\overline{z}_N$ ο δειγματικός μέσος μιας ανεξάρτητης και ισόνομης ακολουθίας $(z_i)_{i=\{1,\ldots,N\}}$, που φέρει την κατανομή μιας μεταβλητής Z με μέση τιμή 0 και πεπερασμένη q-οστή ροπή για $q \geq 2$. Τότε $\forall\ z > 0$ θα $\exists$ σταθερά C_q, εξαρτώμενη μόνο από την τιμή q, τέτοια ώστε:

- $\mathbb{E}\left[|\overline{z}_N|^q\right] \leq C_q N^{-q/2}\mathbb{E}\left[|Z|^q\right]$
- $\mathbb{P}\left[|\overline{z}_N| > z\right] \leq \min\left(1, C_q z^{-q} N^{-q/2}\mathbb{E}\left[|Z|^q\right]\right)$.

Για να μπορέσουμε να αποδείξουμε το παραπάνω λήμμα θα στηριχτούμε σε δύο προτάσεις, οι οποίες θα διατυπωθούν στη συνέχεια αυτής της ενότητας. Ειδικότερα, έχουμε ότι:

Πρόταση 3.3.3(Burkholder, Gundy)

Έστω $z_1, z_2, \dots$ μια ανεξάρτητη και ισόνομη ακολουθία τυχαίων μεταβλητών με μέση τιμή ίση με 0 και Φ μια μη αρνητική συνάρτηση που ικανοποιεί τις παρακάτω ιδιότητες:

- $\exists$ μη αρνητική και μετρήσιμη συνάρτηση ϕ, ορισμένη στο $(0, \infty)$, ώστε:

$\Phi(b) = \int_0^b \phi(\lambda) d\lambda$, $\forall\ b \in [0, \infty]$,

η οποία πληροί τη συνθήκη:

$\phi(2\lambda) \leq \widetilde{c}\phi(\lambda)$, $\forall\ \lambda > 0$, για κάποια σταθερά $\widetilde{c} > 0$

- $\Phi(1) < +\infty$.

Αν επιπλέον η συνάρτηση Φ είναι κυρτή, τότε $\forall\ n \geq 1$ θα ισχύει ότι:

$$c\mathbb{E}\left[\Phi\left(\left(\sum_{i=1}^{n} z_i^2\right)^{1/2}\right)\right] \leq \mathbb{E}\left[\Phi\left(\left|\sum_{i=1}^{n} z_i\right|\right)\right] \leq C\mathbb{E}\left[\Phi\left(\left(\sum_{i=1}^{n} z_i^2\right)^{1/2}\right)\right],$$

όπου οι τιμές των σταθερών c, C εξαρτώνται αποκλειστικά από τη συνάρτηση Φ.

Είναι σημαντικό να αναφέρουμε πως η παραπάνω πρόταση μαζί με την αντίστοιχη απόδειξή της εμπεριέχονται στη δημοσίευση [5].

Πρόταση 3.3.4

Έστω $(z_i)_{i=\{1,\dots,N\}}$ μια μη αρνητική ακολουθία και δυο θετικές τιμές p, r με $p \leq r$. Τότε ικανοποιείται η ακόλουθη ανισοτική σχέση:

$$\left(\frac{1}{N}\sum_{i=1}^{N} z_i^p\right)^{1/p} \leq \left(\frac{1}{N}\sum_{i=1}^{N} z_i^r\right)^{1/r}.$$

Απόδειξη: Παρατηρούμε ότι στην περίπτωση που $p = r$ η ζητούμενη σχέση ισχύει με τετριμμένο τρόπο, οπότε θα προχωρήσουμε τη μελέτη μας θεωρώντας πως ισχύει αυστηρά η ανισοτική σχέση $p < r$. Σε ένα αρχικό στάδιο, θα εξετάσουμε τη μονοτονία της συνάρτησης $f(x) = x^{p/r}$, προχωρώντας στην παραδοχή πως η μεταβλητή x λαμβάνει θετικές τιμές. Έτσι λοιπόν παίρνουμε πως ισχύει ότι:

$f(x) = x^{p/r} \Longrightarrow f'(x) = \frac{p}{r}x^{p/r-1} \Longrightarrow f''(x) = \frac{p}{r}\left(\frac{p}{r} - 1\right)x^{p/r-2}$.

Βλέπουμε όμως από τα δεδομένα μας πως ικανοποιείται η συνθήκη:

$p < r \Longrightarrow \frac{p}{r} < 1 \Longrightarrow \frac{p}{r} - 1 < 0$,

με αποτέλεσμα να καταλήγουμε πως η συνάρτηση f είναι κοίλη στο διάστημα μελέτης που εξετάζουμε. Έτσι, λαμβάνοντας υπόψη μας το γεγονός πως η ακολουθία $(z_i)_{i=\{1,\dots,N\}}$ είναι μη αρνητική και αξιοποιώντας την ανισότητα του Jensen για κοίλες συναρτήσεις μπορούμε να πάρουμε πως ισχύει ότι:

$$\left(\frac{1}{N}\sum_{i=1}^{N} z_i^r\right)^{p/r} \overset{\text{ορ.}}{=} f\left(\frac{1}{N}\sum_{i=1}^{N} z_i^r\right) \geq \frac{1}{N}\sum_{i=1}^{N} f\left(z_i^r\right) \overset{\text{ορ.}}{=} \frac{1}{N}\sum_{i=1}^{N} z_i^p.$$

Συνεπώς, με βάση τα παραπάνω δείξαμε τελικά ότι:

$$\left(\frac{1}{N}\sum_{i=1}^{N} z_i^r\right)^{p/r} \geq \frac{1}{N}\sum_{i=1}^{N} z_i^p \Longrightarrow \left(\frac{1}{N}\sum_{i=1}^{N} z_i^r\right)^{1/r} \geq \left(\frac{1}{N}\sum_{i=1}^{N} z_i^p\right)^{1/p}. \qquad \square$$

Απόδειξη λήμματος: Ορίζουμε αρχικά τη συνάρτηση $\Phi(x) = x^q$, θεωρώντας πως $x \geq 0$, και ισχυριζόμαστε ότι αυτή ανήκει στην κλάση των συναρτήσεων που είδαμε στην πρόταση 3.3.3. Για να δεχτούμε το συγκεκριμένο ισχυρισμό αρκεί να δείξουμε πως η παραπάνω συνάρτηση πληροί τις δυο βασικές υποθέσεις που αναφέραμε. Έτσι, βλέπουμε ότι:

• Θεωρούμε τη μη αρνητική και μετρήσιμη συνάρτηση $\phi(\lambda) = q\lambda^{q-1}$, $\lambda > 0$, και για το τυχόν $b \in [0, +\infty]$ λαμβάνουμε πως:

$$\int_0^b \phi(\lambda)d\lambda = \int_0^b q\lambda^{q-1}d\lambda = q\left.\frac{\lambda^q}{q}\right|_0^b = b^q \overset{\text{ορ.}}{=} \Phi(b).$$

Επιπλέον, για το τυχόν $\lambda > 0$ παίρνουμε πως ισχύει ότι:

$$\phi(2\lambda) = q(2\lambda)^{q-1} = q2^{q-1}\lambda^{q-1} \overset{\text{ορ.}}{=} 2^{q-1}\phi(\lambda) \equiv c\phi(\lambda).$$

• Ισχύει προφανώς πως $\Phi(1) = 1^q = 1 < +\infty$.

Επιπλέον, είναι εύκολο να δούμε πως για τα θετικά x που ορίστηκε η συνάρτηση Φ είναι κυρτή, με αποτέλεσμα από την πρόταση 3.3.3 να υπάρχει σταθερά C που εξαρτάται αποκλειστικά από τη συνάρτηση Φ (θα τη συμβολίσουμε με C_q) τέτοια ώστε:

$$\mathbb{E}\left[\Phi\left(\left|\sum_{i=1}^{N} z_i\right|\right)\right] \leq C_q\mathbb{E}\left[\Phi\left(\left(\sum_{i=1}^{N} z_i^2\right)^{1/2}\right)\right] \Longrightarrow \mathbb{E}\left[\left|\sum_{i=1}^{N} z_i\right|^q\right] \leq C_q\mathbb{E}\left[\left(\sum_{i=1}^{N} z_i^2\right)^{q/2}\right]. \quad (3.11)$$

Για τη ζητούμενη τώρα αναμενόμενη τιμή έχουμε ότι:

$$\begin{aligned}
\mathbb{E}\left[|\overline{z}_N|^q\right] &= \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N} z_i\right|^q\right] \\
&= N^{-q}\mathbb{E}\left[\left|\sum_{i=1}^{N} z_i\right|^q\right] \\
&\overset{(3.11)}{\leq} C_qN^{-q}\mathbb{E}\left[\left(\sum_{i=1}^{N} z_i^2\right)^{q/2}\right] \\
&= C_q\mathbb{E}\left[\left(N^{-2}\sum_{i=1}^{N} z_i^2\right)^{q/2}\right] \\
&= C_q\mathbb{E}\left[\left(N^{-1}\frac{1}{N}\sum_{i=1}^{N} z_i^2\right)^{q/2}\right] \\
&= C_qN^{-q/2}\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} z_i^2\right)^{q/2}\right]. \quad (3.12)
\end{aligned}$$

Βλέπουμε όμως πως κάτω από το πλαίσιο στο οποίο εργαζόμαστε ισχύει ότι $2 \leq q$, συνεπώς αξιοποιώντας τη σχέση (3.12) και την πρόταση 3.3.4 προκύπτει ότι:

$$\begin{aligned}
\mathbb{E}\left[|\overline{z}_N|^q\right] &\leq C_q N^{-q/2}\mathbb{E}\left[\left(\left(\frac{1}{N}\sum_{i=1}^{N} z_i^2\right)^{1/2}\right)^q\right] \\
&= C_q N^{-q/2}\mathbb{E}\left[\left(\left(\frac{1}{N}\sum_{i=1}^{N} |z_i|^2\right)^{1/2}\right)^q\right] \\
&\leq C_q N^{-q/2}\mathbb{E}\left[\left(\left(\frac{1}{N}\sum_{i=1}^{N} |z_i|^q\right)^{1/q}\right)^q\right] \\
&= C_q N^{-q/2} N^{-1}\mathbb{E}\left[\sum_{i=1}^{N} |z_i|^q\right] \\
&= C_q N^{-q/2}\mathbb{E}\left[|Z|^q\right]. \quad (3.13)
\end{aligned}$$

Για τη δεύτερη ζητούμενη σχέση, κάνοντας χρήση της ανισότητας Markov για το τυχόν $z > 0$ παίρνουμε ότι:

$$\mathbb{P}\left[|\overline{z}_N| > z\right] = \mathbb{P}\left[|\overline{z}_N|^q > z^q\right] \leq \frac{\mathbb{E}[|\overline{z}_N|^q]}{z^q} \overset{(3.13)}{\leq} C_q z^{-q} N^{-q/2}\mathbb{E}\left[|Z|^q\right].$$

Προφανώς, η παραπάνω πιθανότητα φράσσεται και από την τιμή 1 με αποτέλεσμα να ισχύει τελικά πως:

$$\mathbb{P}\left[|\overline{z}_N| > z\right] \leq \min\left(1, C_q z^{-q} N^{-q/2}\mathbb{E}\left[|Z|^q\right]\right),$$

όπως ακριβώς θέλαμε. □

Παρατήρηση 3.3.5

Το λήμμα 3.3.2 που παραθέσαμε και αποδείξαμε παραπάνω μπορεί να διατυπωθεί με ένα λίγο γενικότερο τρόπο. Ειδικότερα, σύμφωνα με τη δημοσίευση [6] παίρνουμε πως τα αποτελέσματα αυτού του λήμματος έχουν πλήρη ισχύ και στην περίπτωση που η εξεταζόμενη μεταβλητή έχει πεπερασμένη ροπή μεγαλύτερη ή ίση της μονάδας διατηρώντας ταυτόχρονα τις υπόλοιπες βασικές υποθέσεις. Παρ΄ όλα αυτά, όπως θα δούμε στη συνέχεια αυτής της εργασίας η εκδοχή του λήμματος που εξετάσαμε μας καλύπτει σχεδόν στον απόλυτο βαθμό.

Εφαρμογή Λήμματος

Δοθέντος της μεταβλητής Y, θεωρούμε την ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών $(z_i)_{i=\{1,\dots,N\}}$ οι οποίες ορίζονται μέσω της σχέσης $z_i = x_i(Y) - \mathbb{E}[X|Y]$, $i = 1, \dots, N$, όπου με $x_i(Y)$ συμβολίζουμε μια μεταβλητή που ακολουθεί την κατανομή της X δεσμεύοντας ως προς Y. Τότε, για το δειγματικό μέσο $\overline{z}_N$ της συγκεκριμένης ακολουθίας θα ισχύει ότι:

$$\overline{z}_N = \frac{1}{N}\sum_{i=1}^{N} z_i = \frac{1}{N}\sum_{i=1}^{N}\left(x_i(Y) - \mathbb{E}[X|Y]\right) = \frac{1}{N}\sum_{i=1}^{N} x_i(Y) - \mathbb{E}[X|Y] = \overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y].$$

Επιπλέον, δεχόμαστε πως $\exists$ τιμή $2 < q < \infty$ ώστε να ικανοποιείται η υπόθεση 3.3.1. Κάτω από το πλαίσιο αυτό και με βάση τον ορισμό της ποσότητας κ_q λαμβάνουμε ότι:

$\kappa_q = \sup_y \mathbb{E}\left[\sigma^{-q}\left|X(Y) - \mathbb{E}[X|Y]\right|^q |Y = y\right] \geq \sigma^{-q}\mathbb{E}\left[\left|X(Y) - \mathbb{E}[X|Y]\right|^q\right] \overset{\text{ορ.}}{=} \sigma^{-q}\mathbb{E}\left[|Z|^q\right]$, οπότε:

$\kappa_q \geq \sigma^{-q}\mathbb{E}\left[|Z|^q\right] \Longrightarrow \mathbb{E}\left[|Z|^q\right] \leq \sigma^q\kappa_q \overset{\text{υποθ.}}{<} +\infty.$

Βλέπουμε λοιπόν πως η μεταβλητή $Z = X(Y) - \mathbb{E}[X|Y]$ έχει πεπερασμένη q-οστή ροπή για κάποιο $q \geq 2$ και μέση τιμή ίση με το 0, οπότε αξιοποιώντας το λήμμα 3.3.2 για την τιμή $z = d = |\mathbb{E}[X|Y]| > 0$ παίρνουμε ότι:

$\mathbb{P}\left[|\overline{z}_N| > z\right] \leq \min\left(1, C_q z^{-q} N^{-q/2}\mathbb{E}\left[|Z|^q\right]\right)$ ή μέσω της κατάλληλης αντικατάστασης:

$\mathbb{P}\left[\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y\right] \leq \min\left(1, C_q d^{-q} N^{-q/2}\mathbb{E}\left[|Z|^q\right]\right).$

Τέλος, κάνοντας χρήση του άνω φράγματος που προσδιορίσαμε για την q-οστή ροπή της μεταβλητής Z καταλήγουμε πως:

$\mathbb{P}\left[\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y\right] \leq \min\left(1, C_q d^{-q} N^{-q/2}\sigma^q\kappa_q\right)$ ή ισοδύναμα:

$$\mathbb{P}\left[\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y\right] \leq \min\left(1, C_q N^{-q/2}(d/\sigma)^{-q}\kappa_q\right) \equiv \min\left(1, C_q N^{-q/2}\delta^{-q}\kappa_q\right). \quad (3.14)$$

Αξίζει στο σημείο αυτό να θυμηθούμε πως σε μια προγενέστερη ανάλυσή μας είχαμε καταλήξει στη σχέση:

$\mathbb{P}\left[\left|\overline{\mathbb{E}}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y\right] \leq \min\left(1, \delta^{-2}N^{-1}\right).$

Έτσι, βλέπουμε πως η σχέση (3.14) δεν είναι τίποτα παραπάνω από μια γενίκευση του προηγούμενου αποτελέσματός μας.

3.4 Προσαρμοσμένο Δείγμα στη Μέθοδο Multilevel Monte Carlo Αριθμητική Ανάλυση

Όπως είδαμε σε προγενέστερη μελέτη μας, για τον προσδιορισμό της ποσότητας $\eta = \mathbb{E}[H(\mathbb{E}[X|Y])]$ στηριχτήκαμε στον εκτιμητή Multilevel Monte Carlo ο οποίος δίνεται με βάση τον τύπο:

$$\widetilde{I} = \frac{1}{M_0}\sum_{j=1}^{M_0} H\left(\overline{\mathbb{E}}^0[X|Y = y_j]\right) + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{j=1}^{M_l}\left(H\left(\overline{\mathbb{E}}^l[X|Y = y_j]\right) - H\left(\overline{\mathbb{E}}^{l-1}[X|Y = y_j]\right)\right),$$

όπου ισχύει πως:

$$\overline{\mathbb{E}}^l[X|Y = y_j] = \overline{\mathbb{E}}^{l,y_j}[X] = \frac{1}{N_l}\sum_{i=1}^{N_l} x_i^{l,y_j}$$

για μια αύξουσα και ντετερμινιστική ακολουθία N_l, $l = 0, \ldots, L$. Επιπλέον, όπως εξετάσαμε αναλυτικά στην ενότητα 3.2 η επιλογή $N_l = N_0 2^l$ μας οδηγεί σε μια ολική αριθμητική πολυπλοκότητα της τάξης του $\epsilon^{-5/2}$, όπου ϵ είναι η ζητούμενη ακρίβεια της μεθόδου μας. Αυτό που θα επιχειρήσουμε να κάνουμε παρακάτω είναι να εφαρμόσουμε το αριθμητικό σχήμα Multilevel Monte Carlo για μια προσαρμοσμένη ακολουθία N_l και να επιτύχουμε ένα μειωμένο αριθμητικό κόστος της τάξης του $\epsilon^{-2}|\log\epsilon|^2$, σύμφωνα με το θεώρημα ανάλυσης πολυπλοκότητας. Βέβαια, αξίζει να αναφέρουμε πως η διαδικασία που περιγράψαμε είναι στην πραγματικότητα ιδιαίτερα πολύπλοκη. Για το λόγο αυτό, στη συγκεκριμένη ενότητα θα εξάγουμε τα αποτελέσματά μας κάτω από ιδανικές/μη ρεαλιστικές συνθήκες, μεταφέροντας τη γενική και σαφώς δυσκολότερη μελέτη στο κεφάλαιο που θα επακολουθήσει. Έχοντας πλέον διασαφηνίσει τα παραπάνω προχωράμε στο κυρίως κομμάτι της ενότητας αυτής.

Ορίζουμε καταρχήν, δοθέντος της μεταβλητής Y, το εσωτερικό πλήθος δείγματος μέσω της σχέσης $N_l = \lceil \mathcal{N}_l \rceil$, όπου δεχόμαστε πως ισχύει ότι:

$$\mathcal{N}_l = N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{d}{\sigma}\right)^{-r}\right)\right).$$

Αξίζει να αναφέρουμε πως οδηγηθήκαμε στη συγκεκριμένη μορφή με τη βοήθεια των εκφράσεων που είχαμε πάρει για το προσαρμοσμένο δείγμα στο προηγούμενο κεφάλαιο και τροποποιώντας στα ακόλουθα σημεία:

- Η ποσότητα $\mathcal{N}_l$ είναι άνω φραγμένη από την τιμή $N_0 4^l$. Πράγματι, εύκολα παίρνουμε πως:

$$\mathcal{N}_l \leq N_0 4^l \max\left(2^{-l}, 1\right) = N_0 4^l.1 = N_0 4^l, \; \forall \; l = 0, 1, \ldots, L.$$

- Η ποσότητα $\mathcal{N}_l$ είναι κάτω φραγμένη από την τιμή $N_0 2^l$. Πράγματι, βλέπουμε ότι:

$$\mathcal{N}_l \geq N_0 4^l . 2^{-l} = N_0 2^l, \; \forall \; l = 0, 1, \ldots, L.$$

- Η τιμή C είναι μια σταθερά εμπιστοσύνης, σαν αυτή που είδαμε στην έκφραση για το προσαρμοσμένο πλήθος δείγματος της ενότητας 2.5, η οποία θεωρούμε πως ξεπερνά την τιμή 1.

