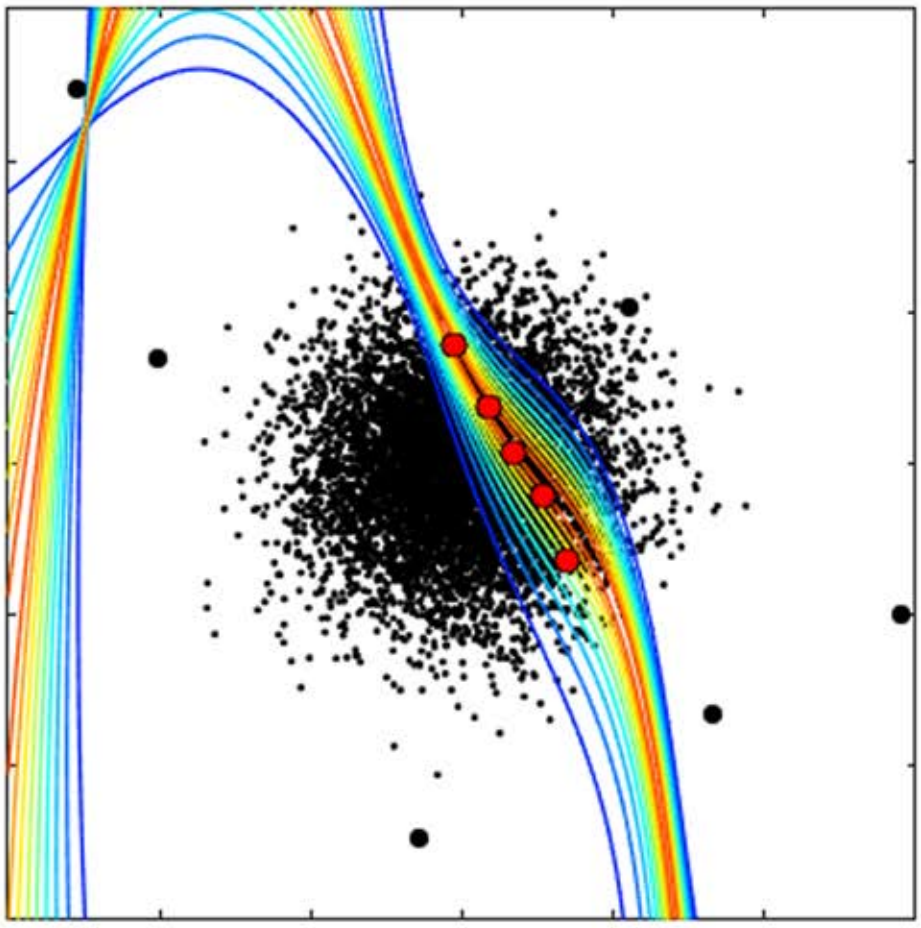
NATIONAL TECHNICAL UNIVERSITY OF ATHENS

School of Civil Engineering

POST GRADUATE THESIS

DEVELOPMENT OF KRIGING SURROGATE MODELS FOR THE RELIABILITY ANALYSIS OF STRUCTURES

Ανάπτυξη Αδρομερών Μοντέλων με την Μέθοδο Kriging για την Ανάλυση Αξιοπιστίας των Κατασκευών



Mantelos Michalis

Supervisor: Fragiadakis Michalis

Athens, 2017

Contents

Abstract 3

Abstract in Greek 5

1 Theoretical Framework 9

1.1 Random Variables 9

1.2 Common probability distributions of continuous variables 9

1.2.1 Normal distribution 9

1.2.2 Extreme type I distribution (or Gumbel distribution) 10

1.2.3 Lognormal distribution 10

1.3 Random Vectors 11

1.4 Stochastic Processes 11

1.4.1 Definition 11

1.4.2 Correlation functions 12

1.5 Kriging 17

1.5.1 Least-squares linear regression model 17

1.5.2 The best linear unbiased predictor – Kriging 19

1.5.3 Estimation of parameters $\boldsymbol{\vartheta}$ 21

1.5.4 Computational difficulties 22

1.5.5 One-dimensional example of the kriging predictor 22

1.5.6 Probabilistic classification function $\boldsymbol{\pi}(\mathbf{x})$ 26

2 Reliability analysis 29

2.1 Introduction 29

2.2 Problem definition 29

2.3 Monte Carlo Simulation 32

2.4 Importance Sampling 34

2.5 Subset Simulation 36

2.5.1 Principle 36

2.5.2 Algorithm 38

2.5.3 Example 41

2.5.4 Comparison of the Subset Simulation and the Monte Carlo Simulation 42

2.6 Meta-Model Based Importance Sampling 43

2.6.1 Introduction 43

2.6.2 Implementation 44

2.6.3 Example 45

3 Adaptive Refinement 49
3.1 Introduction 49
3.2 Refinement Criteria 51
3.2.1 The margin indicator function 51
3.2.2 The margin probability function 51
3.2.3 The function of most probable misclassification 52
3.3 Sampling-based Adaptive Refinement 53
3.3.1 Principle 53
3.3.2 Proposed Refinement Method 54
3.3.3 Stopping criterion of the refinement procedure 55
3.4 Complete Algorithm of Kriging Based Reliability Analysis 57
3.5 Refinement Example 58
4 Applications 63
4.1 One-dimensional problem 63
4.2 2-dimensional, piecewise performance function 64
4.3 High-dimensional problem 67
4.4 Truss Bridge 72
4.5 Tall building under seismic loads 74
5 Conclusion 77
Appendix A – Weighted K-means Clustering 78
Appendix B – Latin Hypercube Sampling 79
Appendix C – Markov Chain Monte Carlo 80
Modified Metropolis-Hasting Algorithm 80
Burn-in and thinning 81
Numerical estimation of coefficient $\boldsymbol{\gamma}$ 81
References 82

Abstract

Reliability analysis is a scientific field, that deals with problems with uncertainties in a probabilistic way. In structural engineering such uncertainties are common, and originate from the properties of the materials, the geometry of a structure, the construction, the mathematical modeling, and the loading. The most common approach is to distinguish those uncertainties to favorable and unfavorable, and apply reductive and augmentative coefficients respectively. Such an approach though, may be impossible in non-linear problems, where the superposition principle does not hold, and also it leads to very conservative designs. The probabilistic approach on the other hand, only requires the definition of those uncertainties with probability density functions, and provides a measurement of probable scenarios, such as the failure and safety, and ultimately the reliability of a structure.

Solving a reliability problem, is achieved with various methods. The analytical ones are straightforward but restricted to a point that they are rendered impractical. Other gradient-based method were developed through the years, but they are also restricted, and provide approximations that often are not acceptable. The most prevailing techniques, are numerical approximations called simulation experiments, or computer experiments. The most common and robust is the known Monte Carlo Simulation (MCS), which produces numerous results with purpose to classify them in the correspondent domain (e.g. failure or safety) and make conclusions on the frequency of a phenomenon. The MCS and its variations, such as the Subset Simulation (SS), are the most preferred methods to deal with a reliability analysis problem.

This thesis presents a way to deal with a great disadvantage of the simulation techniques. As mentioned earlier, these techniques are based on an iterative algorithm, and therefore the function that indicates the classification of a probable state, also called the performance function $g(\mathbf{x})$, has to be evaluated many times. This means that the computational cost is high, and the procedure time-consuming, because a performance function may rely on solving large finite elements models, or a time-history analysis. The main idea to treat this inefficiency, is to replace the performance function with a new one, that provides accurate classification and is computationally easy to evaluate. Such a function is referred to as a surrogate model and is reliant on very few information of the performance function. This thesis focuses on the application of the kriging predictor as a surrogate model in structural reliability.

In the first chapter, the necessary theoretical framework is presented. This includes the definition of a random variable and vector, and the most commonly used distributions that define a random variable, the definition of a stochastic process and the kriging predictor. The kriging predictor is a procedure that creates a Gaussian stochastic process G based on a specific dataset or observations, which derive from the performance function. Making use of correlation functions that express the confidence of the prediction in the vicinity of the observations, the kriging predictor estimates the performance function, providing a mean value $\mu_G(\mathbf{x})$ and variance $\sigma_G(\mathbf{x})^2$ at any point of the input space. In reliability analysis only the classification of a state is important. This classification is expressed from the probabilistic classification function $\pi(\mathbf{x})$, which depends on the kriging mean and variance, and provides the probability, a state to reside in the safety or the failure domain.

In the second chapter, the most efficient simulation techniques to estimate the failure probability, are presented. The application of those techniques differ slightly when performed

on the kriging model. More precisely, the probabilistic classification function $\pi(\mathbf{x})$ replaces the indicator function $\mathrm{I}_{\mathrm{F}}(\mathbf{x})$ which relies on evaluating the performance function. The two approaches, meaning the classic $\mathrm{g}(\mathbf{x})$-based and the Gaussian process $\mathrm{G}\sim\mathcal{N}(\mu_{\mathrm{G}}(\mathbf{x}),\sigma_{\mathrm{G}}(\mathbf{x})^2)$ based are shown in parallel, to highlight to similarities and the differences. Lastly, a new simulation technique is introduced, that combines the kriging model based simulation techniques with Importance Sampling. The hybrid method Meta-model based Importance Sampling, focuses on approximating the optimal instrumental PDF from the kriging model, and sampling efficiently with evaluations of the $\mathrm{g}(\mathbf{x})$ function. This technique can validate the kriging model estimation, or corrected it if the model is inaccurate.

The third chapter focuses on the strategies that should be implemented, in order to enrich the dataset of the kriging model and improve its prediction. As mentioned above, the kriging model is not accurate in areas away from the observations. However, this is not important as long as the classification is accurate. Consequently, the model is enriched with new points at areas where the classification is not confident, with priority to those that affect more the failure probability. This is achieved, by solving a clustering problem of a large amount of candidate points, and choosing one of each cluster. This refinement procedure is repeated until the kriging model is sufficient.

In the fourth chapter, validation examples and applications are presented. The kriging-based reliability method is tested on problems with irregular limit state function, problems with many dimensions and lastly on two structural reliability problems. The Meta-model Importance Sampling is also implemented and illustrated at each application.

In the last chapter, the final conclusions are drawn. The kriging model approach, is proven to be very reliable in reliability analysis, especially in problems with less than 100 dimensions, a continuous performance function and low probabilities. Its application can be extended in optimization, and big data regression.