- Η σταθερά r κυμαίνεται από 1 έως 2 και αυξάνει το πλήθος του δείγματος καθώς αυτή φθίνει.

Παρά το γεγονός πως τη συγκεκριμένη στιγμή η επιλογή για το εσωτερικό δείγμα δε φαντάζει προφανής θα δούμε στη συνέχεια πως μέσα από το συγκεκριμένο ορισμό ικανοποιούνται ιδιότητες όπως $\mathbb{E}\left[N_l\right] = \mathcal{O}\left(2^l\right)$, παρουσιάζοντας έτσι μια συνέπεια με την προγενέστερη επιλογή $N_l = N_0 2^l$, ενώ ταυτόχρονα μειώνεται η τάξη της διασποράς της διαφοράς μεταξύ διαδοχικών μεταβλητών διακριτοποίησης.

Παρατήρηση 3.4.1

Ακριβώς όπως συνέβαινε και στις εκφράσεις για το προσαρμοσμένο πλήθος που μελετήσαμε σε προηγούμενη ανάλυσή μας, παρατηρούμε πως για κάθε τιμή y της μεταβλητής Y η ποσότητα d/σ δεν είναι a-priori γνωστή και πρέπει να εκτιμηθεί κατάλληλα. Έτσι, υποθέτουμε πως οι τιμές d, σ προσδιορίζονται αριθμητικά από τις ποσότητες $\overline{d}$ και $\overline{\sigma}$ αντίστοιχα, δίχως να εστιάζουμε προς το παρόν στον τρόπο υπολογισμού των συγκεκριμένων εκτιμήσεων. Προφανώς τότε, μέσω του κλάσματος $\overline{d}/\overline{\sigma}$ ορίζεται μια αριθμητική προσέγγιση της τιμής $\mathcal{N}_l$, για την οποία θα εισάγουμε το συμβολισμό $\overline{\mathcal{N}}_l$. Επομένως, δοθέντος της μεταβλητής Y, ορίζουμε τελικά το εσωτερικό πλήθος δείγματος N_l σύμφωνα με τον τύπο $N_l = \left\lceil \overline{\mathcal{N}}_l \right\rceil$, όπου ισχύει πως:

$$\overline{\mathcal{N}}_l = N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}\right)\right), \text{ για } l = 0, 1, \ldots, L.$$

Ιδιότητα 3.4.2

Για το προσαρμοσμένο πλήθος δείγματος που ορίσαμε παραπάνω ικανοποιείται η ανισοτική σχέση $\overline{\mathcal{N}}_l \leq N_l < 2\overline{\mathcal{N}}_l$.

Απόδειξη: Βλέπουμε πως εξ' ορισμού έχουμε ότι $N_l = \left\lceil \overline{\mathcal{N}}_l \right\rceil$, συνεπώς από βασικές ιδιότητες του ακέραιου μέρους ενός πραγματικού αριθμού λαμβάνουμε την πρώτη ανισότητα:

$$\overline{\mathcal{N}}_l \leq N_l. \quad (3.15)$$

Επιπλέον, από το κάτω φράγμα της ποσότητας $\mathcal{N}_l$ και άρα της $\overline{\mathcal{N}}_l$ παίρνουμε ότι:

$\overline{\mathcal{N}}_l \geq N_0 2^l \geq N_0 \geq 1.$

Έτσι, για μια ακόμα φορά από τη σχέση $N_l = \lceil \overline{\mathcal{N}}_l \rceil$ έχουμε πως:

$N_l < \overline{\mathcal{N}}_l + 1 \leq \overline{\mathcal{N}}_l + \overline{\mathcal{N}}_l = 2\overline{\mathcal{N}}_l$. (3.16)

Από τις σχέσεις (3.15), (3.16) λαμβάνουμε τη ζητούμενη διπλή ανισότητα:

$\overline{\mathcal{N}}_l \leq N_l < 2\overline{\mathcal{N}}_l$. □

Είμαστε πλέον έτοιμοι να προχωρήσουμε στη βασική απόδειξη της συγκεκριμένης ενότητας η οποία και θα οδηγήσει στο συμπέρασμα πως $\mathcal{C} = \mathcal{O}\left(\epsilon^{-2} \left|\log \epsilon\right|^2\right)$ κατά μέσο όρο. Όπως αναφέραμε ήδη στην εισαγωγή μας, θα δείξουμε το αποτέλεσμα αυτό κάνοντας μια παραδοχή και συγκεκριμένα θεωρώντας πως έχουμε πλήρη γνώση των εκτιμητών $\overline{d}$, $\overline{\sigma}$, δηλαδή θεωρώντας πως $\overline{d} = d$ και $\overline{\sigma} = \sigma$. Παρά το γεγονός πως το εξεταζόμενο σενάριο είναι μη ρεαλιστικό, η απόδειξη του σχετικού θεωρήματος είναι μη τετριμμένη και στηρίζεται πάνω σε αρκετά αποτελέσματα που έχουμε δει έως τώρα. Ειδικότερα λοιπόν έχουμε ότι:

Θεώρημα 3.4.3(Giles, Haji-Ali)

Έστω μια μονοδιάστατη μεταβλητή X, η οποία δεσμεύεται από την τιμή μιας πολυδιάστατης μεταβλητής Y. Θεωρούμε τότε πως $\exists\ q > 2$ για το οποίο ικανοποιείται η υπόθεση 3.3.1, δηλαδή ισχύει πως:

$\kappa_q = \sup\limits_y \mathbb{E}\left[\sigma^{-q} \left|X(Y) - \mathbb{E}[X\,|Y]\right|^q |Y = y\right] < +\infty.$

Για τον προσδιορισμό της ποσότητας $\eta = \mathbb{E}[H(\mathbb{E}[X\,|Y])]$ ορίζουμε τον εκτιμητή Multilevel Monte Carlo $\widetilde{I}$ με προσαρμοσμένο πλήθος δείγματος N_l, ακριβώς όπως είδαμε παραπάνω. Επιπλέον, υποθέτουμε πως:

$1 < r < 2 - 2/q\,,\ \ d = \overline{d} = |\mathbb{E}[X|Y]|$ και $\ \sigma^2 = \overline{\sigma}^2 = Var[X|Y].$

Τότε, αν δεχτούμε πως η συνάρτηση πυκνότητας πιθανότητας ρ της μεταβλητής $\delta = d/\sigma$ είναι τοπικά φραγμένη κοντά στο 0 παίρνουμε πως:

$\mathbb{E}\left[N_l\right] = \mathcal{O}\left(2^l\right)$ και $Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] = \mathcal{O}\left(2^{-l}\right).$

Απόδειξη: Βλέπουμε πως από τα δεδομένα μας ισχύει ότι $d = \overline{d}$ και $\sigma = \overline{\sigma}$, με αποτέλεσμα για δοσμένη τιμή της μεταβλητής Y να παίρνουμε εξ' ορισμού τη σχέση $\overline{\mathcal{N}}_l = \mathcal{N}_l$. Σε ένα αρχικό στάδιο, θα εστιάσουμε στον προσδιορισμό της τάξης για την αναμενόμενη τιμή της μεταβλητής N_l, η οποία με βάση αυτά που είδαμε έως τώρα θα δίνεται μέσω της σχέσης:

$N_l = \lceil \overline{\mathcal{N}}_l \rceil \overset{\text{υποθ.}}{\Longrightarrow} N_l = \lceil \mathcal{N}_l \rceil$, όπου $\mathcal{N}_l = N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{d}{\sigma}\right)^{-r}\right)\right).$

Έτσι, αξιοποιώντας τη σχέση $N_l < 2\overline{\mathcal{N}}_l \Longrightarrow N_l < 2\mathcal{N}_l$ που προσδιορίσαμε στην ιδιότητα 3.4.2 λαμβάνουμε ότι:

$\mathbb{E}\left[N_l\right] \leq 2\mathbb{E}\left[\mathcal{N}_l\right] = 2\mathbb{E}\left[N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{d}{\sigma}\right)^{-r}\right)\right)\right].$

Εφόσον όμως γνωρίζουμε πως η μεταβλητή δ έχει συνάρτηση πυκνότητας πιθανότητας ρ, για την παραπάνω συνάρτηση της μεταβλητής αυτής έχουμε ότι:

$\mathbb{E}[N_l] \leq 2\mathbb{E}[\mathcal{N}_l] = 2N_0 4^l \int_0^\infty \max\left(2^{-l}, \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)\right)\rho(\delta)d\delta.$

Επιπλέον, εύκολα παρατηρούμε πως για οποιουσδήποτε θετικούς αριθμούς α, β ισχύει η ανισοτητα $\max(\alpha,\beta) < \alpha+\beta$, γεγονός που μας οδηγεί στη σχέση:

$$\begin{aligned}
\mathbb{E}[N_l] &\leq 2N_0 4^l \int_0^\infty \max\left(2^{-l}, \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)\right)\rho(\delta)d\delta \\
&\leq 2N_0 4^l \int_0^\infty \left(2^{-l} + \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)\right)\rho(\delta)d\delta \\
&= 2N_0 4^l 2^{-l} \int_0^\infty \rho(\delta)d\delta + 2N_0 4^l \int_0^\infty \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)\rho(\delta)d\delta \\
&= 2N_0 2^l + 2N_0 4^l \int_0^\infty \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)\rho(\delta)d\delta \\
&\equiv 2N_0 2^l + 2N_0 4^l \int_0^\infty \alpha(\delta)\rho(\delta)d\delta. \quad (3.17)
\end{aligned}$$

Από την υπόθεσή μας τώρα, έχουμε πως η συνάρτηση ρ είναι τοπικά φραγμένη κοντά στο 0, δηλαδή $\exists\ \rho_0,\ \delta_0 > 0$ τέτοια ώστε:

$\rho(\delta) \leq \rho_0,\ \forall\ \delta \in [0, \delta_0].$

Έτσι, με τη βοήθεια του λήμματος 3.2.1 και λαμβάνοντας υπόψη μας το γεγονός πως η συνάρτηση α της σχέσης (3.17) είναι μη αρνητική και μη αύξουσα, παίρνουμε ότι:

$\mathbb{E}[N_l] \leq 2N_0 2^l + 2N_0 4^l \rho_0 \int_0^\infty \alpha(\delta)d\delta + 2N_0 4^l \alpha(\delta_0)$ ή μέσω κατάλληλης αντικατάστασης:

$\mathbb{E}[N_l] \leq 2N_0 2^l + 2N_0 4^l \rho_0 \int_0^\infty \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)d\delta + 2N_0 4^l \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta_0\right)^{-r}\right).$

Επιπλέον, από την υπόθεσή μας ισχύει ότι $r > 1$, με αποτέλεσμα από το λήμμα 3.2.2 για το ολοκλήρωμα που εμφανίζεται στην τελευταία σχέση να καταλήγουμε πως:

$\int_0^\infty \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)d\delta \equiv \int_0^\infty \min\left(1, b\delta^{-r}\right)d\delta = \frac{r}{r-1}b^{1/r} \overset{\text{ορ.}}{=} \frac{r}{r-1}CN_0^{-1/2}2^{-l}.$

Τελικά λοιπόν, η αναμενόμενη τιμή της μεταβλητής N_l φράσσεται ως εξής:

$$\begin{aligned}
\mathbb{E}[N_l] &\leq 2N_0 2^l + 2N_0 4^l \rho_0 \frac{r}{r-1} CN_0^{-1/2}2^{-l} + 2N_0 4^l \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta_0\right)^{-r}\right) \\
&= 2N_0 2^l + 2CN_0^{1/2}\rho_0 \frac{r}{r-1}2^l + 2N_0 4^l \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta_0\right)^{-r}\right) \\
&\leq 2N_0 2^l + 2CN_0^{1/2}\rho_0 \frac{r}{r-1}2^l + 2N_0 4^l C^r N_0^{-r/2} 2^{-lr}\delta_0^{-r} \\
&= 2N_0 2^l + 2CN_0^{1/2}\rho_0 \frac{r}{r-1}2^l + 2C^r N_0^{(2-r)/2}\delta_0^{-r} 2^{(2-r)l}.
\end{aligned}$$

Τέλος όμως, παρατηρούμε ότι $r > 1 \implies -r < -1 \implies 2 - r < 1$, γεγονός που μας οδηγεί στο επιθυμητό αποτέλεσμα $\mathbb{E}[N_l] = \mathcal{O}\left(2^l\right)$.

Προχωράμε τώρα στον προσδιορισμό της τάξης για τη ζητούμενη διασπορά. Καταρχήν αξίζει να αναφέρουμε ότι, εφόσον υποθέσαμε πως η ποσότητα $\delta = d/\sigma$ είναι πλήρως καθορισμένη για την κάθε τιμή y της μεταβλητής Y, η αντίστοιχη ποσότητα N_l θα συμπεριφέρεται ως μια σταθερή τιμή που μπορούμε να υπολογίσουμε αναλυτικά. Με βάση το σκεπτικό αυτό και αξιοποιώντας ιδέες που αναπτύχθηκαν στην απόδειξη του λήμματος 3.2.3, βλέπουμε πως ισχύει ότι:

$$\begin{aligned} Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] &\leq \mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2 |Y\right]\right] \\ &\leq \mathbb{E}\left[\mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y\right]\right]. \quad (3.18) \end{aligned}$$

Επειδή τώρα από την υπόθεσή μας η τιμή κ_q είναι φραγμένη για κάποιο $q > 2$, από την εφαρμογή του λήμματος 3.3.2 και ειδικότερα τη σχέση (3.14) συμπεραίνουμε πως έχει ισχύ η παρακάτω ανισότητα:

$$\mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y\right] \leq \min\left(1, C_q\kappa_q\left(\delta N_l^{1/2}\right)^{-q}\right). \quad (3.19)$$

Έτσι, συνδυάζοντας τις σχέσεις (3.18) και (3.19) εύκολα έχουμε ότι:

$$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \leq \mathbb{E}\left[\min\left(1, C_q\kappa_q\left(\delta N_l^{1/2}\right)^{-q}\right)\right].$$

Από εδώ και στο εξής, ορίζουμε την ποσότητα $\nu(\delta) := C^{-1}N_0^{1/2}2^l\delta$, όπου C και N_0 είναι οι σταθερές που εμφανίζονται στον τύπο του εσωτερικού δείγματος N_l. Τότε, η τελευταία σχέση που πήραμε μπορεί να γραφτεί με τον ακόλουθο τρόπο:

$$\begin{aligned} Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] &\leq \mathbb{E}\left[\min\left(1, C_q\kappa_q\left(\delta N_l^{1/2}\right)^{-q}\right)\right] \\ &= \mathbb{E}\left[\min\left(1, C_q\kappa_q C^q N_0^{-q/2}2^{-lq}\delta^{-q}C^{-q}N_0^{q/2}2^{lq}N_l^{-q/2}\right)\right] \\ &\overset{\text{ορ.}}{=} \mathbb{E}\left[\min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}N_0^{q/2}2^{lq}N_l^{-q/2}\right)\right] \\ &= \mathbb{E}\left[\min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}N_l^{-q/2}N_0^{q/2}4^{lq/2}\right)\right] \\ &= \mathbb{E}\left[\min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}\left(\frac{N_l}{N_0 4^l}\right)^{-q/2}\right)\right] \end{aligned}$$

Ισχύει όμως πως $N_l \geq \overline{N}_l \Longrightarrow N_l \geq \mathcal{N}_l$, απ' όπου και προκύπτει εξ' ορισμού ότι:

$$\begin{aligned} &N_l \geq N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}\right)\right) \Longrightarrow \\ &N_l \geq N_0 4^l \max\left(2^{-l}, \min\left(1, \nu(\delta)^{-r}\right)\right) \Longrightarrow \\ &\frac{N_l}{N_0 4^l} \geq \max\left(2^{-l}, \min\left(1, \nu(\delta)^{-r}\right)\right) \Longrightarrow \\ &\left(\frac{N_l}{N_0 4^l}\right)^{-q/2} \leq \min\left(2^{lq/2}, \max\left(1, \nu(\delta)^{rq/2}\right)\right). \end{aligned}$$

Επομένως, αξιοποιώντας το τελευταίο μας αποτέλεσμα βλέπουμε πως για τη ζητούμενη διασπορά ισχύει ότι:

$$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \leq \mathbb{E}\left[\min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}\min\left(2^{lq/2}, \max\left(1, \nu(\delta)^{rq/2}\right)\right)\right)\right].$$

Εφόσον τώρα η συνάρτηση ν ορίστηκε μέσω της μεταβλητής δ θεωρούμε τη νέα συνάρτηση β η οποία δίνεται με τη βοήθεια του τύπου:

$$\beta(\delta) := \min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}\min\left(2^{lq/2}, \max\left(1, \nu(\delta)^{rq/2}\right)\right)\right).$$

Σε ένα πρώτο στάδιο, βλέπουμε εύκολα πως η β είναι μια μη αρνητική συνάρτηση του δ. Επιπλέον, παρατηρούμε πως ισχύει ότι:

$$r < 2 \Longrightarrow \tfrac{r}{2} < 1 \Longrightarrow \tfrac{r}{2}q < q,$$

με αποτέλεσμα η β ως συνάρτηση της μεταβλητής δ να φθίνει ταχύτερα από ότι αυξάνει. Το διαισθητικό αυτό επιχείρημα σε συνδυασμό με τη διάκριση διάφορων περιπτώσεων μπορούν να οδηγήσουν στο συμπέρασμα πως η συνάρτηση β είναι μη αύξουσα. Έτσι, για μια ακόμα φορά από το λήμμα 3.2.1 μπορούμε να πάρουμε πως ισχύει ότι:

$$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \leq \mathbb{E}[\beta(\delta)] = \textstyle\int_0^\infty \beta(\delta)\rho(\delta)d\delta \leq \rho_0\int_0^\infty \beta(\delta)d\delta + \beta(\delta_0).$$

Αν τώρα αντικαταστήσουμε τη συνάρτηση β που εμπεριέχεται στην τελευταία μας σχέση προκύπτει ότι:

$$\begin{aligned}
Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] &\leq \rho_0\textstyle\int_0^\infty \beta(\delta)d\delta + \beta(\delta_0)\\
&= \rho_0\textstyle\int_0^\infty \min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}\min\left(2^{lq/2}, \max\left(1, \nu(\delta)^{rq/2}\right)\right)\right)d\delta\\
&\quad + \min\left(1, C_q\kappa_q\nu(\delta_0)^{-q}C^{-q}\min\left(2^{lq/2}, \max\left(1, \nu(\delta_0)^{rq/2}\right)\right)\right)\\
&\leq \rho_0\textstyle\int_0^\infty \min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}\max\left(1, \nu(\delta)^{rq/2}\right)\right)d\delta\\
&\quad + \min\left(1, C_q\kappa_q\nu(\delta_0)^{-q}C^{-q}2^{lq/2}\right)\\
&\leq \rho_0\textstyle\int_0^\infty \min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}\max\left(1, \nu(\delta)^{rq/2}\right)\right)d\delta\\
&\quad + C_q\kappa_q C^{-q}\nu(\delta_0)^{-q}2^{lq/2}\\
&\overset{\text{ορ.}}{=} \rho_0\textstyle\int_0^\infty \min\left(1, C_q\kappa_q\nu(\delta)^{-q}C^{-q}\max\left(1, \nu(\delta)^{rq/2}\right)\right)d\delta\\
&\quad + C_q\kappa_q N_0^{-q/2}\delta_0^{-q}2^{-lq/2}.
\end{aligned}$$

Βλέπουμε όμως εξ' ορισμού πως:

$$\nu = C^{-1}N_0^{1/2}2^l\delta \Longrightarrow \delta = CN_0^{-1/2}2^{-l}\nu \Longrightarrow d\delta = CN_0^{-1/2}2^{-l}d\nu,$$

συνεπώς προκύπτει τελικά ότι:

$$\begin{aligned}
Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] &\leq \rho_0 CN_0^{-1/2}2^{-l}\textstyle\int_0^\infty \min\left(1, C_q\kappa_q C^{-q}\nu^{-q}\max\left(1, \nu^{rq/2}\right)\right)d\nu\\
&\quad + C_q\kappa_q N_0^{-q/2}\delta_0^{-q}2^{-lq/2}\\
&\leq \rho_0 CN_0^{-1/2}2^{-l}\textstyle\int_0^\infty \min\left(1, C_q\kappa_q C^{-q}\nu^{-q}\left(1 + \nu^{rq/2}\right)\right)d\nu\\
&\quad + C_q\kappa_q N_0^{-q/2}\delta_0^{-q}2^{-lq/2}\\
&= \rho_0 CN_0^{-1/2}2^{-l}\textstyle\int_0^\infty \min\left(1, C_q\kappa_q C^{-q}\nu^{-q} + C_q\kappa_q C^{-q}\nu^{rq/2-q}\right)d\nu\\
&\quad + C_q\kappa_q N_0^{-q/2}\delta_0^{-q}2^{-lq/2}.
\end{aligned}$$

Γνωρίζουμε όμως πως για οποιουσδήποτε δύο θετικούς αριθμούς a, b ικανοποιείται η ανισοτική σχέση:

$\min(1, a+b) \leq \min(1, a) + \min(1, b)$

με αποτέλεσμα να παίρνουμε ότι:

$$\begin{aligned} Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \leq \rho_0 C N_0^{-1/2} 2^{-l} &\int_0^\infty \min\left(1, C_q \kappa_q C^{-q} \nu^{-q}\right) d\nu \\ + \rho_0 C N_0^{-1/2} 2^{-l} &\int_0^\infty \min\left(1, C_q \kappa_q C^{-q} \nu^{-(q-rq/2)}\right) d\nu \\ + C_q \kappa_q N_0^{-q/2} \delta_0^{-q} 2^{-lq/2}. \quad (3.20) \end{aligned}$$

Από τα δεδομένα μας όμως λαμβάνουμε ότι:

$r < 2 - \frac{2}{q} \Longrightarrow \frac{r}{2} < 1 - \frac{1}{q} \Longrightarrow \frac{rq}{2} < q - 1 \Longrightarrow q - \frac{rq}{2} > 1,$

συνεπώς από το λήμμα 3.2.2 τα ολοκληρώματα της σχέσης (3.20) υπολογίζονται ως εξής:

- $\int_0^\infty \min\left(1, C_q \kappa_q C^{-q} \nu^{-q}\right) d\nu = \frac{q}{q-1}\left(C_q \kappa_q C^{-q}\right)^{1/q} = \frac{q}{q-1} C^{-1} \left(C_q \kappa_q\right)^{1/q}$
- $\int_0^\infty \min\left(1, C_q \kappa_q C^{-q} \nu^{-(q-rq/2)}\right) d\nu = \frac{q-rq/2}{(q-rq/2)-1}\left(C_q \kappa_q C^{-q}\right)^{\frac{1}{q-rq/2}}.$

Έτσι λοιπόν, δείξαμε πως τα ολοκληρώματά μας είναι πεπερασμένα και ανεξάρτητα του επιπέδου μελέτης l. Τέλος, ισχύει και η συνθήκη:

$q > 2 \Longrightarrow \frac{q}{2} > 1 \Longrightarrow 2^{-lq/2} < 2^{-l},$

γεγονός που οδηγεί στο συμπέρασμα:

$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] = \mathcal{O}\left(2^{-l}\right),$

ολοκληρώνοντας την απόδειξή μας. □

Προτού προχωρήσουμε τη μελέτη μας, θα παραθέσουμε δίχως απόδειξη μια γενίκευση του θεωρήματος ανάλυσης πολυπλοκότητας, όπου υποθέτουμε πως το κόστος διακριτοποίησης δεν είναι ντετερμινιστικό αλλά συμπεριφέρεται ως μια τυχαία μεταβλητή. Αξίζει να αναφέρουμε πως το θεώρημα αυτό εμπεριέχεται στην παραπομπή [7]. Προφανώς, το συμπέρασμα στο οποίο οδηγούμαστε είναι ανάλογο με εκείνο της απλής περίπτωσης που εξετάσαμε στην ενότητα 1.3. Ειδικότερα έχουμε ότι:

Θεώρημα 3.4.4(Giles)

Έστω $G : \Omega \to \mathbb{R}$ μια μονοδιάστατη τυχαία μεταβλητή και $\overline{G}^l$ η αριθμητική της προσέγγιση στο $l-$οστό επίπεδο μελέτης. Αν υπάρχουν ανεξάρτητοι εκτιμητές I_l βασισμένοι σε M_l το πλήθος προσομοιώσεις, η κάθε μία με κόστος (διακριτοποίησης) C_l και διασπορά V_l, και θετικές σταθερές $a, \beta, \gamma, c_1, c_2, c_3$ τέτοιες ώστε $a \geq \frac{1}{2}\min(\beta, \gamma)$ και

i) $\left|\mathbb{E}\left[\overline{G}^l - G\right]\right| \leq c_1 2^{-al}$

$ii)\ \mathbb{E}\Big[I_l\Big] = \begin{cases} \mathbb{E}\Big[\overline{G}^0\Big], & l = 0 \\ \mathbb{E}\Big[\overline{G}^l - \overline{G}^{l-1}\Big], & l > 0 \end{cases}$

$iii)\ V_l \leq c_2 2^{-\beta l}$

$iv)\ \mathbb{E}\left[C_l\right] \leq c_3 2^{\gamma l},$

τότε υπάρχει θετική σταθερά c_4 ώστε $\forall\ \epsilon < 1$ να υπάρχουν τιμές L και M_l για τις οποίες ο multilevel εκτιμητής $I = \sum\limits_{l=0}^{L} I_l$ έχει μέσο τετραγωνικό σφάλμα με φράγμα:

$MSE = \mathbb{E}\left[(I - \mathbb{E}[G])^2\right] < \epsilon^2$

και αναμενόμενο υπολογιστικό κόστος με φράγμα:

$$\mathbb{E}\left[\mathcal{C}\right] \leq \begin{cases} c_4\epsilon^{-2}, & \beta > \gamma \\ c_4\epsilon^{-2}(\log\epsilon)^2, & \beta = \gamma \\ c_4\epsilon^{-2-(\gamma-\beta)/\alpha}, & \beta < \gamma \end{cases}.$$

Πόρισμα 3.4.5

Κάτω από τις υποθέσεις του θεωρήματος 3.4.3 ισχύει ότι:

$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\overline{\mathbb{E}}_{N_{l-1}}[X|Y]\right)\right] = \mathcal{O}\left(2^{-l}\right)$

και λαμβάνουμε μια υπολογιστική πολυπλοκότητα για τη μέθοδό μας με την ιδιότητα:

$\mathbb{E}\left[\mathcal{C}\right] = \mathcal{O}\left(\epsilon^{-2}\left(\log\epsilon\right)^2\right).$

Απόδειξη: Όπως έχουμε δει έως τώρα, κάτω από την υπόθεση πως η μεταβλητή $\delta = d/\sigma$ είναι πλήρως καθορισμένη για την κάθε τιμή y της μεταβλητής Y, παίρνουμε πως η ποσότητα N_l συμπεριφέρεται ως μια σταθερή τιμή. Επιπλέον, υπενθυμίζουμε πως σύμφωνα με το λήμμα 3.2.4 για δυο οποιεσδήποτε τυχαίες μεταβλητές V, W ικανοποιείται η ανισότητα:

$Var[V - W] \leq 2Var[V] + 2Var[W].$

Έτσι, προσθαφαιρώντας τον όρο $H\left(\mathbb{E}[X|Y]\right)$ στη ζητούμενη διασπορά και κάνοντας χρήση της παραπάνω σχέσης έχουμε ότι:

$$\begin{aligned} Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\overline{\mathbb{E}}_{N_{l-1}}[X|Y]\right)\right] \leq\ & 2Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \\ & + 2Var\left[H\left(\overline{\mathbb{E}}_{N_{l-1}}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \end{aligned}$$

Αξιοποιώντας τώρα το αποτέλεσμα του θεωρήματος 3.4.3 καταλήγουμε πως:

$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\overline{\mathbb{E}}_{N_{l-1}}[X|Y]\right)\right] = \mathcal{O}\left(2^{-l}\right) + \mathcal{O}\left(2^{-(l-1)}\right) = \mathcal{O}\left(2^{-l}\right).$

Συνεπώς, εξ' ορισμού θα $\exists$ σταθερά $\overline{c}_2$ ωστε να ισχύει ότι:

$$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right)-H\left(\overline{\mathbb{E}}_{N_{l-1}}[X|Y]\right)\right]\leq \overline{c}_2 2^{-l}.$$

Στη συνέχεια, εφόσον για κάθε επίπεδο l και για κάθε τιμή y η ποσότητα N_l είναι πλήρως καθορισμένη θα ικανοποιεί τη βασική ιδιότητα που είδαμε στη μέθοδο Multilevel Monte Carlo ή με άλλα λόγια, δοθέντος της μεταβλητής Y, θα ισχύει πως:

$$\left|\mathbb{E}\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right)-H\left(\mathbb{E}[X|Y]\right)\right]\right|=\mathcal{O}\left(N_l^{-1}\right)\leq c_1 N_l^{-1}\leq c_1\mathcal{N}_l^{-1}\leq c_1 N_0^{-1}2^{-l}=\overline{c}_1 2^{-l},$$

όπου $\overline{c}_1=c_1N_0^{-1}$. Τέλος, αντιλαμβανόμαστε πως το κόστος διακριτοποίησης στο κάθε επίπεδο είναι μια τυχαία μεταβλητή που εξαρτάται αποκλειστικά από το εσωτερικό δείγμα N_l. Έτσι λοιπόν θα ικανοποιείται η σχέση:

$$C_l=\mathcal{O}\left(N_l\right)\Longrightarrow C_l\leq c_3N_l\Longrightarrow \mathbb{E}\left[C_l\right]\leq c_3\mathbb{E}\left[N_l\right]=\mathcal{O}(2^l),$$

με αποτέλεσμα να υπάρχει σταθερά $\overline{c}_3>0$, ώστε:

$$\mathbb{E}\left[C_l\right]\leq \overline{c}_3 2^l.$$

Από τη γενίκευση τώρα του θεωρήματος ανάλυσης πολυπλοκότητας, για τις τιμές $a=\beta=\gamma=1$ $(a\geq\min(\beta,\gamma)/2)$ και τις σταθερές $\overline{c}_1,\ \overline{c}_2,\ \overline{c}_3>0$ συμπεραίνουμε πως ισχύει ότι:

$$\mathbb{E}\left[\mathcal{C}\right]=\mathcal{O}\left(\epsilon^{-2}\left(\log\epsilon\right)^2\right).\qquad\square$$

4 Μέθοδος Multilevel Monte Carlo με Προσαρμοσμένο Δείγμα

4.1 Εισαγωγή Αλγορίθμου για το Προσαρμοσμένο Δείγμα

Στο προηγούμενο κεφάλαιο επιχειρήσαμε να εξετάσουμε το πρόβλημα αριθμητικής εκτίμησης της τιμής $\eta = \mathbb{P}\left[\mathbb{E}[X\,|\,Y] \geq 0\right]$ μέσα από την εφαρμογή ενός αριθμητικού σχήματος Multilevel Monte Carlo. Ειδικότερα, όπως μελετήσαμε αναλυτικά, εργαζόμενοι στο πλαίσιο αυτό και επιλέγοντας το εσωτερικό δείγμα μέσω μιας κατάλληλης ντετερμινιστικής ακολουθίας, μπορούμε να πάρουμε πως η ολική πολυπλοκότητα της μεθόδου μας έχει τάξη ίση με $\epsilon^{-5/2}$, όπου ϵ είναι η ακρίβεια που επιδιώκουμε να επιτύχουμε. Στη συνέχεια, προσπαθήσαμε να μειώσουμε περαιτέρω το αριθμητικό μας κόστος εισάγοντας στην υπάρχουσα μέθοδο Multilevel Monte Carlo ένα προσαρμοσμένο πλήθος εσωτερικού δείγματος στο κάθε επίπεδο, το οποίο δοθέντος της μεταβλητής Y δινόταν με τη βοήθεια του τύπου:

$$N_l = \left\lceil N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{d}{\sigma}\right)^{-r}\right)\right)\right\rceil,$$

όπου $d = \mathbb{E}[X|Y]$ και $\sigma^2 = Var[X|Y]$. Όπως εξηγήσαμε, το βασικό μειονέκτημα που εμφανίζει ο παραπάνω τύπος, είναι πως οι εκάστοτε τιμές d και σ δεν είναι a-priori γνωστές και πρέπει να αντικατασταθούν από κατάλληλες αριθμητικές τους προσεγγίσεις, τις οποίες συμβολίσαμε με $\overline{d}$ και $\overline{\sigma}$ αντίστοιχα. Συνεπώς, παρατηρήσαμε πως η σχέση προσδιορισμού του εσωτερικού προσαρμοσμένου δείγματος τροποποιείται και παίρνει τη μορφή:

$$N_l = \left\lceil N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}\right)\right)\right\rceil \equiv \lceil \overline{\mathcal{N}}_l \rceil. \quad (4.1)$$

Το ερώτημα λοιπόν που γεννιέται είναι με ποιόν τρόπο θα μπορέσουμε να οδηγηθούμε στην εύρεση των παραπάνω αριθμητικών τιμών $\overline{d}$, $\overline{\sigma}$. Υπενθυμίζουμε πως στο κεφάλαιο που προηγήθηκε, αντιμετωπίσαμε το πρόβλημα αυτό θεωρώντας ένα μη ρεαλιστικό σενάριο, στο οποίο έχουμε πλήρη γνώση των ζητούμενων ποσοτήτων και οδηγηθήκαμε σε μια έκφραση για την τάξη της ολικής πολυπλοκότητας, μέσω του θεωρήματος ανάλυσης πολυπλοκότητας του Giles. Αυτό που θα επιχειρήσουμε τώρα είναι να περάσουμε στη γενική περίπτωση, όπου υποθέτουμε πως η ποσότητα $\delta = d/\sigma$ δεν είναι γνωστή για την κάθε τιμή y της μεταβλητής Y και πρέπει να εκτιμηθεί αριθμητικά. Για τη συγκεκριμένη ανάλυση θα στηριχτούμε σε έναν αλγόριθμο ο οποίος επιτελεί ένα διπλό ρόλο. Ειδικότερα, όπως θα δούμε, ο αλγόριθμος αυτός προσεγγίζει κατάλληλα τις τιμές $\overline{d}$, $\overline{\sigma}$ και εν συνεχεία τις χρησιμοποιεί για τον προσδιορισμό του εσωτερικού δείγματος N_l. Επιπλέον, αξίζει να αναφέρουμε πως το αποτέλεσμα του αλγορίθμου που θα παραθέσουμε, δεν είναι ισοδύναμο με εκείνο που λαμβάνουμε μέσω της σχέσης (4.1), αλλά πληροί τη βασική ιδιότητα:

$$\overline{\mathcal{N}}_l \leq N_l < 2\overline{\mathcal{N}}_l, \quad (4.2)$$

που απορρέι από αυτή. Προφανώς, ο βασικός λόγος που επιλέγουμε να αποφύγουμε τη χρήση του τύπου (4.1) στον αλγοριθμό μας, είναι η πολυπλοκότητα που αυτός περικλύει, ενώ αν εξετάσουμε προσεκτικά την απόδειξη του θεωρήματος 3.4.3, αντιλαμβανόμαστε πως η συνθήκη (4.2) πιθανότατα είναι ικανή να μας οδηγήσει από μόνη της στη ζητούμενη τάξη πολυπλοκότητας.

Με βάση τα παραπάνω, είμαστε έτοιμοι να εισάγουμε τον αλγόριθμο με τον οποίο θα εργαστούμε στο συγκεριμένο κεφάλαιο με σκοπό την εύρεση του εκάστοτε εσωτερικού δείγματος N_l, αλλά και της ολικής πολυπλοκότητας που εξάγεται από αυτό. Συγκεριμένα, για δοσμένο επίπεδο μελέτης l, καθορισμένη τιμή της μεταβλητής Y ίση με y, αρχικό πλήθος δείγματος N_0 και εκθετική σταθερά r ορίζουμε το πλήθος του εσωτερικού δείγματος N_l σαν τη μεταβλητή η οποία προκύπτει ως αποτέλεσμα εκτέλεσης του παρακάτω αλγορίθμου:

Αλγόριθμος 4.1.1: Αλγόριθμος εύρεσης προσαρμοσμένου πλήθους δείγματος

Input: l, y, N_0, r
Output: N_l

/* Αρχικοποίηση μεταβλητών

$N_l = N_0 2^l$
completed = false

/* Επαναληπτική διαδικασία προσδιορισμού δείγματος

while completed = false
    if $N_l \geq N_0 4^l$
        completed = true
    else
        Δημιουργία ενός ανεξάρτητου και ισόνομου δείγματος της μεταβλητής X δοθέντος $Y = y$ και εύρεση νέων εκτιμήσεων για τις τιμές $\overline{d}$ και $\overline{\sigma}$
        if $N_l \geq N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}} \right)^{-r}$
            completed = true
        else
            $N_l = 2N_l$
        end
    end
end

return N_l

Θα προχωρήσουμε τώρα σε μια σειρά παρατηρήσεων με σκοπό να διασαφηνίσουμε τα κύρια σημεία του παραπάνω αλγορίθμου. Έτσι, έχουμε ότι:

Παρατήρηση 4.1.2

Για να είμαστε συνεπείς με προγενέστερη ανάλυσή μας, βλέπουμε πως το πλήθος του εσωτερικού δείγματος που εξάγεται από τον αλγόριθμο 4.1.1 κυμαίνεται μεταξύ των τιμών $N_0 2^l$ και $N_0 4^l$, όπως συνέβαινε για τη μεταβλητή $\mathcal{N}_l$ που εξετάσαμε στο προηγούμενο κεφάλαιο. Συγκεκριμένα, το κάτω φράγμα επιτυγχάνεται μέσω της αρχικοποίησης της μεταβλητής N_l, ενώ το άνω φράγμα διασφαλίζεται με βάση τον πρώτο υπό συνθήκη έλεγχο (if statement) που εφαρμόζουμε.

Παρατήρηση 4.1.3

Ο αλγόριθμος τερματίζει ακριβώς τη στιγμή που ικανοποιείται η συνθήκη $N_l \geq \overline{\mathcal{N}}_l$ για πρώτη φορά, όπου με $\overline{\mathcal{N}}_l$ συμβολίζουμε την τιμή που εμφανίζεται στη σχέση (4.1) χρησιμοποιώντας τις τελικές προσεγγίσεις των $\overline{d}$ και $\overline{\sigma}$. Βλέπουμε καταρχάς ότι, στην περίπτωση που επαληθεύεται ο πρώτος υπό συνθήκη έλεγχος που εφαρμόζουμε, ισχύει αναγκαστικά πως $N_l = N_0 4^l$, απ΄ όπου και μπορούμε εύκολα να πάρουμε ότι:

$$N_l = N_0 4^l \geq \overline{\mathcal{N}}_l.$$

Για να ολοκληρωθεί η αποδείξη του ισχυρισμού μας αρκεί να συμπεράνουμε πως η συνθήκη:

$$N_l \geq N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r} \quad (4.3)$$

οδηγεί στη σχέση $N_l \geq \overline{\mathcal{N}}_l$ καθώς και πως ισχύει κατ΄ ανάγκη ότι $N_l < \overline{\mathcal{N}}_l$ στην περίπτωση που η ανισοτική σχέση (4.3) δεν ικανοποιείται. Για το πρώτο σκέλος λοιπόν θεωρούμε πως ισχύει ότι:

$$N_l \geq N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r},$$

ενώ με βάση την αρχική μας συνθήκη θα ισχύει σίγουρα και $N_l \geq N_0 2^l$. Συνδυάζοντας τις δυο τελευταίες σχέσεις παρατηρούμε πως προκύπτει ότι:

$$N_l \geq N_0 4^l \max\left(2^{-l}, \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}\right) \geq N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}\right)\right) \overset{\text{ορ.}}{=} \overline{\mathcal{N}}_l.$$

Έστω τώρα πως η τιμή N_l που εξετάζουμε πληροί η συνθήκη:

$$N_l < N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}.$$

Θα δείξουμε τότε πως υποχρεωτικά καταλήγουμε στη σχέση $N_l < \overline{\mathcal{N}}_l$. Στο πλαίσιο που μελετάμε, βλέπουμε πως για το εσωτερικό πλήθος δείγματος ισχύει ότι $N_l < N_0 4^l$ (διαφορετικά θα είχαμε σταματήσει τον αλγόριθμο) με αποτέλεσμα να παίρνουμε πως:

$$N_l < N_0 4^l \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}\right) \leq N_0 4^l \max\left(2^{-l}, \min\left(1, \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}\right)\right) = \overline{\mathcal{N}}_l.$$

Εύκολα τώρα βλέπουμε πως με βάση τα παραπάνω ο ισχυρισμός μας έχει αποδειχτεί.