Abstract in Greek

Η ανάλυση αξιοπιστίας είναι ένα επιστημονικό πεδίο, που σκοπό έχει να αντιμετωπίσει προβλήματα με αβεβαιότητες, με τρόπο πιθανοτικό. Στην αξιοπιστία των κατασκευών τέτοιες αβεβαιότητες είναι συνήθεις, και πηγάζουν από τις ιδιότητες των υλικών, τη γεωμετρία του φορέα, την κατασκευή, τη μαθηματική προσομοίωση, τα επιβαλλόμενα φορτία. Η πιο κοινή αντιμετώπιση τέτοιων προβλημάτων είναι ο διαχωρισμός των αβέβαιων μεγεθών σε ευμενή και δυσμενή, και η χρήση μειωτικών και αυξητικών συντελεστών αντίστοιχα. Αυτή η προσέγγιση, ωστόσο, είναι αδύνατη σε μη-γραμμικά προβλήματα, όπου η αρχή της επαλληλίας δεν ισχύει, και ενδεχομένως οδηγεί σε υπερσυντηρητικό σχεδιασμό. Από την άλλη, η πιθανοτική αντιμετώπιση, απαιτεί μονάχα τον προσδιορισμών των αβεβαιοτήτων, με συναρτήσεις συχνότητας πιθανότητας (ΣΣΠ), προκειμένου να παρέχει την μέτρηση ενός πιθανού σεναρίου, όπως αυτά της αστοχίας και της ασφάλειας, και εν τέλει της δομικής αξιοπιστίας.

Η επίλυση ενός προβλήματος αξιοπιστίας, επιτυγχάνεται με διάφορους τρόπους. Από αυτούς, οι αναλυτικοί τρόποι είναι ευθείς, όμως εξαιρετικά περιορισμένοι και ανεφάρμοστοι σε πολύπλοκα προβλήματα. Άλλες μέθοδοι, όπως αυτές που βασίζονται στη γνώση της παραγώγου, είναι αποτελεσματικές υπό προϋποθέσεις, και μπορούν να αντιμετωπίσουν προσεγγιστικά, μόνο κάποια είδη προβλημάτων. Οι επικρατέστερες μέθοδοι, είναι αριθμητικές, και προσομοιώνουν πειράματα τύχης σε υπολογιστή. Η πιο γνωστή και εύρωστη μέθοδος είναι η προσομοίωση Monte Carlo (MCS), που παρέχει πολυπληθή αποτελέσματα από προσομοιώσεις, τα οποία ταξινομούνται στο αντίστοιχο πεδίο (π.χ. της ασφάλειας ή της αστοχίας), ώστε να ληφθεί συμπέρασμα για τη συχνότητα εμφάνισης ενός φαινομένου. Η προσομοίωση Monte Carlo και οι παραλλαγές της, όπως η προσομοίωση με τη μέθοδο των υποσυνόλων (Subset Simulation ή SS), είναι οι πιο προτιμητέες για την αντιμετώπιση προβλημάτων αξιοπιστίας.

Η εργασία αυτή παρουσιάζει ένα τρόπο αντιμετώπισης του μεγαλύτερου μειονεκτήματος των μεθόδων προσομοίωσης. Όπως προαναφέρθηκε, αυτές οι μέθοδοι βασίζονται σε επαναληπτικές διαδικασίες. Συνεπώς η συνάρτηση αστοχίας $g(\mathbf{x})$, που καθορίζει την ταξινόμηση μίας πιθανής κατάστασης, υπολογίζεται διαδοχικά πολλές φορές. Αυτό σημαίνει, ότι το υπολογιστικό κόστος είναι υψηλό, και η διαδικασία χρονοβόρα, διότι η συνάρτηση αστοχίας μπορεί να περιλαμβάνει πολύπλοκά μοντέλα πεπερασμένων στοιχείων, ανάλυση χρονοϊστορίας κ.α.. Η κύρια ιδέα για την αντιμετώπιση αυτής της δυσκολίας, είναι η αντικατάσταση της συνάρτησης αστοχίας, με μία νέα, που μπορεί να παρέχει αξιόπιστη ταξινόμηση και εύκολα υπολογίσιμη. Αυτή η συνάρτηση ονομάζεται αδρομερές μοντέλο και είναι βασιζόμενο σε πολύ περιορισμένη γνώση της πραγματικής συνάρτησης αστοχίας. Η εργασία, λοιπόν, επικεντρώνεται στη γραμμική πρόβλεψη kriging ως αδρομερές μοντέλο στη δομική αξιοπιστία.

Στο πρώτο κεφάλαιο παρουσιάζονται όλα τα βασικά θεωρητικά στοιχεία στα οποία βασίζεται η ανάλυση αξιοπιστίας σε μοντέλο kriging. Αρχικά ορίζονται οι τυχαίες μεταβλητές και διανύσματα, καθώς οι βασικότερες κατανομές, όπως η κανονική, η λογαριθμοκανονική και η ακροτάτων (Gumbel), με βάση τις οποίες προσομοιώνονται οι αβεβαιότητες των μεγεθών. Στη συνέχεια ορίζεται η στοχαστική ανέλιξη, που αποτελεί τη βάση της πρόβλεψης kriging. Στα πλαίσια αυτής της εργασίας οι στοχαστικές ανελίξεις περιορίζονται στις Γκαουσιανές. Επιπλέον, γίνεται η απλοποίηση, ότι το μητρώο συσχετίσεως, προκύπτει από συναρτήσεις συσχέτισης, οι οποίες συναρτάνε τη συσχέτιση με την απόσταση στο χώρο των εισαγόμενων

μεταβλητών, και ρυθμίζονται από τους συντελεστές συσχέτισης. Η πρόβλεψη kriging, ή αλλιώς η βέλτιστη αμερόληπτη (unbiased) γραμμική πρόβλεψη, στοχεύει στον προσδιορισμό μίας στοχαστικής ανέλιξης, βασιζόμενη σε παρατηρήσεις, δηλαδή ένα πεπερασμένο σύνολο σημείων του πεδίου ορισμού της συνάρτησης αστοχίας, με τις αντίστοιχες τιμές της. Η εύρεση των συντελεστών συσχέτισης γίνεται επιλύοντας ένα υπολογιστικά δύσκολο πρόβλημα βελτιστοποίησης. Με αυτόν τον τρόπο, η πρόβλεψη kriging παρέχει το μοντέλο $\mathrm{G}\sim\mathcal{N}(\mu_{\mathrm{G}}(\mathbf{x}),\sigma_{\mathrm{G}}(\mathbf{x})^2)$, δηλαδή για το τυχόν σημείου του πεδίου ορισμού, δίνεται η μέση τιμή $\mu_{\mathrm{G}}(\mathbf{x})$ και η τυπική απόκλιση $\sigma_{\mathrm{G}}(\mathbf{x})$, που περιγράφουν την εκτίμηση της συνάρτησης αστοχίας, μαζί με την αντίστοιχη αβεβαιότητα. Ιδιότητα αυτού του μοντέλου είναι ότι στα υποστηρικτικά σημεία (παρατηρήσεις), η μέση τιμή $\mu_{\mathrm{G}}(\mathbf{x})$ ταυτίζεται με την $\mathrm{g}(\mathbf{x})$, ενώ η τυπική απόκλιση $\sigma_{\mathrm{G}}(\mathbf{x})$ είναι μηδενική. Συνεπώς το μοντέλο kriging παρέχει στις περιοχές κοντά στα υποστηρικτικά σημεία πιο βέβαιη πρόβλεψη απ' ότι πιο μακριά.

Στα προβλήματα ανάλυσης αξιοπιστίας, απαιτείται η σωστή ταξινόμηση κάθε σημείου του πεδίου ορισμού. Συνεπώς, το μοντέλο kriging θεωρείται επαρκές όχι όταν παρέχει με βεβαιότητα κάθε τιμή της $\mathrm{g}(\mathbf{x})$, αλλά όταν παρέχει σωστή ταξινόμηση. Για τον υπολογισμό της πιθανότητας αστοχίας, στην αξιοπιστία κατασκευών, ορίζεται η συνάρτηση πιθανοτικής ταξινόμησης $\pi(\mathbf{x})$ που εκφράζει την πιθανότητα μία κατάσταση ή αλλιώς ένα σημείο του χώρου να βρίσκεται στην περιοχή της αστοχίας. Λαμβάνει τιμές από το μηδέν έως το ένα και υπολογίζεται από τις προαναφερθέντες ποσότητες $\mu_{\mathrm{G}}(\mathbf{x}),\sigma_{\mathrm{G}}(\mathbf{x})$.

Στο δεύτερο κεφάλαιο παρουσιάζονται αναλυτικά οι μέθοδοι προσομοίωσης για την αριθμητική προσέγγιση της πιθανότητας αστοχίας. Η προσομοίωση Monte Carlo και η προσομοίωση με τη μέθοδο των υποσυνόλων, εφαρμόζονται είτε με βάση την πραγματική συνάρτηση αστοχίας είτε με βάση το μοντέλο kriging $\mathrm{G}\sim\mathcal{N}(\mu_{\mathrm{G}}(\mathbf{x}),\sigma_{\mathrm{G}}(\mathbf{x})^2)$. Οι δύο αυτές προσεγγίσεις παρουσιάζονται παράλληλα, ώστε να φαίνεται η αντιστοιχία. Η κύρια διαφορά είναι ότι η συνάρτηση πιθανοτικής ταξινόμησης $\pi(\mathbf{x})$ αντικαθιστά το δείκτη $\mathrm{I}_{\mathrm{F}}(\mathbf{x})$, όταν η προσομοίωση βασίζεται στο μοντέλο kriging.

Στη συνέχεια, γίνεται λόγος για τη χρήσιμη ιδιότητα της Importance Sampling (IS). Η IS επιτρέπει την επιλογή διαφορετικής συνάρτησης συχνότητας πιθανότητας (instrumental PDF), για την παραγωγή του τυχαίου δείγματος στην εκτίμηση της πιθανότητας αστοχίας. Με τη σωστή επιλογή εναλλακτικής ΣΣΠ, η επαναληπτική διαδικασία της MCS μπορεί να επιταχυνθεί σημαντικά. Επίσης αποδεικνύεται ποια είναι η βέλτιστη ΣΣΠ, δηλαδή αυτή που μηδενίζει τη διακύμανση της εκτιμώμενης πιθανότητας. Ωστόσο, η βέλτιστη ΣΣΠ απαιτεί τη γνώση της πιθανότητας και συνεπώς είναι μη-εφαρμόσιμη. Σε αυτό το σημείο παρουσιάζεται η νέα υβριδική μέθοδος Meta-model based Importance Sampling (MIS). Σκοπός της είναι η προσέγγιση της βέλτιστης ΣΣΠ μέσω του μοντέλου kriging, και η παραγωγή περιορισμένου τυχαίου δείγματος για την εκτίμηση της πιθανότητας. Με άλλα λόγια, αν το μοντέλο kriging μπορεί να παρέχει μία μέτρια έως καλή εκτίμηση της πιθανότητας αστοχίας, η MIS διορθώνει σε δεύτερη φάση την πιθανότητα αστοχίας με μικρό δείγμα που απαιτεί υπολογισμούς της συνάρτησης αστοχίας. Η μέθοδος MIS συνδυάζει το μοντέλο kriging με την Importance Sampling και είναι εξαιρετικά χρήσιμη όταν απαιτείται επιβεβαίωση της εκτίμησης ή όταν το μοντέλο είναι ανεπαρκές.