Παρατήρηση 4.1.4

Για το αποτέλεσμα του αλγορίθμου θα ικανοποιείται η συνθήκη $N_l < 2\overline{\mathcal{N}}_l$, όπου, όπως και πριν, με $\overline{\mathcal{N}}_l$ έχουμε συμβολίσει την τελευταία αριθμητική εκτίμηση της ποσότητας που εμπεριέχεται στη σχέση (4.1). Για να αποδείξουμε τη συγκεκριμένη σχέση θα προχωρήσουμε στην παραδοχή πως καθώς το πλήθος των προσομοιώσεων N_l διπλασιάζεται η τιμή $\overline{d}/\overline{\sigma}$ εμφανίζει πολύ μικρή διαφοροποίηση. Προφανώς ένα τέτοιο σενάριο είναι απόλυτα ρεαλιστικό εάν θεωρήσουμε πως η μεταβλητή εισόδου N_0 λαμβάνει μια αρκετά υψηλή τιμή. Κάτω από τη συγκερκριμένη θεώρηση εύκολα παρατηρούμε πως ισχύει ότι:

$$\frac{\overline{d}_{old}}{\overline{\sigma}_{old}} \approx \frac{\overline{d}_{new}}{\overline{\sigma}_{new}} \Longrightarrow N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}_{old}}{\overline{\sigma}_{old}}\right)^{-r} \approx N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}_{new}}{\overline{\sigma}_{new}}\right)^{-r}. \quad (4.4)$$

Σε ένα πρώτο στάδιο, θα εξετάσουμε την ισχύ του ζητούμενου αποτελέσματος στην περίπτωση που ο αλγόριθμος ολοκληρώνεται μετά από ακριβώς 1 επανάληψη, δηλαδή στην περίπτωση που έχουμε μία μόνο προσέγγιση του κλάσματος $\overline{d}/\overline{\sigma}$. Κάτω από το πλαίσιο αυτό, εύκολα παρατηρούμε πως ισχύει κατ΄ ανάγκη ότι $N_l = N_0 2^l < 2N_0 2^l \leq 2\overline{\mathcal{N}}_l$ και η σχέση έχει αποδειχτεί. Συνεχίζουμε τώρα τη μελέτη μας για το εσωτερικό πλήθος προσομοιώσεων στην περίπτωση που η συνθήκη τερματισμού δεν εκπληρώνεται στην πρώτη επανάληψη. Τότε, με βάση αυτά που είδαμε έως τώρα, αντιλαμβανόμαστε πως στο προτελευταίο βήμα του αλγορίθμου ικανοποιείται η ανισοτική σχέση:

$$N_{old} < N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}_{old}}{\overline{\sigma}_{old}}\right)^{-r} \Longrightarrow 2N_{old} < 2N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}_{old}}{\overline{\sigma}_{old}}\right)^{-r}.$$

Παρατηρούμε όμως ότι εκ κατασκευής ισχύει πως $N_{new} = 2N_{old}$, οπότε με τη βοήθεια της τελευταίας μας σχέσης αλλά και της (4.4) προκύπτει ότι:

$$N_{new} < 2N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}_{new}}{\overline{\sigma}_{new}}\right)^{-r} \overset{\text{συμ.}}{\Longrightarrow} N_l < 2N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r}.$$

Τέλος, ακολουθώντας αντίστοιχα βήματα με αυτά που είδαμε κατά την ολοκλήρωση της αμέσως προηγούμενης παρατήρησης καταλήγουμε πως:

$$N_l < 2\overline{N}_l$$

και η ζητούμενη ανισότητα έχει αποδειχτεί.

Παρατήρηση 4.1.5

Ως άμεσο πόρισμα των παρατηρήσεων 4.1.3 και 4.1.4 βλέπουμε πως για το πλήθος του εσωτερικού δείγματος που εξάγεται από τον αλγόριθμό μας ικανοποιείται η σχέση $\overline{N}_l \leq N_l < 2\overline{N}_l$.

4.2 Μελέτη Ολικού Υπολογιστικού Κόστους - Γενική Θεώρηση

Το βασικό αντικείμενο με το οποίο θα ασχοληθούμε στην ενότητα αυτή είναι η εύρεση της τάξης για την ολική πολυπλοκότητας της μεθόδου Multilevel Monte Carlo με κατάλληλη προσαρμοσμένη επιλογή εσωτερικού δείγματος. Αξίζει να θυμηθούμε πως για το ζητούμενο μέγεθος έχουμε ήδη εργαστεί στο τέλος του προηγούμενου κεφαλαίου, κάτω από ιδανικές συνθήκες, απ΄ όπου και συμπεράναμε πως ικανοποιείται η σχέση $\mathbb{E}[\mathcal{C}] = \mathcal{O}\left(\epsilon^{-2} (\log \epsilon)^2\right)$. Αυτό που θα επιχειρήσουμε να κάνουμε τώρα είναι να δείξουμε πως το αποτέλεσμα αυτό μεταφέρεται φυσιολογικά και σε ένα πιο ευρύ πλαίσιο στο οποίο και δεχόμαστε πως δε γνωρίζουμε εκ των προτέρων τις αναμενόμενες τιμές και διασπορές της μεταβλητής X δεσμεύοντας ως προς Y. Ειδικότερα λοιπόν, θα εξετάσουμε τη μέση πολυπλοκότητα της μεθόδου Multilevel Monte Carlo θεωρώντας πως το εσωτερικό πλήθος δείγματος δίνεται με τη βοήθεια του αλγορίθμου 4.1.1, όπου όπως είδαμε οι εκάστοτε άγνωστες ποσότητες d, σ εκτιμώνται αριθμητικά σε κάθε επανάληψη. Για να είμαστε όμως σε θέση να διατυπώσουμε το βασικό θεώρημα της συγκεκριμένης ενότητας θα χρειαστούμε δύο λήμματα, τα οποία διατυπώνονται και αποδεικνύονται στο κομμάτι της εργασίας που ακολουθεί.

Λήμμα 4.2.1

Έστω X μια μονοδιάστατη τυχαία μεταβλητή, η οποία δεσμεύεται από τις τιμές μιας πολυδιάστατης τυχαίας μεταβλητής Y. Θεωρούμε επιπλέον πως υπάρχει θετικός ακέραιος $q > 2$ ώστε:

$$\kappa_q = \sup_y \mathbb{E}\left[\sigma^{-q} \left|X(Y) - \mathbb{E}[X|Y]\right|^q | Y = y\right] < +\infty,$$

όπου όπως έχουμε ήδη εξηγήσει η παραπάνω ποσότητα αναπαριστά την κανονικοποιημένη και κεντραρισμένη q-οστή ροπή της μεταβλητής $X(Y)$. Τέλος, υποθέτουμε πως για την αριθμητική εκτίμηση της διασποράς $\sigma^2 = Var[X|Y]$ στηριζόμαστε σε $N_l = N_0 2^{l'}$ πλήθος δείγματος, με $l' = l, \ldots, 2l$, και πως η προσδιοριζόμενη τιμή δίνεται μέσω της σχέσης:

$\overline{\sigma}^2_{l'} = \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] \right)^2,$

για ένα ανεξάρτητο και ισόνομο δείγμα της μεταβλητής $X(Y)$ και για:

$\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] = \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} x_j^Y.$

Τότε, για κάθε σταθερά $c_1 > 0$ υπάρχει σταθερά $c_2 > 0$ εξαρτώμενη αποκλειστικά από τις ποσότητες $N_0,\ \kappa_q,\ q$ και c_1 ώστε να ικανοποιείται η σχέση:

$\mathbb{P}\left[\left| \overline{\sigma}^2_{l'} - \sigma^2 \right| > c_1 \sigma^2 | Y \right] \leq c_2 2^{-ql'/4}.$

Απόδειξη: Με βάση τα δεδομένα μας βλέπουμε πως, δοθέντος της μεταβλητής Y, για την εκτιμήτρια της διασποράς που χρησιμοποιούμε ισχύει ότι:

$$\begin{aligned}
\overline{\sigma}^2_{l'} &= \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] \right)^2 \\
&= \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] + \mathbb{E}[X|Y] - \overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] \right)^2 \\
&= \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right)^2 - \frac{2}{N_0 2^{l'}} \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right) \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right) \\
&\quad + \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right)^2 \\
&= \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right)^2 - 2 \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right)^2 + \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right)^2 \\
&= \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right)^2 - \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right)^2.
\end{aligned}$$

Κάνοντας χρήση της παραπάνω ισότητας παίρνουμε ότι:

$$\begin{aligned}
&\mathbb{P}\left[\left| \overline{\sigma}^2_{l'} - \sigma^2 \right| > c_1 \sigma^2 | Y \right] = \mathbb{P}\left[\left| \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] \right)^2 - \sigma^2 \right| > c_1 \sigma^2 | Y \right] \\
&= \mathbb{P}\left[\left| \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right)^2 - \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right)^2 - \sigma^2 \right| > c_1 \sigma^2 | Y \right] \\
&= \mathbb{P}\left[\left\{ \left| \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right)^2 - \sigma^2 \right| > \frac{c_1}{2} \sigma^2 \right\} \cup \left\{ \left| \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right)^2 \right| > \frac{c_1}{2} \sigma^2 \right\} | Y \right] \\
&\leq \mathbb{P}\left[\left| \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right)^2 - \sigma^2 \right| > \frac{c_1}{2} \sigma^2 | Y \right] + \mathbb{P}\left[\left| \left(\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right)^2 \right| > \frac{c_1}{2} \sigma^2 | Y \right] \\
&= \mathbb{P}\left[\left| \frac{1}{N_0 2^{l'}} \sum_{j=1}^{N_0 2^{l'}} \left(x_j^Y - \mathbb{E}[X|Y] \right)^2 - \sigma^2 \right| > \frac{c_1}{2} \sigma^2 | Y \right] + \mathbb{P}\left[\left| \overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y] \right| > \sqrt{\frac{c_1}{2}} \sigma | Y \right],
\end{aligned}$$

όπου για την εξαγωγή της προτελευταίας σχέσης κάναμε χρήση της γνωστής ιδιότητας:

$\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$,

η οποία και ισχύει για οποιαδήποτε δύο ενδεχόμενα A, B. Για την ολοκλήρωση της απόδειξής μας θα αξιοποιήσουμε το λήμμα 3.3.2 καθώς και την εφαρμογή που προκύπτει από αυτό. Έτσι, για το δεύτερο όρο της τελευταίας σχέσης, λαμβάνοντας υπόψη μας το γεγονός πως η q-οστή ροπή της μεταβλητής $Z = X(Y) - E[X|Y]$ είναι πεπεραμένη καταλήγουμε πως ισχύει ότι:

$$\begin{aligned}
\mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y]\right| > \sqrt{\tfrac{c_1}{2}}\sigma|Y\right] &\leq \min\left(1, C_q\left(\sqrt{\tfrac{c_1}{2}}\sigma\right)^{-q}\left(N_0 2^{l'}\right)^{-q/2}\mathbb{E}\left[|Z|^q\right]\right) \\
&\leq C_q\left(\sqrt{\tfrac{c_1}{2}}\sigma\right)^{-q}\left(N_0 2^{l'}\right)^{-q/2}\mathbb{E}\left[|Z|^q\right] \\
&\leq C_q 2^{q/2} c_1^{-q/2}\sigma^{-q}\left(N_0 2^{l'}\right)^{-q/2}\sigma^q\kappa_q \\
&= 2^{q/2} C_q c_1^{-q/2}\kappa_q\left(N_0 2^{l'}\right)^{-q/2}. \quad (4.5)
\end{aligned}$$

Αξίζει στο σημείο αυτό να αναφέρουμε πως για την ποσότητα $\mathbb{E}\left[|Z|^q\right] = \mathbb{E}\left[|X(Y) - \mathbb{E}[X|Y]|^q\right]$ χρησιμοποιήσαμε το άνω της φράγμα $\sigma^q\kappa_q$, το οποίο και προσδιορίστηκε στην εφαρμογή που εξετάσαμε στο τέλος της ενότητας 3.3. Για την πρώτη τώρα πιθανότητα θα στηριχτούμε σε μια ανάλογη διαδικασία, η οποία ωστόσο εμφανίζει μια επιπρόσθετη πολυπλοκότητα. Ειδικότερα, σε ένα πρώτο στάδιο, δοθέντος της Y, ορίζουμε τη νέα μεταβλητή $Z' = (X(Y) - \mathbb{E}[X|Y])^2 - \sigma^2$, για την οποία ισχύει πως:

$$\overline{z}'_{N_0 2^{l'}} \equiv \overline{z}' = \frac{1}{N_0 2^{l'}}\sum_{j=1}^{N_0 2^{l'}}\left(x_j^Y - \mathbb{E}[X|Y]\right)^2 - \sigma^2,$$

ενώ λαμβάνουμε και εξ' ορισμού της διασποράς σ^2 ότι $\mathbb{E}[Z'] = 0$. Επιπλέον, κάτω από την υπόθεση πως η μεταβλητή $Z = X(Y) - \mathbb{E}[X|Y]$ εχει πεπερασμένη q-οστή ροπή μπορούμε να πάρουμε ότι η μεταβλητή Z' πληροί την ιδιότητα:

$$\mathbb{E}\left[|Z'|^{q/2}\right] < \infty.$$

Για να καταλήξουμε στο αποτέλεσμα αυτό θα δείξουμε καταρχάς πως:

$$|Z'|^{q/2} \overset{\text{ορ.}}{=} \left|(X(Y) - \mathbb{E}[X|Y])^2 - \sigma^2\right|^{q/2} < 2^{q/2-1}\left(|X(Y) - \mathbb{E}[X|Y]|^q + \sigma^q\right).$$

Πράγματι, βλέπουμε ότι $q > 2 \implies \frac{q}{2} > 1$, απ' όπου μπορούμε να πάρουμε πως η συνάρτηση $f(x) = x^{q/2}$, $x \in \mathbb{R}_+$ είναι κυρτή. Συνεπώς, αξιοποιώντας τον ορισμό της κυρτότητας έχουμε ότι:

$$\begin{aligned}
\left|\frac{1}{2}(X(Y) - \mathbb{E}[X|Y])^2 - \frac{1}{2}\sigma^2\right|^{q/2} &< \left|\frac{1}{2}(X(Y) - \mathbb{E}[X|Y])^2 + \frac{1}{2}\sigma^2\right|^{q/2} \\
&= \left(\frac{1}{2}(X(Y) - \mathbb{E}[X|Y])^2 + \frac{1}{2}\sigma^2\right)^{q/2} \\
&\overset{\text{ορ.}}{=} f\left(\frac{1}{2}(X(Y) - \mathbb{E}[X|Y])^2 + \frac{1}{2}\sigma^2\right) \\
&\leq \frac{1}{2}f\left((X(Y) - \mathbb{E}[X|Y])^2\right) + \frac{1}{2}f\left(\sigma^2\right) \\
&= \frac{1}{2}|X(Y) - \mathbb{E}[X|Y]|^q + \frac{1}{2}\sigma^q. \quad (4.6)
\end{aligned}$$

Επιπλέον, παρατηρούμε πως:

$$\begin{aligned}\left|(X(Y)-\mathbb{E}[X|Y])^2-\sigma^2\right|^{q/2} &= \left|2\frac{1}{2}\,(X(Y)-\mathbb{E}[X|Y])^2-2\frac{1}{2}\sigma^2\right|^{q/2}\\ &= 2^{q/2}\left|\frac{1}{2}\,(X(Y)-\mathbb{E}[X|Y])^2-\frac{1}{2}\sigma^2\right|^{q/2}\\ &\overset{(4.6)}{<} 2^{q/2-1}\left(|X(Y)-\mathbb{E}[X|Y]|^q+\sigma^q\right).\end{aligned}$$

Έτσι, με βάση το τελευταίο μας αποτέλεσμα καταλήγουμε πως:

$$\begin{aligned}\mathbb{E}\left[|Z'|^{q/2}\right] &= \mathbb{E}\left[\left|(X(Y)-\mathbb{E}[X|Y])^2-\sigma^2\right|^{q/2}\right]\\ &\le 2^{q/2-1}\left(\mathbb{E}\left[|X(Y)-\mathbb{E}[X|Y]|^q\right]+\sigma^q\right)\\ &\overset{\text{ορ.}}{=} 2^{q/2-1}\sigma^q\left(\sigma^{-q}\mathbb{E}\left[|Z|^q\right]+1\right)\\ &\le 2^{q/2-1}\sigma^q\left(\kappa_q+1\right). \quad (4.7)\end{aligned}$$

Είναι άμεσο τώρα να δούμε πως το άνω φράγμα που προσδιορίσαμε είναι πεπερασμένο, συνεπώς λαμβάνοντας υπόψη μας την παρατήρηση 3.3.5 και εφαρμόζοντας το λήμμα 3.3.2 για τη μεταβλητή Z' παίρνουμε ότι:

$$\begin{aligned}\mathbb{P}\left[\left|\frac{1}{N_0 2^{l'}}\sum_{j=1}^{N_0 2^{l'}}\left(x_j^Y-\mathbb{E}[X|Y]\right)^2-\sigma^2\right|>\tfrac{c_1}{2}\sigma^2|Y\right] &= \mathbb{P}\left[\left|\overline{z}'_{N_0 2^{l'}}\right|>\tfrac{c_1}{2}\sigma^2|Y\right]\\ &\le \min\left(1, C_{q/2}\left(\tfrac{c_1}{2}\sigma^2\right)^{-q/2}\left(N_0 2^{l'}\right)^{-q/4}\mathbb{E}\left[|Z'|^{q/2}\right]\right)\\ &\le 2^{q/2}C_{q/2}c_1^{-q/2}\sigma^{-q}\left(N_0 2^{l'}\right)^{-q/4}\mathbb{E}\left[|Z'|^{q/2}\right]\\ &\overset{(4.7)}{\le} 2^{q/2}C_{q/2}c_1^{-q/2}\sigma^{-q}\left(N_0 2^{l'}\right)^{-q/4}2^{q/2-1}\sigma^q\left(\kappa_q+1\right)\\ &= 2^{q-1}C_{q/2}c_1^{-q/2}\left(N_0 2^{l'}\right)^{-q/4}\left(\kappa_q+1\right). \quad (4.8)\end{aligned}$$

Τελικά λοιπόν από τις σχέσεις (4.5) και (4.8) συμπεραίνουμε πως υπάρχει σταθερά $c_2>0$ εξαρτώμενη αποκλειστικά από τις τιμές $N_0,\ \kappa_q,\ c_1$ και q τέτοια ώστε:

$$\mathbb{P}\left[\left|\overline{\sigma}^2_{l'}-\sigma^2\right|>c_1\sigma^2|Y\right]\le c_2 2^{-ql'/4}.$$ □

Λήμμα 4.2.2

Έστω πως υπάρχει $q>2$ τέτοιο ώστε:

$$\kappa_q=\sup_y\mathbb{E}\left[\sigma^{-q}\,|X(Y)-\mathbb{E}[X|Y]|^q\,|Y=y\right]<+\infty,$$

ακριβώς όπως είδαμε και στο λήμμα που προηγήθηκε. Υποθέτουμε επιπλέον πως, δοθέντος της μεταβλητής Y και του επιπέδου l, η άγνωστη διασπορά σ^2 εκτιμάται αριθμητικά μέσω της σχέσης:

$\overline{\sigma}^2 = \frac{1}{N_l} \sum\limits_{j=1}^{N_l} \left(x_j^Y - \overline{\mathbb{E}}_{N_l}[X|Y] \right)^2,$