Το τρίτο κεφάλαιο περιλαμβάνει τις στρατηγικές με τις οποίες η πρόβλεψη kriging μπορεί να βελτιωθεί, ώστε να παρέχει πιο βέβαιη ταξινόμηση. Όπως προαναφέρθηκε, η κατασκευή του μοντέλου βασίζεται μονάχα σε ορισμένες παρατηρήσεις ή υποστηρικτικά σημεία, τα οποία λαμβάνονται τυχαία από το πεδίο ορισμού της συνάρτησης αστοχίας $\mathrm{g}(\mathbf{x})$. Η βελτίωση του

μοντέλου επιτυγχάνεται λαμβάνοντας νέες παρατηρήσεις και επανυπολογίζοντάς το. Τα νέα υποστηρικτικά σημεία επιλέγονται, με βάση μία συνάρτηση-κριτήριο, που λαμβάνει μεγάλες τιμές όταν η μέση τιμή $\mu_G(\mathbf{x})$ πλησιάζει το μηδέν, και όταν η τυπική απόκλιση $\sigma_G(\mathbf{x})$ αυξάνεται. Με αυτόν τον τρόπο στοχεύονται οι περιοχές στις οποίες γίνεται η μετάβαση από την αστοχία στην ασφάλεια και αντίστροφα. Είναι οι περιοχές, δηλαδή, στις οποίες η εκτίμηση της ταξινόμησης είναι πιο δύσκολη. Στην παρούσα εργασία, δύο είναι τα κριτήρια στα οποία δίνεται μεγαλύτερη έμφαση: Η συνάρτηση πιθανοτικού εύρους (margin probability function $\mathrm{MP}(\mathbf{x})$) και η προτεινόμενη συνάρτηση πιθανέστερης εσφαλμένης ταξινόμησης (most probable misclassification function $\mathrm{MF}(\mathbf{x})$). Στη συνέχεια, τα άνω κριτήρια που κατευθύνουν στη μεταβατική περιοχή της αστοχίας και της ασφάλειας, πολλαπλασιάζονται με τη ΣΣΠ $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$ του προβλήματος αξιοπιστίας. Με αυτό τον τρόπο, το τελικό κριτήριο $\mathrm{f}_{\mathbf{X}}(\mathbf{x}) \cdot \mathrm{MF}(\mathbf{x})$ λαμβάνει επιπρόσθετα υπόψιν ότι περιοχές με μεγάλες τιμές της ΣΣΠ είναι πιο καθοριστικές στον υπολογισμό της πιθανότητας αστοχίας.

Σε αυτό το σημείο, το κριτήριο για την επιλογή νέων παρατηρήσεων στην κατασκευή του μοντέλου, χρησιμοποιείται ως ΣΣΠ για την παραγωγή δείγματος. Το δείγμα αυτό είναι ένα σύνολο υποψήφιων σημείων. Ωστόσο το μέγεθός του είναι μεγάλο και είναι μη-αποδοτικό να επιλεγεί ολόκληρό. Προκειμένου να επιλεγούν τα πιο αντιπροσωπευτικά σημεία, λύνεται το πρόβλημα συσταδοποίησης (clustering problem) με τον αλγόριθμο k-means. Έτσι το σύνολο των παρατηρήσεων ανανεώνεται και το μοντέλο kriging είναι πιο ακριβές σε κρίσιμες περιοχές.

Αντί της παραπάνω προσέγγισης, δηλαδή της παραγωγής δείγματος με βάση την συνάρτηση πυκνότητας πιθανότητας $\propto \mathrm{f}_{\mathbf{X}}(\mathbf{x}) \cdot \mathrm{MF}(\mathbf{x})$, προτείνεται, επιπρόσθετα, μία εναλλακτική λύση. Δεδομένου ότι η διαδικασία ανανέωσης των παρατηρήσεων, είναι μέρος ενός επαναληπτικού αλγορίθμου, και ότι η πιθανότητα αστοχίας παρατηρείται σε κάθε βήμα, τότε υπάρχει διαθέσιμο δείγμα που έχει παραχθεί από την προσομοίωση MCS ή SS. Αυτό το δείγμα χρησιμοποιείται για να λυθεί το πρόβλημα συσταδοποίησης, όμως κάθε σημείο έχει διαφορετική βαρύτητα, που καθορίζεται από τη συνάρτηση $\mathrm{w}(\mathbf{x}) = \mathrm{MF}(\mathbf{x})$. Με αυτόν τον τρόπο, αποφεύγεται η παραγωγή δείγματος από την $\propto \mathrm{f}_{\mathbf{X}}(\mathbf{x}) \cdot \mathrm{MF}(\mathbf{x})$, που θα απαιτούσε δειγματοληψία από μαρκοβιανές αλυσίδες και γίνεται περεταίρω εκμετάλλευση του υπάρχοντος δείγματος από την προσομοίωση MCS ή SS.

Όπως προαναφέρθηκε, η διαδικασία ανανέωσης των παρατηρήσεων γίνεται πολλές φορές διαδοχικά με λίγα σημεία, ώστε το μοντέλο να μπορεί να δίνει σε κάθε βήμα νέα κατεύθυνση. Το προτεινόμενο κριτήριο για τη λήξη της διαδικασίας βασίζεται στην ποσότητα p_{MF} που ισούται με την ολοκλήρωση της συνάρτησης $\mathrm{f}_{\mathbf{X}}(\mathbf{x}) \cdot \mathrm{MF}(\mathbf{x})$ σε όλο το πεδίο ορισμού. Αυτό επιτυγχάνεται με τον ίδιο τρόπο που υπολογίζεται η πιθανότητα αστοχίας, και με το ίδιο δείγμα. Αν, ιδανικά, το μοντέλο αποκτήσει τέλεια ακρίβεια στην ταξινόμηση, η συνάρτηση $\mathrm{MF}(\mathbf{x})$ αποκτά μηδενική τιμή, και συνεπώς μηδενίζεται το p_{MF}.

Στο τέταρτο κεφάλαιο, γίνεται εφαρμογή της μεθόδου αξιοπιστίας που βασίζεται σε μοντέλο kriging. Αυτή περιλαμβάνει την αρχική κατασκευή του μοντέλου από τυχαία επιλεγμένες παρατηρήσεις, την ανανέωση με νέα αποδοτικά υποστηρικτικά σημεία, την τελική εκτίμηση της πιθανότητας αστοχίας, και την επαλήθευση με την υβριδική μέθοδο Meta-model based Importance Sampling. Τα προβλήματα στα οποία εφαρμόζεται είναι προβλήματα με ανώμαλη (αλλά συνεχή) συνάρτηση αστοχίας, πολλών διαστάσεων, και τα τελευταία δύο αφορούν περιπτώσεις δομικής αξιοπιστίας. Το πρώτο είναι ένα δικτύωμα μεγάλου ανοίγματος, με τυχαίες μεταβλητές τα φορτία, τη γεωμετρία και τις ιδιότητες των υλικών, στο οποίο μετράται η πιθανότητα υπέρβασης του επιτρεπόμενου βέλους. Στο δεύτερο

πρόβλημα, ένα κτίριο ύψους $80\,\mathrm{m}$ προσομοιώνεται με την παραδοχή συνεχούς καμπτοδιατμητικού προβόλου, με τυχαίες συνεχείς μεταβλητές τη δυσκαμψία και την ανηγμένη καθ' ύψος μάζα. Επίσης, το φάσμα των επιταχύνσεων ακολουθεί λογαριθμοκανονική κατανομή, όπως προκύπτει από στατιστική επεξεργασία των Boore-Atkinson. Ως συνάρτηση αστοχίας ορίζεται η υπέρβαση του drift κατά 0.75%.

Το τελευταίο κεφάλαιο περιλαμβάνει τα τελικά συμπεράσματα. Η προσέγγιση των προβλημάτων αξιοπιστίας των κατασκευών με τη γραμμική πρόβλεψη kriging αποδεικνύεται αποδοτική, ειδικά σε προβλήματα κάτω των 100 διαστάσεων, με συνεχή συνάρτηση αστοχίας και μικρές πιθανότητες. Το μοντέλο kriging μπορεί να επεκταθεί σε προβλήματα βελτιστοποίησης και σε παλινδρόμηση μεγάλων δεδομένων (big data).

1 Theoretical Framework

1.1 Random Variables

The function $\mathbf{X}: \boldsymbol{\Omega} \to \mathbb{R}$**,** defined in probability space $(\Omega, \mathcal{F}, \mathcal{P})$, is a real-valued random variable if $\{\omega: \mathrm{X}(\omega) \leq \mathrm{r}\} \in \mathcal{F},\ \forall \mathrm{r} \in \mathbb{R}$, where Ω is the sample space containing all possible outcomes, $\mathcal{F}$ the set of events containing possible outcomes and $\mathcal{P}$ the probability measure.

A continuous real-valued random variable can be entirely defined by its cumulative distribution function:

$$\mathrm{F_X}(\mathrm{x}) = \mathrm{P}(\mathrm{X} \leq \mathrm{x}), \mathrm{x} \in \mathbb{X} \subseteq \mathbb{R} \tag{1. 1}$$

Alternatively, it can also be defined by its probability distribution function:

$$\mathrm{f_X}(\mathrm{x}) = \frac{\mathrm{dF_X}(\mathrm{x})}{\mathrm{dx}} \tag{1. 2}$$

Most commonly the following values are considered to describe a random variable.

The expected value, or mean value:

$$\mu_\mathrm{X} = \mathbb{E}(\mathrm{X}) = \int_{\mathbb{X}} \mathrm{x} \cdot \mathrm{f_X}(\mathrm{x})\,\mathrm{dx} \tag{1. 3}$$

and the standard deviation, which is the square root of the variance of X:

$$\sigma_\mathrm{X} = \sqrt{\int_{\mathbb{X}} (\mathrm{x} - \mu_\mathrm{X})^2 \cdot \mathrm{f_X}(\mathrm{x})\,\mathrm{dx}} \tag{1. 4}$$

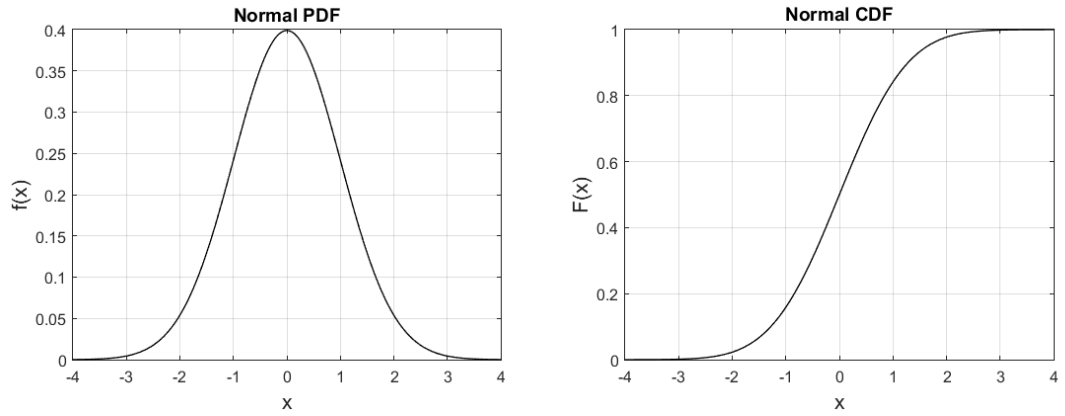
1.2 Common probability distributions of continuous variables