όπου N_l είναι το εσωτερικό πλήθος δείγματος που λαμβάνουμε από τον αλγόριθμό 4.1.1. Τότε θα $\exists$ σταθερά $c_3 > 0$ για την οποία:

$\mathbb{P}\left[\left| \overline{\sigma}^2 - \sigma^2 \right| > c_1 \sigma^2 | Y \right] \leq c_3 2^{-ql/4}.$

Απόδειξη: Εφόσον το εσωτερικό πλήθος δείγματος είναι αποτέλεσμα του αλγορίθμου που έχουμε εισαγάγει θα φέρει υποχρεωτικά τη μορφή $N_l = N_0 2^{l'}$ για κάποια τιμή $l' = l,\ l+1, \ldots,\ 2l-1,\ 2l$. Έτσι λοιπόν, με βάση τον εκτιμητή διασποράς που έχουμε στα δεδομένα μας αλλά και εκείνον που παραθέσαμε στο λήμμα 4.2.1 παίρνουμε ότι:

$$\begin{aligned}
\mathbb{P}\left[\left| \overline{\sigma}^2 - \sigma^2 \right| > c_1 \sigma^2 | Y \right] &\leq \mathbb{P}\left[\bigcup_{l'=l}^{2l} \left| \overline{\sigma}_{l'}^2 - \sigma^2 \right| > c_1 \sigma^2 | Y \right] \\
&\leq \sum_{l'=l}^{2l} \mathbb{P}\left[\left| \overline{\sigma}_{l'}^2 - \sigma^2 \right| > c_1 \sigma^2 | Y \right] \\
&\overset{4.2.1}{\leq} \sum_{l'=l}^{2l} c_2 2^{-ql'/4} \\
&\leq c_2 \sum_{l'=l}^{2l} \left(2^{-q/4} \right)^{l'} \\
&\leq c_2 \frac{\left(2^{-q/4} \right)^{l}}{1-2^{-q/4}} \\
&= \frac{c_2}{1-2^{-q/4}} 2^{-ql/4}.
\end{aligned}$$

Αξίζει να αναφέρουμε πως για να εξάγουμε την προτελευταία σχέση στηριχτήκαμε στο γνωστό αποτέλεσμα που αφορά πεπερασμένο άθροισμα όρων γεωμετρικής προόδου με λόγο μικρότερο της μονάδας. Αν τώρα θέσουμε $c_3 = \frac{c_2}{1-2^{-q/4}}$ με βάση την τελική μας σχέση καταλήγουμε πως:

$\mathbb{P}\left[\left| \overline{\sigma}^2 - \sigma^2 \right| > c_1 \sigma^2 | Y \right] \leq c_3 2^{-ql/4}.$ □

Είμαστε πλέον έτοιμοι να προχωρήσουμε στο θεώρημα που θα μας οδηγήσει στην εύρεση του ζητούμενου υπολογιστικού κόστους. Ειδικότερα, έχουμε ότι:

Θεώρημα 4.2.3(Giles, Haji-Ali)

Έστω X μια μονοδιάστατη τυχαία μεταβλητή, η οποία δεσμεύεται από τις τιμές μιας πολυδιάστατης τυχαίας μεταβλητής Y. Θεωρούμε επίσης, δοθέντος της μεταβλητής Y, τις ποσότητες $d = |\mathbb{E}[X|Y]|$, $\sigma^2 = Var[X|Y]$ και υποθέτουμε πως η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής $\delta := d/\sigma$ είναι τοπικά φραγμένη κοντά στο 0. Δεχόμαστε επιπλέον πως υπάρχει θετικός ακέραιος $q > 2$ ώστε:

$\kappa_q = \sup\limits_{y} \mathbb{E}\left[\sigma^{-q} \left| X(Y) - \mathbb{E}[X|Y] \right|^q | Y = y \right] < +\infty.$

Επιπρόσθετα, θεωρούμε πως η τιμή $\mathbb{E}[X|Y = y_j]$ προσεγγίζεται αριθμητικά σε κάθε επανάληψη του αλγορίθμου 4.1.1 μέσω της εκτιμήτριας:

$$\overline{\mathbb{E}}_{N_l}[X|Y=y_j] = \overline{\mathbb{E}}_{N_l}^{y_j}[X] = \frac{1}{N_l}\sum_{i=1}^{N_l} x_i^{y_j},$$

όπου N_l είναι το πλήθος δείγματος τη δοσμένη στιγμή και $\left\{x_i^{y_j}\right\}_{i=1}^{N_l}$ είναι ένα ανεξάρτητο τυχαίο δείγμα της μεταβλητής X, δοθέντος $Y=y_j$. Τέλος, με βάση την παραπάνω αριθμητική προσέγγιση και την τιμή N_l που έχουμε στην κάθε επανάληψη, ορίζουμε την εκτιμήτρια της διασποράς $Var[X|Y=y_j]$ στο κάθε βήμα του αλγορίθμου μέσω της σχέσης:

$$\overline{\sigma}^2 = \frac{1}{N_l}\sum_{i=1}^{N_l}\left(x_i^{y_j} - \overline{\mathbb{E}}_{N_l}[X|Y=y_j]\right)^2.$$

Τότε κάτω από την παραδοχή πως:

$$1 < r < 2 - \frac{\sqrt{4q+1}-1}{q}$$

λαμβάνουμε ότι:

$$\mathbb{E}[N_l] = \mathcal{O}\left(2^l\right) \text{ και } Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] = \mathcal{O}\left(2^{-l}\right).$$

<u>**Απόδειξη**</u>: Λόγω της πολυπλοκότητας που εμφανίζει η συγκεκριμένη απόδειξη, θα την παρουσιάσουμε μέσα από μια σειρά βημάτων με σκοπό να διευκολύνουμε τον αναγνώστη στην κατανόηση των βασικών ιδεών της. Έτσι λοιπόν έχουμε ότι:

Βήμα 1

Στο πρώτο βήμα της απόδειξής μας, θα αναζητήσουμε ένα άνω φράγμα για την πιθανότητα ο αλγόριθμός μας να τερματίσει σε ένα αρκετά υψηλότερο επίπεδο από εκείνο που θα παίρναμε έχοντας πλήρη γνώση των ποσοτήτων d και σ. Με βάση το σκεπτικό αυτό και για δοσμένο Y, θεωρούμε πως στην περίπτωση που οι τιμές d, σ ήταν καθορισμένες, ο αλγόριθμος θα ολοκληρωνόταν στο επίπεδο l^*, δίνοντας ένα πλήθος εσωτερικού δείγματος $N_l = N_0 2^{l^*}$, όπου $l \leq l^* \leq 2l$. Τότε σύμφωνα με τις συνθήκες τερματισμού του αλγορίθμου και τις παρατηρήσεις που διατυπώσαμε γύρω από αυτές, συμπεραίνουμε πως θα ίσχυαν οι ακόλουθες σχέσεις:

$$N_l \geq N_0 4^l \left(C^{-1}N_0^{1/2}2^l\tfrac{d}{\sigma}\right)^{-r} \overset{\text{ορ.}}{\Longrightarrow} N_0 2^{l^*} \geq N_0 4^l\left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r} \Longrightarrow \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r} \leq \frac{2^{l^*}}{4^l},$$

$$N_l < 2N_0 4^l \left(C^{-1}N_0^{1/2}2^l\tfrac{d}{\sigma}\right)^{-r} \Longrightarrow N_0 2^{l^*} < 2N_0 4^l\left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r} \Longrightarrow \frac{2^{l^*-1}}{4^l} < \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r}$$

Συνδυάζοντας τα δύο μας τελευταία αποτελέσματα λαμβάνουμε τελικά ότι:

$$\frac{2^{l^*-1}}{4^l} < \left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r} \leq \frac{2^{l^*}}{4^l}, \quad (4.9)$$

όπου η αριστερή ανισοτήτα ικανοποιείται για $l < l^* \leq 2l$ και η δεξιά για $l \leq l^* < 2l$. Υποθέτουμε τώρα πως ο αλγόριθμός μας, εκτιμώντας τις τιμές d και σ, τερματίζει σε ένα επίπεδο $\hat{l}$, με $l \leq \hat{l} \leq 2l$. Όπως ήδη αναφέραμε, στο σημείο αυτό θα εστιάσουμε στην περίπτωση που το παραπάνω επίπεδο τερματισμού είναι σημαντικά μεγαλύτερο από το l^* που εισαγάγαμε. Για το σκοπό αυτό, θα ορίσουμε ένα νέο επίπεδο l', το οποίο θα δεχτούμε πως βρίσκεται αρκετά μακριά από το l^*. Ειδικότερα, για τις ανάγκες της απόδειξής μας αρκεί να υποθέσουμε πως για το επίπεδο αυτό ικανοποιείται η ανισοτική σχέση $l' \geq l^* + 3$. Αυτό που θα κάνουμε λοιπόν είναι να επικεντρωθούμε στην πιθανότητα να ολοκληρωθεί ο αλγόριθμος που εξετάζουμε σε ένα τέτοιο επίπεδο ή με άλλα λόγια στην πιθανότητα να ισχύει η ισότητα $\hat{l} = l'$. Προφανώς, για το επίπεδο l' θα πρέπει να ικανοποιείται και η ανισότητα $l' \leq 2l \Longrightarrow l^* + 3 \leq 2l \Longrightarrow l^* \leq 2l - 3$, η οποία οδηγεί τελικά στον περιορισμό $l \leq l^* \leq 2l - 3$. Επειδή τώρα εξ' ορισμού ισχύει πως $l' \geq l^* + 3$ και επιβάλλουμε τη συνθήκη $\hat{l} = l'$ εύκολα αντιλαμβανόμαστε πως για όλα τα επίπεδα l'' με $l^*+2 \leq l'' < l'$ ο αλγόριθμος δεν έχει ακόμα σταματήσει, με αποτέλεσμα για το αντίστοιχο πλήθος δείγματος να παίρνουμε ότι:

$$N_l < N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r} \overset{\text{ορ.}}{\Longrightarrow} N_0 2^{l''} < N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}_{l''}\right)^{-r} \Longrightarrow \frac{2^{l''}}{4^l} < \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}_{l''}\right)^{-r}.$$

Ισχύει όμως εξ' υποθέσεως και η ανισότητα $l'' \geq l^* + 2$, συνπώς λαμβάνουμε ότι:

$$\frac{2^{l^*+2}}{4^l} \leq \frac{2^{l''}}{4^l} < \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}_{l''}\right)^{-r} \Longrightarrow 4\,\frac{2^{l^*}}{4^l} \leq \frac{2^{l''}}{4^l} < \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}_{l''}\right)^{-r}.$$

Με βάση την τελευταία σχέση αλλά και την (4.9) συμπεραίνουμε τελικά πως ισχύει ότι:

$$4\left(C^{-1} N_0^{1/2} 2^l \delta\right)^{-r} \leq 4\,\frac{2^{l^*}}{4^l} \leq \frac{2^{l''}}{4^l} < \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}_{l''}\right)^{-r},$$

με αποτέλεσμα:

$$4\left(C^{-1} N_0^{1/2} 2^l \delta\right)^{-r} < \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}_{l''}\right)^{-r} \Longrightarrow 4\delta^{-r} < \left(\overline{\delta}_{l''}\right)^{-r} \Longrightarrow \overline{\delta}_{l''} < 4^{-1/r}\delta. \quad (4.10)$$

Σύμφωνα με αυτά που είδαμε παραπάνω, κατανοούμε πως η συνθήκη $\hat{l} = l'$ μας οδηγεί στην ανισοτική σχέση:

$$\overline{\delta}_{l''} < 4^{-1/r}\delta,\ \forall\ l^*+2 \leq l'' < l',$$

επομένως βλέπουμε ότι:

$$\mathbb{P}\left[\hat{l} = l' | Y\right] \leq \mathbb{P}\left[\bigcap_{l''=l^*+2}^{l'-1} \left\{\overline{\delta}_{l''} < 4^{-1/r}\delta\right\} |Y\right] = \mathbb{P}\left[\bigcap_{l''=l^*+2}^{l'-1} \left\{\frac{d}{\overline{d}_{l''}} \frac{\overline{\sigma}_{l''}}{\sigma} > 4^{1/r}\right\} |Y\right].$$

Επειδή όμως το δείγμα που χρησιμοποιούμε για την εκτίμηση των ποσοτήτων d, σ είναι ανεξάρτητο μεταξύ των διαφόρων επαναλήψεων παίρνουμε ότι:

$$\mathbb{P}\left[\bigcap_{l''=l^*+2}^{l'-1} \left\{\frac{d}{\overline{d}_{l''}} \frac{\overline{\sigma}_{l''}}{\sigma} > 4^{1/r}\right\} |Y\right] = \prod_{l''=l^*+2}^{l'-1} \mathbb{P}\left[\frac{d}{\overline{d}_{l''}} \frac{\overline{\sigma}_{l''}}{\sigma} > 4^{1/r} |Y\right],$$

οπότε τελικά:

$$\mathbb{P}\left[\hat{l} = l' | Y\right] \leq \prod_{l''=l^*+2}^{l'-1} \mathbb{P}\left[\frac{d}{\overline{d}_{l''}} \frac{\overline{\sigma}_{l''}}{\sigma} > 4^{1/r} |Y\right].$$

Γνωρίζουμε τώρα πως στην περίπτωση που το γινόμενο δυο μη αρνητικών αριθμών είναι μεγαλύτερο από ένα $c > 0$ υποχρεωτικά τουλάχιστον ο ένας από τους δύο θα υπερβαίνει τη ρίζα του αριθμού αυτού. Έτσι, έχουμε ότι:

$$\begin{aligned}
\mathbb{P}\left[\hat{l} = l' | Y\right] &\leq \prod_{l''=l^*+2}^{l'-1} \mathbb{P}\left[\frac{d}{\overline{d}_{l''}} \frac{\overline{\sigma}_{l''}}{\sigma} > 4^{1/r} |Y\right] \\
&\leq \prod_{l''=l^*+2}^{l'-1} \mathbb{P}\left[\left\{\frac{d}{\overline{d}_{l''}} > 2^{1/r}\right\} \cup \left\{\frac{\overline{\sigma}_{l''}}{\sigma} > 2^{1/r}\right\} |Y\right] \\
&\leq \prod_{l''=l^*+2}^{l'-1} \left(\mathbb{P}\left[\frac{d}{\overline{d}_{l''}} > 2^{1/r} |Y\right] + \mathbb{P}\left[\overline{\sigma}^2_{l''} > 4^{1/r}\sigma^2 |Y\right]\right),
\end{aligned}$$

για κάθε επίπεδο l' με $l' \geq l^* + 3$. Στο σημείο αυτό θα εξετάσουμε την κάθε πιθανότητα χωριστά.

α) $$\begin{aligned}\mathbb{P}\left[\frac{d}{\overline{d}_{l''}} > 2^{1/r}|Y\right] &= \mathbb{P}\left[\overline{d}_{l''} < 2^{-1/r}d|Y\right] \\ &= \mathbb{P}\left[-\overline{d}_{l''} > -2^{-1/r}d|Y\right] \\ &= \mathbb{P}\left[d-\overline{d}_{l''} > \left(1-2^{-1/r}\right)d|Y\right] \\ &\leq \mathbb{P}\left[\left|d-\overline{d}_{l''}\right| > \left(1-2^{-1/r}\right)d|Y\right] \\ &= \mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_0 2^{l''}}[X|Y]-\mathbb{E}[X|Y]\right| > \left(1-2^{-1/r}\right)d|Y\right]\end{aligned}$$

Με τη βοήθεια όμως της σχέσης (4.9) λαμβάνουμε ότι:

$$\begin{aligned}&\left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r} \leq \frac{2^{l^*}}{4^l} &&\Longrightarrow \\ &\left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r} \leq 2^{l^*-2l} &&\Longrightarrow \\ &C^{-1}N_0^{1/2}2^l\delta \geq 2^{(2l-l^*)/r} &&\Longrightarrow \\ &C^{-1}N_0^{1/2}2^l\frac{d}{\sigma} \geq 2^{(2l-l^*)/r} &&\Longrightarrow \\ &d \geq C\sigma N_0^{-1/2}2^{-l}2^{(2l-l^*)/r}.\end{aligned}$$

Με βάση τώρα το τελευταίο μας αποτέλεσμα βλέπουμε πως:

$$\begin{aligned}\mathbb{P}\left[\frac{d}{\overline{d}_{l''}} > 2^{1/r}|Y\right] &\leq \mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_0 2^{l''}}[X|Y]-\mathbb{E}[X|Y]\right| > \left(1-2^{-1/r}\right)d|Y\right] \\ &\leq \mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_0 2^{l''}}[X|Y]-\mathbb{E}[X|Y]\right| > \left(1-2^{-1/r}\right)C\sigma N_0^{-1/2}2^{-l}2^{(2l-l^*)/r}|Y\right].\end{aligned}$$

Αξιοποιώντας όμως το λήμμα 3.3.2 για το δειγματικό μέσο $\overline{z}_{N_0 2^{l''}} = \overline{\mathbb{E}}_{N_0 2^{l''}}[X|Y] - \mathbb{E}[X|Y]$, παρατηρούμε με βάση την τελευταία μας σχέση πως ισχύει ότι:

$$\begin{aligned}\mathbb{P}\left[\frac{d}{\overline{d}_{l''}} > 2^{1/r}|Y\right] &\leq \min\left(1, C_q\left[\left(1-2^{-1/r}\right)C\sigma N_0^{-1/2}2^{-l}2^{(2l-l^*)/r}\right]^{-q}\left(N_0 2^{l''}\right)^{-q/2}\sigma^q\kappa_q\right) \\ &\leq C_q\left[\left(1-2^{-1/r}\right)C\sigma N_0^{-1/2}2^{-l}2^{(2l-l^*)/r}\right]^{-q}\left(N_0 2^{l''}\right)^{-q/2}\sigma^q\kappa_q \\ &= \left(1-2^{-1/r}\right)^{-q}C_q\kappa_q C^{-q}2^{-q(2l-l^*)/r}2^{lq}2^{-l''q/2} \\ &= \left(1-2^{-1/r}\right)^{-q}C_q\kappa_q C^{-q}2^{-q\left((2l-l^*)/r-\left(2l-l''\right)/2\right)}.\end{aligned}$$

Τέλος, από τα δεδομένα μας ισχύει πως:

$$r < 2 \Longrightarrow \frac{1}{r} > \frac{1}{2} \Longrightarrow -\frac{q}{r} < -\frac{q}{2},$$

απ' όπου γίνεται άμεσα αντιληπτό ότι:

$$-\frac{q}{r}\left(2l-l^*\right)+\left(2l-l''\right)\frac{q}{2} < -\frac{q}{2}\left(2l-l^*\right)+\left(2l-l''\right)\frac{q}{2} = -ql+\frac{q}{2}l^*+ql-\frac{q}{2}l'' = -q\left(l''-l^*\right)/2,$$

με αποτέλεσμα:

$\mathbb{P}\left[\frac{d}{d_{l''}} > 2^{1/r}|Y\right] = \mathcal{O}\left(2^{-q\left(l''-l^*\right)/2}\right).$

β) $\mathbb{P}\left[\overline{\sigma}^2_{l''} > 4^{1/r}\sigma^2|Y\right] = \mathbb{P}\left[\overline{\sigma}^2_{l''} - \sigma^2 > \left(4^{1/r}-1\right)\sigma^2|Y\right] \leq \mathbb{P}\left[\left|\overline{\sigma}^2_{l''} - \sigma^2\right| > \left(4^{1/r}-1\right)\sigma^2|Y\right]$

Κάνοντας τώρα χρήση του λήμματος 4.2.1 παίρνουμε την ανισοτική σχέση:

$\mathbb{P}\left[\overline{\sigma}^2_{l''} - \sigma^2 > \left(4^{1/r}-1\right)\sigma^2|Y\right] \leq \mathbb{P}\left[\left|\overline{\sigma}^2_{l''} - \sigma^2\right| > \left(4^{1/r}-1\right)\sigma^2|Y\right] \leq c_2 2^{-ql''/4},$

όπου $c_2 > 0$ είναι μια σταθερά εξαρτώμενη από τις τιμές N_0, κ_q, q και την ποσότητα $c_1 = 4^{1/r}-1$, η οποία είναι αυστηρά θετική. Παρατηρούμε όμως πως για το επίπεδο l^* ικανοποιείται και η προφανής σχέση $2^{ql^*/4} \geq 1$, οδηγώντας μας τελικά στο αποτέλεσμα:

$\mathbb{P}\left[\overline{\sigma}^2_{l''} - \sigma^2 > \left(4^{1/r}-1\right)\sigma^2|Y\right] = \mathcal{O}\left(2^{-ql''/4}\right)\mathcal{O}\left(2^{ql^*/4}\right) = \mathcal{O}\left(2^{-q\left(l''-l^*\right)/4}\right).$

Τελικά λοιπόν, μέσα από την μελέτη των δύο πιθανοτήτων, παίρνουμε ότι:

$$\begin{aligned}
\mathbb{P}\left[\hat{l} = l'|Y\right] &\leq \prod_{l''=l^*+2}^{l'-1}\left(\mathbb{P}\left[\frac{d}{d_{l''}} > 2^{1/r}|Y\right] + \mathbb{P}\left[\overline{\sigma}^2_{l''} > 4^{1/r}\sigma^2|Y\right]\right)\\
&= \prod_{l''=l^*+2}^{l'-1}\left(\mathcal{O}\left(2^{-q\left(l''-l^*\right)/2}\right) + \mathcal{O}\left(2^{-q\left(l''-l^*\right)/4}\right)\right)\\
&= \prod_{l''=l^*+2}^{l'-1}\mathcal{O}\left(2^{-q\left(l''-l^*\right)/4}\right)\\
&= \mathcal{O}\left(\prod_{l''=l^*+2}^{l'-1} 2^{-q\left(l''-l^*\right)/4}\right)\\
&= \mathcal{O}\left(2^{-q\left(l'-l^*\right)^2/8}\right).
\end{aligned}$$

Συνεπώς, με βάση τα παραπάνω δείξαμε ότι:

$\mathbb{P}\left[\hat{l} = l'|Y\right] = \mathcal{O}\left(2^{-q\left(l'-l^*\right)^2/8}\right),\ \forall\ l \leq l^* \leq 2l-3.$

Εύκολα όμως αντιλαμβανόμαστε πως το άνω φράγμα που προσδιορίσαμε ισχύει και στην περίπτωση που $l^*+3 \leq l' \leq 2l$ με $2l-3 < l^* \leq 2l$. Συγκεκριμένα, αν θεωρήσουμε πως $2l-3 < l^*$ παρατηρούμε πως ο αλγόριθμός μας δεν μπορεί να σταματήσει στο επίπεδο l' μιας και το τελευταίο ξεπερνά υποχρεωτικά την τιμή $2l$. Με άλλα λόγια, κάτω από το πλαίσιο που εξετάζουμε λαμβάνουμε ότι:

$\mathbb{P}\left[\hat{l} = l'|Y\right] = 0 = \mathcal{O}\left(2^{-q\left(l'-l^*\right)^2/8}\right)$

και τελικά το άνω φράγμα που βρήκαμε έχει ισχύ για όλα τα επίπεδα $l \leq l^* \leq 2l$.