1.2.1 Normal distribution

The normal distribution is the most common distribution used to describe random variables, in natural sciences.

$$\mathrm{f_X}(\mathrm{x}) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1. 5}$$

Below, the normal probability density function (PDF) and the normal cumulative distribution function are plotted. In this case μ=0 and σ=1, the distribution is called standard. Also, such a variable is called standard normal variable and noted commonly with Z.



PDF and CDF of random variable $\mathcal{N}(0,1)$

1.2.2 Extreme type I distribution (or Gumbel distribution)

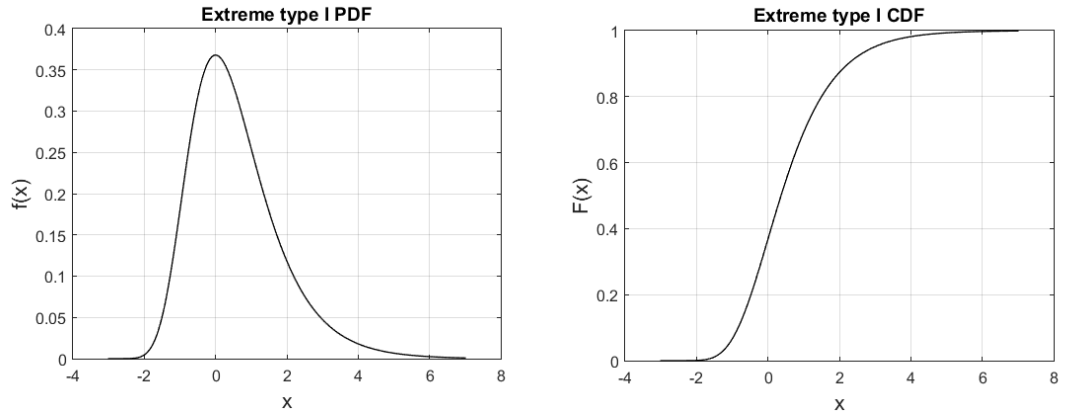
The extreme type I is used more often to describe random variables that are attributed to loads on a structure, such a wind load or a snow load.

$$f_X(x) = e^{-e^{-\frac{x-\xi}{\beta}}} \tag{1. 6}$$

with:

$$\mu = \xi + \gamma\beta,\ \gamma \approx 0.5772, \sigma = \frac{\beta\pi}{\sqrt{6}}$$

Considering $\xi = 0$, $\beta = 1$ the PDF and CDF are plotted below:



PDF and CDF of random variable Gum(0,1)

1.2.3 Lognormal distribution

A log-normally distributed variable X is defined as:

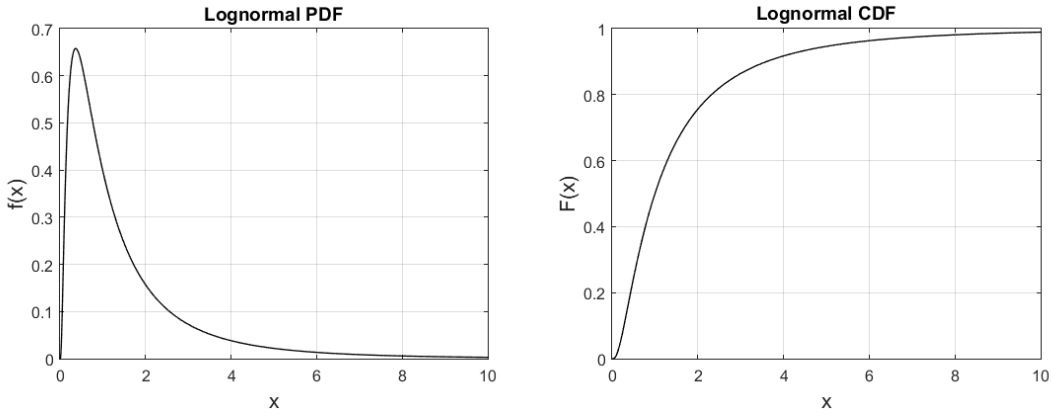
$$X = e^{\mu+\sigma Z} \tag{1. 7}$$

where:

μ, σ: parameters of the distribution

Z: standard normal variable

Considering $\mu = 0, \sigma = 1$ the PDF and CDF are plotted below:



PDF and CDF of random variable lnN(0,1)

1.3 Random Vectors

A random vector **X** is a vector, whose m components are random variables. The vector $\boldsymbol{\mu}_{\mathbf{X}}$ consists of the expected value of each component.