Βήμα 2

Σε αυτό το στάδιο της απόδειξής μας θα οδηγηθούμε στην πρώτη ζητούμενη σχέση, $\mathbb{E}[N_l] = \mathcal{O}\left(2^l\right)$, με τη βοήθεια του άνω φράγματος που εξάγαμε στο αμέσως προηγούμενο βήμα. Ειδικότερα, αν ορίσουμε τη συνάρτηση $\alpha\left(N_l\right) := \frac{N_l}{N_0 4^l}$ βλέπουμε πως ισχύει ότι:

$$\begin{aligned}
\mathbb{E}\left[\tfrac{N_l}{N_0 4^l}|Y\right] &= \mathbb{E}\left[\alpha\left(N_l\right)|Y\right] \\
&= \sum_{l'=l}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] \alpha\left(N_0 2^{l'}\right) \\
&= \sum_{l'=l}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] \tfrac{N_0 2^{l'}}{N_0 4^l} \\
&= \sum_{l'=l}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] 2^{l'-2l}.
\end{aligned}$$

Προφανώς, για τον παραπάνω υπολογισμό στηριχτήκαμε στο γεγονός πως η τυχαία μεταβλητή N_l λαμβάνει τις διακριτές τιμές $N_0 2^{l'}$, $l \leq l' \leq 2l$, όπως φαίνεται καθαρά από την κατασκευή του αλγορίθμου μας. Προχωρώντας τη σκέψη μας ακόμα ένα βήμα παίρνουμε τελικά ότι:

$$\mathbb{E}\left[\tfrac{N_l}{N_0 4^l}|Y\right] = \sum_{l'=l}^{l^*+2} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] 2^{l'-2l} + \sum_{l'=l^*+3}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] 2^{l'-2l}.$$

Για το πρώτο άθροισμα χρησιμοποιούμε την προφανή ανισοτική σχέση:

$$\mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] \leq 1$$

και λαμβάνουμε ότι:

$$\sum_{l'=l}^{l^*+2} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] 2^{l'-2l} \leq \sum_{l'=l}^{l^*+2} 2^{l'-2l} = 2^{-2l} \sum_{l'=l}^{l^*+2} 2^{l'} = 2^{-2l} 2^l \tfrac{2^{l^*+3-l}-1}{2-1} \leq 2^{l^*+3-2l}.$$

Έτσι λοιπόν έχουμε πως:

$$\mathbb{E}\left[\tfrac{N_l}{N_0 4^l}|Y\right] \leq 2^{l^*+3-2l} + \sum_{l'=l^*+3}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] 2^{l'-2l}.$$

Στο σημείο αυτό, παρατηρούμε πως το ενδεχόμενο $N_l = N_0 2^{l'}$ εκπληρώνεται ακριβώς όταν ικανοποιείται η συνθήκη $\hat{l} = l'$ και αντίστροφα, με αποτέλεσμα να παίρνουμε ότι:

$$\mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] = \mathbb{P}\left[\hat{l} = l'|Y\right].$$

Έτσι, για τη μέση τιμή που εξετάζουμε συμπεραίνουμε ότι:

$$\mathbb{E}\left[\tfrac{N_l}{N_0 4^l}|Y\right] \leq 2^{l^*+3-2l} + \sum_{l'=l^*+3}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'}|Y\right] 2^{l'-2l} = 2^{l^*+3-2l} + \sum_{l'=l^*+3}^{2l} \mathbb{P}\left[\hat{l} = l'|Y\right] 2^{l'-2l}.$$

Με τη βοήθεια όμως του προηγούμενου βήματος, βλέπουμε πως για όλες τις τιμές l' του παραπάνω αθροίσματος ισχύει ότι:

$$\mathbb{P}\left[\hat{l} = l'|Y\right] = \mathcal{O}\left(2^{-q\left(l'-l^*\right)^2/8}\right),$$

γεγονός που μας οδηγεί στο αποτέλεσμα:

$$\mathbb{E}\left[\tfrac{N_l}{N_0 4^l}|Y\right] \leq 2^{l^*+3-2l} + \sum_{l'=l^*+3}^{2l} \mathcal{O}\left(2^{-q\left(l'-l^*\right)^2/8}\right) 2^{l'-2l}.$$

Θα επιχειρήσουμε τώρα να να απλοποιήσουμε την έκφραση που εμφανίζεται εντός του παραπάνω αθροίσματος. Ειδικότερα, έχουμε πως:

$$\begin{aligned}\sum_{l'=l^*+3}^{2l} \mathcal{O}\left(2^{-q\left(l'-l^*\right)^2/8}\right) 2^{l'-2l} &= \sum_{l'=l^*+3}^{2l} \mathcal{O}\left(2^{-q\left(l'-l^*\right)^2/8}\right) 2^{l'-2l+l^*-l^*} \\ &\leq \sum_{l'=l^*+3}^{2l} c2^{-q\left(l'-l^*\right)^2/8} 2^{l'-l^*} 2^{l^*-2l} \\ &= c2^{l^*-2l} \sum_{l'=l^*+3}^{2l} 2^{-q\left(l'-l^*\right)^2/8+l'-l^*} \\ &= \mathcal{O}\left(2^{l^*-2l}\right) \sum_{l'=l^*+3}^{2l} 2^{-q\left(l'-l^*\right)^2/8+l'-l^*}.\end{aligned}$$

Για να προχωρήσουμε το βήμα αυτό, ισχυριζόμαστε πως για κάθε θετικό ακέραιο $q > 2$ ισχύει ότι $l' - l^* - q\left(l' - l^*\right)^2/8 < 0$. Πράγματι, βλέπουμε πως:

$$q \geq 3 \Longrightarrow \frac{1}{q} \leq \frac{1}{3} \Longrightarrow \frac{8}{q} \leq \frac{8}{3} < 3 \leq l' - l^* \Longrightarrow \frac{8}{q} < l' - l^* \Longrightarrow l' - l^* - q\left(l' - l^*\right)^2/8 < 0.$$

Αν στο σημείο αυτό αξιοποιήσουμε το γεγονός πως, με βάση τα δεδομένα μας, ικανοποιείται η συνθήκη $q \geq 3$, καταλήγουμε στη σχέση:

$$\sum_{l'=l^*+3}^{2l} 2^{-q\left(l'-l^*\right)^2/8+l'-l^*} = \mathcal{O}(1).$$

Έτσι, λαμβάνουμε τελικά ότι:

$$\mathbb{E}\left[\frac{N_l}{N_0 4^l}|Y\right] \leq 2^{l^*+3-2l} + \mathcal{O}\left(2^{l^*-2l}\right)\mathcal{O}(1) = \mathcal{O}\left(2^{l^*-2l}\right) + \mathcal{O}\left(2^{l^*-2l}\right) = \mathcal{O}\left(2^{l^*-2l}\right).$$

Αυτό που θα επιχειρήσουμε τώρα είναι να γραφουμε το άνω φράγμα της μέσης τιμής που εξετάζουμε μέσα από έναν διαφορετικό τρόπο. Συγκεκριμένα, από τη σχέση (4.9) και από τον ορισμό της ποσότητας ν που δώσαμε σε προγενέστερη ανάλυσή μας έχουμε ότι:

$$\left(C^{-1}N_0^{1/2}2^l\delta\right)^{-r} > \frac{2^{l^*-1}}{4^l} \overset{\text{ορ.}}{\Longrightarrow} 2\nu^{-r} > 2^{l^*-2l} \Longrightarrow 2^{l^*-2l} = \mathcal{O}\left(\nu^{-r}\right).$$

Ισχύει όμως και η ανισοτική σχέση:

$$l^* - 2l \leq 0 \Longrightarrow 2^{l^*-2l} \leq 1 \Longrightarrow 2^{l^*-2l} = \mathcal{O}(1).$$

Συνδυάζοντας τα δύο τελευταία μας αποτελέσματα συμπεραίνουμε πως:

$$\mathcal{O}\left(2^{l^*-2l}\right) = \mathcal{O}\left(\min\left(1, \nu^{-r}\right)\right) = \mathcal{O}\left(\min\left(1, \max\left(2^{-l}, \nu^{-r}\right)\right)\right).$$

Αξίζει να αναφέρουμε πως για την εξαγωγή της τελευταίας ισότητας στηριχτήκαμε στην προφανή σχέση:

$$\nu^{-r} \leq \max\left(2^{-l}, \nu^{-r}\right).$$

Έτσι λοιπόν, δείξαμε τελικά ότι:

$\mathbb{E}\left[\frac{N_l}{N_0 4^l}|Y\right] = \mathcal{O}\left(\min\left(1, \max\left(2^{-l}, \nu^{-r}\right)\right)\right) \Longrightarrow \mathbb{E}\left[N_l|Y\right] = \mathcal{O}\left(\min\left(1, \max\left(2^{-l}, \nu^{-r}\right)\right)\right),$

με αποτέλεσμα να παίρνουμε ότι:

$$\mathbb{E}\left[N_l\right] = \mathbb{E}\left[\mathbb{E}\left[N_l|Y\right]\right] = \mathbb{E}\left[\mathcal{O}\left(\min\left(1, \max\left(2^{-l}, \nu^{-r}\right)\right)\right)\right].$$

Το συγκεκριμένο βήμα ολοκληρώνεται μέσα από την εκτέλεση πανομοιότυπων πράξεων με εκείνες που είδαμε στην απόδειξη του θεωρήματος 3.4.3, δίνοντας τελικά τη σχέση:

$$\mathbb{E}\left[N_l\right] = \mathcal{O}\left(2^l\right).$$

Βήμα 3

Στο τρίτο βήμα θα επιχειρήσουμε να φράξουμε τη ζητούμενη διασπορά από δυο όρους τάξης $\mathcal{O}\left(2^{-l}\right)$, τους οποίους θα αναλύσουμε διεξοδικά στη μελέτη που θα επακολουθήσει. Σε αυτή μας την προσπάθεια θα συμπεριλάβουμε τόσο την περίπτωση που η αριθμητική προσέγγιση $\overline{\sigma}$ υποεκτιμά την τιμή σ όσο και την αντίθετη περίπτωση με τη βοήθεια μιας σταθεράς $0 < b < 1$. Ειδικότερα, σε ένα πρώτο στάδιο, ορίζουμε την τυχαία μεταβλητή $G_l := H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)$ και για τη ζητούμενη διασπορά λαμβάνουμε ότι:

$$Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] \overset{\text{ορ.}}{=} Var\left[G_l\right] = \mathbb{E}\left[G_l^2\right] - \mathbb{E}\left[G_l\right]^2 \leq \mathbb{E}\left[G_l^2\right].$$

Με βάση την παραπάνω ανισότητα και αξιοποιώντας την ταυτότητα $\mathbb{E}\left[\mathbb{E}\left[G_l^2|Y\right]\right] = \mathbb{E}\left[G_l^2\right]$ έχουμε τελικά πως:

$$\begin{aligned} Var\left[H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right] &\leq \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[G_l^2.1|Y\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[G_l^2\left(\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2} + \mathbb{1}_{b^2\sigma^2>\overline{\sigma}^2}\right)|Y\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right]\right] + \mathbb{E}\left[\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2>\overline{\sigma}^2}|Y\right]\right]. \end{aligned}$$

Γνωρίζουμε όμως ότι για κάθε ενδεχόμενο H που εκπληρώνεται με θετική πιθανότητα ικανοποιείται η σχέση:

$$\mathbb{E}[X|H] = \frac{\mathbb{E}[X.\mathbb{1}_H]}{\mathbb{P}[H]} \Longrightarrow \mathbb{E}\left[X.\mathbb{1}_H\right] = \mathbb{E}[X|H].\mathbb{P}[H],$$

συνεπώς η δεύτερη αναμενόμενη τιμή μπορεί να γραφτεί στη μορφή:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2>\overline{\sigma}^2}|Y\right]\right] = \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right]\mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2|Y\right]\right].$$

Έτσι λοιπόν, κάνοντας χρήση του συμβολισμού που εισαγάγαμε, καταλήγουμε πως:

$$Var\left[G_l\right] \leq \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right]\mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2|Y\right]\right] + \mathbb{E}\left[\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right]\right].$$

Αυτό που απομένει τώρα να δείξουμε είναι πως οι δύο παραπάνω αναμενόμενες τιμές είναι τάξης $\mathcal{O}\left(2^{-l}\right)$.

Βήμα 4

Στο σημείο αυτό θα δούμε αναλυτικά πως η μία εκ των δύο αναμενόμενων τιμών του προηγούμενου βήματος είναι τάξης 2^{-l} και ειδικότερα θα οδηγηθούμε στο αποτέλεσμα:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right]\mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2|Y\right]\right] = \mathcal{O}\left(2^{-l}\right).$$

Καταρχάς, για την πιθανότητα εντός της παραπάνω μέσης τιμής έχουμε ότι:

$$\mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2|Y\right] = \mathbb{P}\left[\sigma^2 - \overline{\sigma}^2 > \left(1-b^2\right)\sigma^2|Y\right] \le \mathbb{P}\left[\left|\sigma^2 - \overline{\sigma}^2\right| > \left(1-b^2\right)\sigma^2|Y\right].$$

Με τη βοήθεια όμως του λήμματος 4.2.2 για τη σταθερά $c_1 = 1 - b^2 > 0$ λαμβάνουμε πως:

$$\mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2|Y\right] \le \mathbb{P}\left[\left|\sigma^2 - \overline{\sigma}^2\right| > \left(1-b^2\right)\sigma^2|Y\right] = \mathcal{O}\left(2^{-ql/4}\right). \quad (4.11)$$

Επιπλέον, παρατηρούμε πως το δείγμα που χρησιμοποιείται για τον προσδιορισμό της ποσότητας G_l είναι ανεξάρτητο από εκείνο στο οποίο στηριζόμαστε για την εκτίμηση του εσωτερικού δείγματος N_l. Έτσι, από βασικές ιδιότητες της δεσμευμένης μέσης τιμής παίρνουμε ότι:

$$\mathbb{E}\left[G_l^2|Y\right] = \mathbb{E}\left[G_l^2|Y, N_l\right],$$

με αποτέλεσμα από την ιδιότητα του πύργου να προκύπτει τελικά πως:

$$\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right] = \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y\right]|Y, b^2\sigma^2 > \overline{\sigma}^2\right] = \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, N_l\right]|Y, b^2\sigma^2 > \overline{\sigma}^2\right].$$

Ακολουθώντας τώρα τη λογική του λήμματος 3.2.3 και λαμβάνοντας υπόψη μας το γεγονός πως το εσωτερικό πλήθος δείγματος έχει πλήρως καθοριστεί, λόγω της δέσμευσης που έχουμε εφαρμόσει, βλέπουμε ότι:

$$\begin{aligned}\mathbb{E}\left[G_l^2|Y, N_l\right] &= \mathbb{E}\left[\left(H\left(\overline{\mathbb{E}}_{N_l}[X|Y]\right) - H\left(\mathbb{E}[X|Y]\right)\right)^2|Y, N_l\right] \\ &\le \mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| \ge d|Y, N_l\right] \\ &\le \min\left(1, \delta^{-2}N_l^{-1}\right).\end{aligned}$$

Έτσι, με βάση την τελευταία σχέση έχουμε πως:

$$\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right] = \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, N_l\right]|Y, b^2\sigma^2 > \overline{\sigma}^2\right] \le \mathbb{E}\left[\min\left(1, \delta^{-2}N_l^{-1}\right)|Y, b^2\sigma^2 > \overline{\sigma}^2\right].$$

Από την κατασκευή όμως του αλγορίθμου μας ισχύει ότι:

$$N_l \ge N_0 2^l \Longrightarrow N_l^{-1} \le N_0^{-1}2^{-l} \Longrightarrow \min\left(1, \delta^{-2}N_l^{-1}\right) \le \min\left(1, \delta^{-2}N_0^{-1}2^{-l}\right).$$

Παρατηρούμε τώρα πως ο τελικός μας όρος δοθέντος της μεταβλητής Y προσδιορίζεται πλήρως, με αποτέλεσμα για τη μέση τιμή που εξετάζουμε να προκύπτει ότι:

$$\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right] \le \min\left(1, \delta^{-2}N_0^{-1}2^{-l}\right),$$

απ΄ όπου και παίρνουμε εύκολα την ανισοτική σχέση:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right]\right] \le \mathbb{E}\left[\min\left(1, \delta^{-2}N_0^{-1}2^{-l}\right)\right].$$

Θεωρούμε στο σημείο αυτό πως η τυχαία μεταβλητή δ έχει συνάρτηση πυκνότητας πιθανότητας ρ, η οποία είναι τοπικά φραγμένη κοντά στο 0 σύμφωνα με τα δεδομένα μας. Έτσι, εξ' ορισμού θα υπάρχουν σταθερές $\rho_0 > 0$ και $\delta_0 > 0$ για τις οποίες ικανοποιείται η σχέση:

$\rho(\delta) \leq \rho_0, \, \forall \, \delta \in [0, \delta_0]$.

Αξιοποιώντας το παραπάνω αποτέλεσμα και κάνοντας χρήση των λημμάτων 3.2.1, 3.2.2 λαμβάνουμε ότι:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[G_l^2 | Y, b^2\sigma^2 > \overline{\sigma}^2\right]\right] &\leq \mathbb{E}\left[\min\left(1, \delta^{-2} N_0^{-1} 2^{-l}\right)\right] \\
&= \int_0^\infty \min\left(1, \delta^{-2} N_0^{-1} 2^{-l}\right) \rho(\delta) d\delta \\
&\leq \rho_0 \int_0^\infty \min\left(1, \delta^{-2} N_0^{-1} 2^{-l}\right) d\delta + \min\left(1, \delta_0^{-2} N_0^{-1} 2^{-l}\right) \\
&\leq \rho_0 \int_0^\infty \min\left(1, \delta^{-2} N_0^{-1} 2^{-l}\right) d\delta + \delta_0^{-2} N_0^{-1} 2^{-l} \\
&= 2\rho_0 \left(N_0^{-1} 2^{-l}\right)^{1/2} + \delta_0^{-2} N_0^{-1} 2^{-l} \\
&= 2\rho_0 N_0^{-1/2} 2^{-l/2} + \delta_0^{-2} N_0^{-1} 2^{-l}.
\end{aligned}$$

Συνεπώς, με τη βοήθεια της τελευταίας σχέσης συμπεραίνουμε πως:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2 | Y, b^2\sigma^2 > \overline{\sigma}^2\right]\right] = \mathcal{O}\left(2^{-l/2}\right), \quad (4.12)$$

και για την τάξη του ζητούμενου όρου προκύπτει ότι:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[G_l^2 | Y, b^2\sigma^2 > \overline{\sigma}^2\right] \mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2 | Y\right]\right] &= \mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2 | Y\right] \mathbb{E}\left[\mathbb{E}\left[G_l^2 | Y, b^2\sigma^2 > \overline{\sigma}^2\right]\right] \\
&\underset{(4.12)}{\overset{(4.11)}{=}} \mathcal{O}\left(2^{-ql/4}\right) \mathcal{O}\left(2^{-l/2}\right) \\
&= \mathcal{O}\left(2^{-l/2 - ql/4}\right).
\end{aligned}$$

Βλέπουμε όμως πως, σύμφωνα με τα δεδομένα μας, ισχύει ότι:

$$q > 2 \Longrightarrow \tfrac{q}{4} > \tfrac{1}{2} \Longrightarrow -\tfrac{q}{4} < -\tfrac{1}{2} \Longrightarrow -l\tfrac{q}{4} < -\tfrac{l}{2} \Longrightarrow 2^{-lq/4} < 2^{-l/2},$$

οπότε τελικά:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2 | Y, b^2\sigma^2 > \overline{\sigma}^2\right] \mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2 | Y\right]\right] = \mathcal{O}\left(2^{-l/2 - l/2}\right) = \mathcal{O}\left(2^{-l}\right).$$

Ακολουθούν τώρα δύο βοηθητικά βήματα, τα οποία θα παίξουν καθοριστικό ρόλο στην εξαγωγή του αποτελέσματος:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2 \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2} | Y\right]\right] = \mathcal{O}\left(2^{-l}\right).$$

Βήμα 5

Στο συγκεκριμένο βήμα, θα αναζητήσουμε ένα άνω φράγμα για την πιθανότητα ο αλγόριθμός μας να τερματίσει σε ένα αρκετά χαμηλότερο επίπεδο από εκείνο που θα παίρναμε έχοντας πλήρη γνώση των d, σ, με ταυτόχρονη επιβολή της συνθήκης $b^2\sigma^2 \leq \overline{\sigma}^2$. Για τη διευκόλυνση του αναγνώστη, θα διατηρήσουμε το συμβολισμό l^* από το βήμα 1 και θα ορίσουμε ένα επίπεδο l' το οποίο ικανοποιεί τη σχέση $l' \leq l^* - 3$. Προφανώς, η τελευταία συνθήκη μας εξασφαλίζει πως το νέο αυτό επίπεδο βρίσκεται αρκετά χαμηλότερα του θεωρητικού επιπέδου l^*. Τότε, θα θέλαμε ο αλγόριθμός μας να σταματά στο επίπεδο αυτό ή ισοδύναμα θα θέλαμε να ικανοποιείται η συνθήκη $\hat{l} = l'$, όπου με $\hat{l}$ έχουμε συμβολίσει το επίπεδο τερματισμού. Εφόσον λοιπόν προχωράμε στην παραδοχή πως ο αλγόριθμός μας ολοκληρώνεται στο επίπεδο l', δίνοντας κατ΄ επέκταση ένα πλήθος εσωτερικού δείγματος $N_l = N_0 2^{l'}$, από τη συνθήκη τερματισμού λαμβάνουμε ότι:

$$N_l \geq N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \frac{\overline{d}}{\overline{\sigma}}\right)^{-r} \overset{\text{op.}}{\Longrightarrow} N_0 2^{l'} \geq N_0 4^l \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}\right)^{-r} \Longrightarrow \left(C^{-1} N_0^{1/2} 2^l \overline{\delta}\right)^{-r} \leq \frac{2^{l'}}{4^l}.$$

Αν στο σημείο αυτό εκμεταλλευτούμε το γεγονός πως για το επίπεδο που εξετάζουμε έχει ισχύ η συνθήκη $l' \leq l^* - 3$ οδηγούμαστε στο αποτέλεσμα:

$$\left(C^{-1} N_0^{1/2} 2^l \overline{\delta}\right)^{-r} \leq \frac{2^{l'}}{4^l} \leq \frac{2^{l^*-3}}{4^l} = 2^{-2} \frac{2^{l^*-1}}{4^l} = \frac{1}{4} \frac{2^{l^*-1}}{4^l}.$$

Με τη βοήθεια τώρα του βήματος 1, βλέπουμε πως το επίπεδο l^* πληροί την ανισοτική σχέση:

$$\frac{2^{l^*-1}}{4^l} < \left(C^{-1} N_0^{1/2} 2^l \delta\right)^{-r},$$

οπότε τελικά:

$$\left(C^{-1} N_0^{1/2} 2^l \overline{\delta}\right)^{-r} < \frac{1}{4} \left(C^{-1} N_0^{1/2} 2^l \delta\right)^{-r} \Longrightarrow \overline{\delta} > 4^{1/r} \delta \Longrightarrow \frac{\overline{d}}{\overline{\sigma}} > 4^{1/r} \frac{d}{\sigma} \Longrightarrow \frac{\overline{d}}{d} \frac{\sigma}{\overline{\sigma}} > 4^{1/r},$$

για κάθε $l \leq l' \leq l^* - 3$ και $l + 3 \leq l^* \leq 2l$. Έτσι λοιπόν, δείξαμε πως η συνθήκη $\hat{l} = l'$ οδηγεί κατ΄ ανάγκη στην ανισοτική σχέση:

$$\frac{\overline{d}}{d} \frac{\sigma}{\overline{\sigma}} > 4^{1/r},$$

συνεπώς παίρνουμε ότι:

$$\mathbb{P}\left[\hat{l} = l', b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right] \leq \mathbb{P}\left[\frac{\overline{d}}{d} \frac{\sigma}{\overline{\sigma}} > 4^{1/r}, b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right].$$

Με βάση το δεύτερο ενδεχόμενο της παραπάνω πιθανότητας βλέπουμε πως:

$$b^2\sigma^2 \leq \overline{\sigma}^2 \Longrightarrow \frac{\sigma^2}{\overline{\sigma}^2} \leq \frac{1}{b^2} \Longrightarrow \frac{1}{b} \geq \frac{\sigma}{\overline{\sigma}} \Longrightarrow \frac{\overline{d}}{d} \frac{1}{b} \geq \frac{\overline{d}}{d} \frac{\sigma}{\overline{\sigma}} > 4^{1/r},$$

απ΄ όπου και προκύπτει ότι:

$$\begin{aligned}
\mathbb{P}\left[\hat{l} = l', b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right] &\leq \mathbb{P}\left[\frac{\overline{d}}{d} \frac{\sigma}{\overline{\sigma}} > 4^{1/r}, b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right] \\
&\leq \mathbb{P}\left[\frac{\overline{d}}{d} \frac{1}{b} > 4^{1/r}, b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right] \\
&= \mathbb{P}\left[\overline{d} > 4^{1/r} b d, b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right].
\end{aligned}$$

Παρατηρούμε όμως πως για το πρώτο ενδεχόμενο που εμφανίζεται στην τελευταία πιθανότητα ισχύει ότι:

$$\overline{d} > 4^{1/r} b d \Longrightarrow d < 4^{-1/r} b^{-1} \overline{d} \Longrightarrow -d > -4^{-1/r} b^{-1} \overline{d} \Longrightarrow \overline{d} - d > \left(1 - 4^{-1/r} b^{-1}\right) \overline{d},$$

επομένως καταλήγουμε πως:

$$\begin{aligned}\mathbb{P}\left[\hat{l}=l', b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] &\le \mathbb{P}\left[\overline{d} > 4^{1/r} bd, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] \\ &\le \mathbb{P}\left[\overline{d} - d > \left(1 - 4^{-1/r} b^{-1}\right)\overline{d}, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] \\ &\le \mathbb{P}\left[\left|\overline{d} - d\right| > \left(1 - 4^{-1/r} b^{-1}\right)\overline{d}, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right].\end{aligned}$$

Με βάση τώρα προγενέστερη ανάλυσή μας βλέπουμε πως ισχύει ότι:

$$\left(C^{-1} N_0^{1/2} 2^l \overline{\delta}\right)^{-r} \le \frac{2^{l'}}{4^l} = 2^{l'-2l} \Longrightarrow C^{-1} N_0^{1/2} 2^l \overline{\delta} \ge 2^{\left(2l-l'\right)/r} \Longrightarrow \overline{\delta} \ge C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r}.$$

Αξιοποιώντας τώρα τον ορισμό της ποσότητας $\overline{\delta}$ αλλά και το δεύτερο ενδεχόμενο εντός της πιθανότητάς μας έχουμε πως:

$$\frac{\overline{d}}{\overline{\sigma}} \ge C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r} \Longrightarrow \overline{d} \ge \overline{\sigma} C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r} \ge \sigma b C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r}.$$

Με τη βοήθεια της σχέσης αυτής βλέπουμε ότι:

$$\begin{aligned}\mathbb{P}\left[\hat{l}=l', b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] &\le \mathbb{P}\left[\left|\overline{d} - d\right| > \left(1 - 4^{-1/r} b^{-1}\right)\overline{d}, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] \\ &\le \mathbb{P}\left[\left|\overline{d} - d\right| > \left(1 - 4^{-1/r} b^{-1}\right)\sigma b C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r}, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] \\ &= \mathbb{P}\left[\left|\overline{d} - d\right| > \left(b - 4^{-1/r}\right)\sigma C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r}, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right].\end{aligned}$$

Για να συνεχίσουμε τον υπολογισμό μας, κάνουμε για τη σταθερά b που εισάγαμε στο βήμα 3 την παραδοχή πως:

$$4^{-1/r} < b < 1 \Longrightarrow b - 4^{-1/r} > 0.$$

Τότε, κάνοντας χρήση του λήμματος 3.3.2 παίρνουμε ότι:

$$\begin{aligned}\mathbb{P}\left[\hat{l}=l', b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] &\le \mathbb{P}\left[\left|\overline{d} - d\right| > \left(b - 4^{-1/r}\right)\sigma C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r}, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] \\ &\overset{\text{ορ.}}{=} \mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| > \left(b - 4^{-1/r}\right)\sigma C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r}, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] \\ &\le \mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_0 2^{l'}}[X|Y] - \mathbb{E}[X|Y]\right| > \left(b - 4^{-1/r}\right)\sigma C N_0^{-1/2} 2^{-l} 2^{\left(2l-l'\right)/r} | Y\right] \\ &\le C_q \sigma^q \kappa_q \left(b - 4^{-1/r}\right)^{-q} C^{-q} \sigma^{-q} N_0^{q/2} 2^{lq} 2^{-\left(2l-l'\right)q/r} \left(N_0 2^{l'}\right)^{-q/2} \\ &= \left(b - 4^{-1/r}\right)^{-q} C_q \kappa_q C^{-q} N_0^{q/2} 2^{lq} 2^{-\left(2l-l'\right)q/r} N_0^{-q/2} 2^{-l'q/2} \\ &= \left(b - 4^{-1/r}\right)^{-q} C_q \kappa_q C^{-q} 2^{q\left(2l-l'\right)/2} 2^{-\left(2l-l'\right)q/r} \\ &= \left(b - 4^{-1/r}\right)^{-q} C_q \kappa_q C^{-q} 2^{-q\left(2l-l'\right)(1/r-1/2)}.\end{aligned}$$

Αξίζει να αναφέρουμε ότι για την εξαγωγή της δεύτερης ανισοτικής σχέσης στηριχτήκαμε στο γεγονός πως για οποιαδήποτε δύο ενδεχόμενα A, B ισχύει η ανισότητα:

$$\mathbb{P}[A, B] \le \mathbb{P}[A].$$

Εύκολα αντιλαμβανόμαστε πως η ανάλυση που προηγήθηκε μας οδηγεί τελικά στο αποτέλεσμα:

$$\mathbb{P}\left[\hat{l}=l', b^2\sigma^2 \leq \overline{\sigma}^2|Y\right] = \mathcal{O}\left(2^{-q\left(2l-l'\right)(1/r-1/2)}\right),$$

για κάθε επίπεδο l' με $l \leq l' \leq l^* - 3$, εφόσον ικανοποιείται η συνθήκη $l+3 \leq l^* \leq 2l$. Μπορούμε όμως να δούμε πως το άνω φράγμα που προσδιορίσαμε ισχύει και στην περίπτωση που $l \leq l' \leq l^* - 3$ με $l \leq l^* < l+3$. Συγκεκριμένα, αν θεωρήσουμε πως $l^* < l+3$ παρατηρούμε πως ο αλγόριθμός μας δεν μπορεί να σταματήσει στο επίπεδο l' μιας και το τελευταίο θα είναι υποχρεωτικά μικρότερο από την τιμή l. Με άλλα λόγια, κάτω από το εξεταζόμενο πλαίσιο θα ισχύει ότι:

$$\mathbb{P}\left[\hat{l}=l', b^2\sigma^2 \leq \overline{\sigma}^2|Y\right] = 0 = \mathcal{O}\left(2^{-q\left(2l-l'\right)(1/r-1/2)}\right)$$

και τελικά το άνω φράγμα που βρήκαμε έχει ισχύ για όλα τα επίπεδα $l \leq l^* \leq 2l$.

Βήμα 6

Στο σημείο αυτό, με τη βοήθεια του βήματος που προηγήθηκε, θα αναζητήσουμε ένα άνω φράγμα για την αναμενόμενη τιμή:

$$\mathbb{E}\left[\left(\frac{N_l}{N_0 4^l}\right)^{-p/2} \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right],$$

θεωρώντας καταρχάς πως η σταθερά p ορίζεται κατά τέτοιο τρόπο ώστε να ικανοποιεί τη συνθήκη:

$$\frac{2}{2-r} < p < q.$$

Ο λόγος για τον οποίο εστιάζουμε στη συγκεκριμένη μέση τιμή θα φανεί λεπτομερώς στο επόμενο και προτελευταίο βήμα της απόδειξής μας. Βλέπουμε τώρα πως η μέση τιμή που καλούμαστε να διαχειριστούμε περιλαμβάνει δύο τυχαίες μεταβλητές, V, W, οι οποίες δίνονται μέσω των σχέσεων:

$$V \equiv \left(\frac{N_l}{N_0 4^l}\right)^{-p/2} \text{ και } W \equiv \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}.$$

Όσον αφορά τη δεύτερη τυχαία μεταβλητή είναι σαφές πως αυτή λαμβάνει τις τιμές 0 και 1, ενώ η πρώτη, δοθέντος πως $N_l = N_0 2^{l'}$, θα παίρνει ακριβώς τις τιμές:

$$\left(\frac{N_0 2^{l'}}{N_0 4^l}\right)^{-p/2} = \left(2^{l'-2l}\right)^{-p/2} = 2^{p\left(2l-l'\right)/2},$$

όπου $l \leq l' \leq 2l$. Αν συμβολίσουμε λοιπόν με W_1 την τιμή 1 της μεταβλητής W και με $V_{l'}$ την τιμή $2^{p\left(2l-l'\right)/2}$ της μεταβλητής V έχουμε ότι:

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{N_l}{N_0 4^l}\right)^{-p/2} \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right] &= \sum_{i=1}^{2}\sum_{l'=l}^{2l} W_i V_{l'}\, \mathbb{P}\left[V=V_{l'},\, W=W_i|Y\right] \\
&= \sum_{l'=l}^{2l} 1.V_{l'}\, \mathbb{P}\left[V=V_{l'},\, W=1|Y\right] + \sum_{l'=l}^{2l} 0.V_{l'}\, \mathbb{P}\left[V=V_{l'},\, W=0|Y\right] \\
&= \sum_{l'=l}^{2l} V_{l'}\, \mathbb{P}\left[\left(\frac{N_l}{N_0 4^l}\right)^{-p/2} = \left(\frac{N_0 2^{l'}}{N_0 4^l}\right)^{-p/2},\, \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2} = 1|Y\right] \\
&= \sum_{l'=l}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'},\, b^2\sigma^2 \leq \overline{\sigma}^2|Y\right] 2^{p\left(2l-l'\right)/2}.
\end{aligned}$$

Όπως έχουμε όμως ήδη εξηγήσει το ενδεχόμενο $N_l = N_0 2^{l'}$ εκπληρώνεται ακριβώς όταν $\hat{l} = l'$ και αντίστροφα, οπότε λαμβάνουμε ότι:

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{N_l}{N_0 4^l}\right)^{-p/2} \mathbb{1}_{b^2\sigma^2 \le \overline{\sigma}^2} | Y\right] &= \sum_{l'=l}^{2l} \mathbb{P}\left[N_l = N_0 2^{l'},\, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2} \\
&= \sum_{l'=l}^{2l} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2} \\
&= \sum_{l'=l}^{l^*-3} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2} + \sum_{l'=l^*-2}^{2l} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2}.
\end{aligned}
$$

Θα εξετάσουμε στο σημείο αυτό το κάθε ένα από τα παραπάνω αθροίσματα χωριστά. Έτσι, βλέπουμε πως ισχύει ότι:

α) Αξιοποιώντας το βήμα 5 για το πρώτο άθροισμα παίρνουμε πως:

$$
\begin{aligned}
\sum_{l'=l}^{l^*-3} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2} &= \sum_{l'=l}^{l^*-3} \mathcal{O}\left(2^{-q\left(2l-l'\right)(1/r-1/2)}\right) 2^{p\left(2l-l'\right)/2} \\
&= \sum_{l'=l}^{l^*-3} \mathcal{O}\left(2^{-q\left(2l-l'\right)(1/r-1/2)+p\left(2l-l'\right)/2}\right) \\
&= \sum_{l'=l}^{l^*-3} \mathcal{O}\left(2^{\left(2l-l'\right)(-2q/r+q+p)/2}\right) \\
&= \sum_{l'=l}^{l^*-3} \mathcal{O}\left(2^{\left(2l-l'\right)(q(r-2)/r+p)/2}\right) \\
&\equiv \sum_{l'=l}^{l^*-3} \mathcal{O}\left(2^{\left(2l-l'\right)u/2}\right) \\
&\le c \sum_{l'=l}^{l^*-3} 2^{\left(2l-l'\right)u/2} \\
&= c 2^{ul} \sum_{l'=l}^{l^*-3} \left(2^{-u/2}\right)^{l'}
\end{aligned}
$$

Με τη βοήθεια τώρα της γνωστής ταυτότητας για το πεπερασμένο άθροισμα όρων γεωμετρικής προόδου προκύπτει ότι:

$$
\sum_{l'=l}^{l^*-3} \left(2^{-u/2}\right)^{l'} = 2^{-ul/2} \cdot \frac{2^{-u\left(l^*-2-l\right)/2}-1}{2^{-u/2}-1} = \frac{2^{-u\left(l^*-2\right)/2}-2^{-ul/2}}{2^{-u/2}-1} = \frac{2^{u\left(2-l^*\right)/2}-2^{-ul/2}}{2^{-u/2}-1}.
$$

Έτσι, οδηγούμαστε τελικά στο αποτέλεσμα:

$$
\sum_{l'=l}^{l^*-3} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2} \le c 2^{ul} \cdot \frac{2^{u\left(2-l^*\right)/2}-2^{-ul/2}}{2^{-u/2}-1} = c \cdot \frac{2^{u\left(2l-l^*+2\right)/2}-2^{ul/2}}{2^{-u/2}-1},
$$

το οποίο συνεπάγεται πως:

$$
\sum_{l'=l}^{l^*-3} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \le \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2} = \mathcal{O}\left(\frac{2^{u\left(2l-l^*+2\right)/2}-2^{ul/2}}{2^{-u/2}-1}\right).
$$

β) Για το δεύτερο άθροισμα χρησιμοποιούμε την προφανή ανισοτική σχέση:

$$\mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right] \leq 1$$

και έχουμε ότι:

$$\begin{aligned}
\sum_{l'=l^*-2}^{2l} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right] 2^{p\left(2l-l'\right)/2} &\leq \sum_{l'=l^*-2}^{2l} 2^{p\left(2l-l'\right)/2} \\
&= 2^{pl} \sum_{l'=l^*-2}^{2l} 2^{-pl'/2} \\
&= 2^{pl} \sum_{l'=l^*-2}^{2l} \left(2^{-p/2}\right)^{l'} \\
&< 2^{pl} \sum_{l'=l^*-2}^{\infty} \left(2^{-p/2}\right)^{l'} \\
&= 2^{pl} \,.\, \tfrac{2^{-p\left(l^*-2\right)/2}}{1-2^{-p/2}} \\
&= \tfrac{1}{1-2^{-p/2}} \,.\, 2^{\,p(2l-l^*+2)/2}
\end{aligned}$$

Εξετάζοντας τις παραπάνω πράξεις, εύκολα αντιλαμβανόμαστε πως:

$$\sum_{l'=l^*-2}^{2l} \mathbb{P}\left[\hat{l} = l',\, b^2\sigma^2 \leq \overline{\sigma}^2 | Y\right] = \mathcal{O}\left(2^{\,p(2l-l^*)/2}\right).$$

Έτσι λοιπόν, με βάση το άνω φράγμα που προσδιορίσαμε για τα δύο μας αθροίσματα, η μέση τιμή που μελετάμε ικανοποιεί τη σχέση:

$$\mathbb{E}\left[\left(\tfrac{N_l}{N_0 4^l}\right)^{-p/2} \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2} | Y\right] = \mathcal{O}\left(2^{\,p(2l-l^*)/2}\right) + \mathcal{O}\left(\tfrac{2^{u\left(2l-l^*+2\right)/2} - 2^{ul/2}}{2^{-u/2}-1}\right).$$

Θα προσπαθήσουμε τώρα να απλοποιήσουμε την παραπάνω παράσταση αναζητώντας το πρόσημο της σταθεράς $u = q(r-2)/r+p$. Για να το επιτύχουμε αυτό θα πρχωρήσουμε σε ένα τέχνασμα, το οποίο σε μια πρώτη φάση στηρίζεται στην εύρεση των ριζών του τριωνύμου $(2-r)^2\, q - 2r$. Ειδικότερα, έχουμε ότι:

$$(2-r)^2\, q - 2r = 0 \Longrightarrow \left(r^2 - 4r + 4\right) q - 2r = 0 \Longrightarrow qr^2 - (4q+2)\, r + 4q = 0,$$

οπότε η διακρίνουσα δίνεται μέσω του τύπου:

$$\Delta = (4q+2)^2 - 16q^2 = 16q^2 + 16q + 4 - 16q^2 = 16q + 4.$$

Με βάση το παραπάνω αποτέλεσμα, συμπεραίνουμε πως οι ρίζες του τριωνύμου που εξετάζουμε, θα φέρουν τη μορφή:

$$r_{1,2} = \tfrac{4q+2\pm\sqrt{16q+4}}{2q} = \tfrac{4q+2\pm 2\sqrt{4q+1}}{2q} = 2 + \tfrac{1\pm\sqrt{4q+1}}{q} = 2 - \tfrac{\pm\sqrt{4q+1}-1}{q}.$$

Εφόσον ο μεγιστοβάθμιος όρος του συγκεκριμένου τριωνύμου είναι θετικός, γνωρίζουμε πως αυτό θα είναι αρνητικό εντός των δύο ριζών του και θετικό έξω από αυτές. Από τα δεδομένα μας όμως παίρνουμε πως:

$$r < 2 - \frac{\sqrt{4q+1}-1}{q},$$

όπου η τιμή αυτή δεν είναι τίποτα άλλο από τη μικρότερη ρίζα του τριωνύμου μας. Συνεπώς, εφόσον έχει ισχύ η συγκεκριμένη συνθήκη, λαμβάνουμε υποχρεωτικά ότι:

$$qr^2 - (4q+2)\,r + 4q > 0 \Longrightarrow (2-r)^2\, q - 2r > 0 \Longrightarrow q > \frac{2r}{(2-r)^2} \Longrightarrow \frac{q\,(2-r)}{r} > \frac{2}{2-r}.$$

Επιπλέον, βλέπουμε πως ισχύει και:

$$r > 1 \Longrightarrow 2r > 2 \Longrightarrow r > 2-r \Longrightarrow \frac{2-r}{r} < 1 \Longrightarrow \frac{q\,(2-r)}{r} < q.$$

Έτσι, με βάση τα παραπάνω μπορούμε να επιλέξουμε τη σταθερά p ώστε η τελευταία να ικανοποιεί τη σχέση:

$$\frac{2}{2-r} < p < \frac{q\,(2-r)}{r} \Longrightarrow \frac{q\,(r-2)}{r} + p < 0 \overset{\text{ορ.}}{\Longrightarrow} u < 0.$$

Εύκολα τώρα βλέπουμε πως υπό αυτό το πλαίσιο ισχύει ότι:

$$\frac{2^{u(2l-l^*+2)/2} - 2^{ul/2}}{2^{-u/2}-1} < \frac{1}{2^{-u/2}-1}\,.\,2^{u(2l-l^*+2)/2} = \mathcal{O}\left(2^{u(2l-l^*)/2}\right)$$

και αφού δείξαμε πως $u < 0 < p$ οδηγούμαστε στο αποτέλεσμα:

$$\mathbb{E}\left[\left(\frac{N_l}{N_0 4^l}\right)^{-p/2} \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2} | Y\right] = \mathcal{O}\left(2^{p(2l-l^*)/2}\right) + \mathcal{O}\left(2^{u(2l-l^*)/2}\right) = \mathcal{O}\left(2^{p(2l-l^*)/2}\right).$$

Στο βήμα 1 όμως, δείξαμε πως για το επίπεδο l^* ικανοποιείται η συνθήκη:

$$\left(C^{-1} N_0^{1/2} 2^l \delta\right)^{-r} \leq \frac{2^{l^*}}{4^l} \Longrightarrow \left(C^{-1} N_0^{1/2} 2^l \delta\right)^{-r} \leq 2^{l^*-2l} \overset{\text{ορ.}}{\Longrightarrow} \nu^{-r} \leq 2^{l^*-2l} \Longrightarrow 2^{p(2l-l^*)/2} \leq \nu^{rp/2},$$

ενώ λαμβάνουμε και:

$$l \leq l^* \Longrightarrow -l^* \leq -l \Longrightarrow 2l - l^* \leq l \Longrightarrow 2^{p(2l-l^*)/2} \leq 2^{lp/2}.$$

Συνδυάζοντας τις παραπάνω σχέσεις, παρατηρούμε πως:

$$2^{p(2l-l^*)/2} \leq \min\left(2^{lp/2}, \nu^{rp/2}\right) \leq \max\left(1, \min\left(2^{lp/2}, \nu^{rp/2}\right)\right),$$

γεγονός που μας οδηγεί στο τελικό άνω φράγμα:

$$\mathbb{E}\left[\left(\frac{N_l}{N_0 4^l}\right)^{-p/2} \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2} | Y\right] = \mathcal{O}\left(\max\left(1, \min\left(2^{lp/2}, \nu^{rp/2}\right)\right)\right).$$

Βήμα 7

Στο βήμα αυτό, με τη βοήθεια του αποτελέσματος που μόλις δείξαμε, θα οδηγηθούμε σε ένα φράγμα τάξης 2^{-l} και για τη δεύτερη αναμενόμενη τιμή στην ανισοτική σχέση της ζητούμενης διασποράς. Λόγω του γεγονότος πως μεσολάβησαν τρία αρκετά μεγάλα βήματα υπενθυμίζουμε πως ο συγκεκριμένος όρος δινόταν μέσω της έκφρασης:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2 \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right]\right],$$

όπου η σταθερά b κυμαίνεται μεταξύ των τιμών 0 και 1. Σε ένα πρώτο στάδιο, για την εσωτερική μέση τιμή έχουμε ότι:

$$\mathbb{E}\left[G_l^2 \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right] = \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, \overline{d}, \overline{\sigma}^2\right] \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right].$$

Είναι σημαντικό να αναφέρουμε ότι για την εξαγωγή της παραπάνω ισότητας αξιοποιήσαμε το γεγονός πως το δείγμα για τον υπολογισμό του G_l είναι ανεξάρτητο από εκείνο για την εκτίμηση των τιμών d και σ. Ακολουθώντας τώρα τη λογική που είδαμε τόσο στο βήμα 4, όσο και σε προηγούμενη πρόταση παίρνουμε ότι:

$$\begin{aligned}
\mathbb{E}\left[G_l^2 \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right] &= \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, \overline{d}, \overline{\sigma}^2\right] \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right] \\
&\leq \mathbb{E}\left[\mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y, \overline{d}, \overline{\sigma}^2\right] \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right].
\end{aligned}$$

Είναι τώρα εύκολο να δούμε, με τη βοήθεια του αλγορίθμου μας, πως η δέσμευση που έχουμε επιβάλλει προσδιορίζει πλήρως το πλήθος του εσωτερικού δείγματος N_l. Έτσι, κάνοντας χρήση της εφαρμογής που απορρέει από το λήμμα 3.3.2 λαμβάνουμε πως:

$$\mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y, \overline{d}, \overline{\sigma}^2\right] \leq \min\left(1, C_q \kappa_q \left(\delta N_l^{1/2}\right)^{-q}\right),$$

μιας και από την υπόθεσή μας υπάρχει θετικός ακέραιος q τέτοιος ώστε $\kappa_q < +\infty$. Αν τώρα ορίσουμε μια σταθερά p με:

$$1 < \frac{2}{2-r} < p < \frac{q(2-r)}{r} < q, \quad (4.13)$$

εύκολα αντιλαμβανόμαστε πως για τη μεταβλητή $Z = X(Y) - \mathbb{E}[X|Y]$ θα υπάρχει και η p-οστή ροπή, γεγονός που μας οδηγεί στο αποτέλεσμα:

$$\mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y, \overline{d}, \overline{\sigma}^2\right] \leq \min\left(1, C_p \kappa_p \left(\delta N_l^{1/2}\right)^{-p}\right).$$

Αξίζει να σημειώσουμε πως η ύπαρξη μιας σταθεράς p που πληρεί τη συνθήκη (4.13) εξασφαλίζεται με βάση τη μελέτη που διεξάγαμε στο αμέσως προηγούμενο βήμα. Έτσι λοιπόν, με τη βοήθεια του άνω φράγματος που προσδιορίσαμε για την πιθανότητά μας παίρνουμε πως:

$$\begin{aligned}
\mathbb{E}\left[G_l^2 \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right] &\leq \mathbb{E}\left[\mathbb{P}\left[\left|\overline{\mathbb{E}}_{N_l}[X|Y] - \mathbb{E}[X|Y]\right| > d|Y, \overline{d}, \overline{\sigma}^2\right] \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right] \\
&\leq \mathbb{E}\left[\min\left(1, C_p \kappa_p \left(\delta N_l^{1/2}\right)^{-p}\right) \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}|Y\right] \\
&= \mathbb{E}\left[\min\left(\mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}, C_p \kappa_p \left(\delta N_l^{1/2}\right)^{-p} \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}\right)|Y\right] \\
&\leq \mathbb{E}\left[\min\left(1, C_p \kappa_p \left(\delta N_l^{1/2}\right)^{-p} \mathbb{1}_{b^2\sigma^2 \leq \overline{\sigma}^2}\right)|Y\right].
\end{aligned}$$

Υπενθυμίζουμε όμως πως σε προηγούμενη ανάλυσή μας προχωρήσαμε στον ορισμό:

$$\nu = C^{-1} N_0^{1/2} 2^l \delta \Longrightarrow \delta = C N_0^{-1/2} 2^{-l} \nu \Longrightarrow \delta^{-p} = C^{-p} N_0^{p/2} 2^{lp} \nu^{-p},$$

ο οποίος μας δίνει ότι:

$$\begin{aligned}\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right] &\leq \mathbb{E}\left[\min\left(1, C_p\kappa_p C^{-p}N_0^{p/2}2^{lp}\nu^{-p}N_l^{-p/2}\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}\right)|Y\right]\\ &= \mathbb{E}\left[\min\left(1, C_p\kappa_p C^{-p}\nu^{-p}\left(\frac{N_l}{N_0 4^l}\right)^{-p/2}\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}\right)|Y\right]\\ &\overset{Jen.}{\leq} \min\left(1, C_p\kappa_p C^{-p}\nu^{-p}\mathbb{E}\left[\left(\frac{N_l}{N_0 4^l}\right)^{-p/2}\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right]\right).\end{aligned}$$

Φράσσοντας τώρα την αναμενόμενη τιμή της παραπάνω σχέσης, όπως είδαμε στο βήμα 6, λαμβάνουμε ότι:

$$\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right] \leq \min\left(1, C_p\kappa_p C^{-p}\nu^{-p}\mathcal{O}\left(\max\left(1, \min\left(2^{lp/2}, \nu^{rp/2}\right)\right)\right)\right),$$

ενώ εφαρμόζοντας στο αποτέλεσμα αυτό μέση τιμή καταλήγουμε πως:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right]\right] \leq \mathbb{E}\left[\min\left(1, C_p\kappa_p C^{-p}\nu^{-p}\mathcal{O}\left(\max\left(1, \min\left(2^{lp/2}, \nu^{rp/2}\right)\right)\right)\right)\right].$$

Για να ολοκληρώσουμε το βήμα αυτό, αρκεί να παρατηρήσουμε ότι:

$$p > \tfrac{2}{2-r} \implies 2p - rp > 2 \implies p - r\tfrac{p}{2} > 1,$$

με αποτέλεσμα ανάλογοι υπολογισμοί με εκείνους που είδαμε στο θεώρημα 3.4.3 να μας οδηγούν στη σχέση:

$$\mathbb{E}\left[\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right]\right] = \mathcal{O}\left(2^{-l}\right).$$

Βήμα 8

Στο τελικό αυτό στάδιο, θα συνδυάσουμε τα αποτελέσματα των βημάτων 4 και 7, με σκοπό να πάρουμε το επιθυμητό φράγμα για τη ζητούμενη διασπορά. Υπενθυμίζουμε πως στο τρίτο βήμα δείξαμε ότι:

$$Var\left[G_l\right] \leq \mathbb{E}\left[\mathbb{E}\left[G_l^2|Y, b^2\sigma^2 > \overline{\sigma}^2\right]\mathbb{P}\left[b^2\sigma^2 > \overline{\sigma}^2|Y\right]\right] + \mathbb{E}\left[\mathbb{E}\left[G_l^2\mathbb{1}_{b^2\sigma^2\leq\overline{\sigma}^2}|Y\right]\right].$$

Αντικαθιστώντας στη σχέση αυτή το άνω φράγμα που έχουμε προσδιορίσει, καταλήγουμε πως:

$$Var\left[G_l\right] = \mathcal{O}\left(2^{-l}\right) + \mathcal{O}\left(2^{-l}\right) = \mathcal{O}\left(2^{-l}\right).$$ □

Παρατήρηση 4.2.3

Το θέωρημα που μόλις αποδείξαμε, μπορεί να οδηγήσει άμεσα στην εύρεση του ζητούμενου φράγματος για την πολυπλοκότητα της μεθόδου που εξετάζουμε. Ειδικότερα, λαμβάνοντας υπόψη μας πως τα αποτελέσματα που εξαγάγαμε για τη διασπορά και τη μέση τιμή από το παραπάνω θεώρημα ταυτίζονται με αυτά που προέκυψαν από το θεώρημα 3.4.3, το πόρισμα 3.4.5, μπορεί να μας δώσει πως η μέθοδος Multilevel Monte Carlo με προσαρμοσμένη επιλογή εσωτερικού δείγματος φέρει ένα μέσο υπολογιστικό κόστος τάξης $\epsilon^{-2}\left(\log\epsilon\right)^2$.

Επίλογος

Ξεκινώντας από την ανάγκη προσδιορισμού της πιθανότητας η απώλεια στην αξία ενός χαρτοφυλακίου να ξεπεράσει ένα δοσμένο φράγμα, επιχειρήσαμε να εκτιμήσουμε αριθμητικά την τιμή $\mathbb{E}[H(\mathbb{E}[X\,|\,Y])]$ μέσα από μια σειρά διαφορετικών προσεγγίσεων. Έτσι λοιπόν, σε ένα πρώτο στάδιο, εφαρμόσαμε ένα διπλό σχήμα απλής μεθόδου Monte Carlo και με χρήση κατάλληλων θεωρημάτων καταλήξαμε πως για να εξασφαλίσουμε ένα μέσο τετραγωνικό σφάλμα τάξης ϵ^2 απαιτείται μια ολική πολυπλοκότητα $\mathcal{C}$ τάξης ϵ^{-3}. Στη συνέχεια, προσαρμόσαμε το εσωτερικό μας δείγμα στις τιμές της μεταβλητής Y και οδηγηθήκαμε στο συμπέρασμα πως η νέα μέθοδος, για το ίδιο επίπεδο ακρίβειας, φέρει προσεγγιστικά μια υπολογιστική πολυπλοκότητα τάξης $\epsilon^{-5/2}\log\left(\epsilon^{-1/2}\right)$, βελτιώνοντας κατ΄ επέκταση το προηγούμενο αποτέλεσμά. Σε ένα επόμενο βήμα, εξετάσαμε κατά πόσο η χρήση ενός σχήματος Multilevel Monte Carlo μπορεί να μειώσει περαιτέρω το ολικό υπολογιστικό κόστος και είδαμε πως κάτι τέτοιο πράγματι επιτυγχάνεται καθώς $\mathcal{C} = \mathcal{O}\left(\epsilon^{-5/2}\right)$. Τέλος, επιχειρήσαμε να συνδυάσουμε τις δυο τελευταίες μεθοδολογίες μέσα από την εισαγωγή ενός αλγορίθμου με σκοπό την εύρεση του εσωτερικού δείγματος για δοσμένο επίπεδο l και τιμή y της μεταβλητής Y. Για τη συγκεκριμένη μέθοδο πήραμε πως η ολική πολυπλοκότητα είναι κατά μέσο όρο τάξης $\epsilon^{-2}\left(\log\epsilon\right)^2$, δηλαδή η μικρότερη δυνατή και κατ΄ αυτή την έννοια η τελική αυτή προσέγγιση είναι η βέλτιστη για την επίλυση του προβλήματός μας.

Βιβλιογραφικές Αναφορές

[1] M. B. Giles. (2015). Multilevel Monte Carlo methods, Acta Numerica, 24 , 259-328

[2] N. Collier, A. Haji-Ali, F.Nobile, E. von Schwerin, R. Tempone. (2015). A Coninuation Multilevel Monte Carlo algorithm, BIT Numerical Mathematics, 55 , 399-432

[3] M. Broadie, Y. Du, C. Moallemi. (2011). Efficient Risk Estimation via Nested Sequential Simulation, Management Science, 57 , 1172-1194

[4] M. B. Gordy, S. Juneja. (2010). Nested simulation in portfolio risk management, Management Science, 56 , 1833-1848

[5] D. L. Burkholder, R. F. Gundy. (1970). Extrapolation and interpolation of quasi-linear operators on martingales, Acta Math, 124, 249-304

[6] M. B. Giles, A. Haji-Ali. (2019). Multilevel nested simulation for efficient risk estimation, SIAM/ASA Journal on Uncertainty Quantification, 7 , 497-525

[7] M. B. Giles. (2018). Multilevel Monte Carlo methods, LMS / CRISM Summer School in Computational Stochastics, University of Warwick