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}(\mathrm{X}_i), i = 1,2,\dots,m \tag{1. 8}$$

The covariance matrix **C** is defined as follows:

$$\mathrm{C}_{\mathrm{i,j}} = \mathrm{Cov}\left[\mathrm{X}_\mathrm{i}, \mathrm{X}_\mathrm{j}\right], \qquad \mathrm{i,j} = 1,2,\dots,\mathrm{m} \tag{1. 9}$$

where $\mathrm{Cov}\left[\mathrm{X}_\mathrm{i}, \mathrm{X}_\mathrm{j}\right] = \mathbb{E}\left[\left(\mathrm{X}_\mathrm{i} - \mu_\mathrm{i}\right) \cdot \left(\mathrm{X}_\mathrm{j} - \mu_\mathrm{j}\right)\right]$ is the covariance of any two random variables $\mathrm{X}_\mathrm{i}, \mathrm{X}_\mathrm{j}$, and the correlation matrix **R** is defined as:

$$\mathrm{R}_{\mathrm{i,j}} = \frac{\mathrm{C}_{\mathrm{i,j}}}{\sigma_{\mathrm{X}_\mathrm{i}} \cdot \sigma_{\mathrm{X}_\mathrm{j}}}, \qquad \mathrm{i,j} = 1,2,\dots\mathrm{m} \tag{1. 10}$$

The joint probability density distribution can be defined from the cumulative distribution function:

$$\mathrm{F}_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}(\mathrm{X}_1 \le \mathrm{x}_1, \mathrm{X}_2 \le \mathrm{x}_2, \dots, \mathrm{X}_\mathrm{m} \le \mathrm{x}_\mathrm{m}) \tag{1. 11}$$

In the special case that the components of the random vector are independent, meaning:

$$\mathrm{R}_{\mathrm{i,j}} = \begin{Bmatrix} 1 & \mathrm{i} = \mathrm{j} \\ 0 & \mathrm{i} \neq \mathrm{j} \end{Bmatrix}, \qquad \mathrm{i,j} = 1,2,\dots\mathrm{m} \tag{1. 12}$$

the following expression is true:

$$\mathrm{f}_{\mathbf{X}}(\mathbf{x}) = \mathrm{f}_{\mathrm{X}_1}(\mathrm{x}_1) \cdot \mathrm{f}_{\mathrm{X}_2}(\mathrm{x}_2) \cdot \dots \cdot \mathrm{f}_{\mathrm{X}_\mathrm{m}}(\mathrm{x}_\mathrm{m}) \tag{1. 13}$$

1.4 Stochastic Processes

1.4.1 Definition

A stochastic process **Y** defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ is a collection of infinite or finite random variables, indexed by the input space **X**:

$$\mathrm{Y}(\mathbf{x}), \mathbf{x} \in \mathbb{X}$$

Historically, $\mathbf{x}$ represents time, but in the context of this thesis the stochastic processes are indexed by m-dimensional continuous vector $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^m$ and are limited to Gaussian processes. Therefore, the mean function $\mu(\mathbf{x})$, and the autocovariance function $\mathrm{C}(\mathbf{x}, \mathbf{x}')$ need to be defined:

$$\mu(\mathbf{x}) \equiv \mathbb{E}[\mathrm{Y}(\mathbf{x})], \qquad \mathbf{x} \in \mathbb{X} \tag{1. 14}$$

$$\mathrm{C}(\mathbf{x}, \mathbf{x}') = \mathbb{E}\big[\big(\mathrm{Y}(\mathbf{x}) - \mu(\mathbf{x})\big) \cdot \big(\mathrm{Y}(\mathbf{x}') - \mu(\mathbf{x}')\big)\big], \qquad (\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times \mathbb{X} \tag{1. 15}$$

An $\mathrm{n} \times \mathrm{n}$ covariance matrix $\mathbf{C}$ must be symmetric:

$$\mathbf{C}(\mathrm{i}, \mathrm{j}) = \mathbf{C}(\mathrm{j}, \mathrm{i}), \qquad \mathrm{i}, \mathrm{j} = 1,2, \dots \mathrm{n} \tag{1. 16}$$

and positive definite:

$$\mathrm{z}^{\mathrm{T}} \cdot \mathrm{C} \cdot \mathrm{z} \geq 0, \qquad \forall \mathrm{z} = (\mathrm{z}_1, \mathrm{z}_2, \dots, \mathrm{z}_{\mathrm{n}})^{\mathrm{T}} \in \mathbb{R}^{\mathrm{n}} \tag{1. 17}$$

1.4.2 Correlation functions

Correlation functions are used to define the autocorrelation in Gaussian processes. Those functions are arbitrarily chosen, and are dependent on certain parameters. In structural reliability problems, which this thesis focuses on, it is reasonable that the correlation should be a function of the "distance" in the input space:

$$\mathrm{R} = \mathrm{R}(\mathbf{x} - \mathbf{x}'), \qquad (\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times \mathbb{X} \tag{1. 18}$$

The quantity $\mathbf{d} = \mathbf{x} - \mathbf{x}'$ is known as lag distance in kriging applications. In addition, the Gaussian process is assumed to be componentwise anisotropic, therefore:

$$\mathrm{R}(\mathbf{d}) = \prod_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{R}(\mathrm{d}_\mathrm{i}), \qquad \mathrm{d}_\mathrm{i} = \mathrm{x}_\mathrm{i} - \mathrm{x'}_\mathrm{i}, \mathrm{i} = 1,2, \dots \mathrm{m} \tag{1. 19}$$

Considering anisotropy is very crucial, since specific components may differ considerably in the sensitivity of the physical problem.

Below, the most popular correlation functions are presented. Those are dependent on certain parameters, that are usually more or equal to the number of the components m of the input vector $\mathbf{x}$. The common characteristics of those functions are:

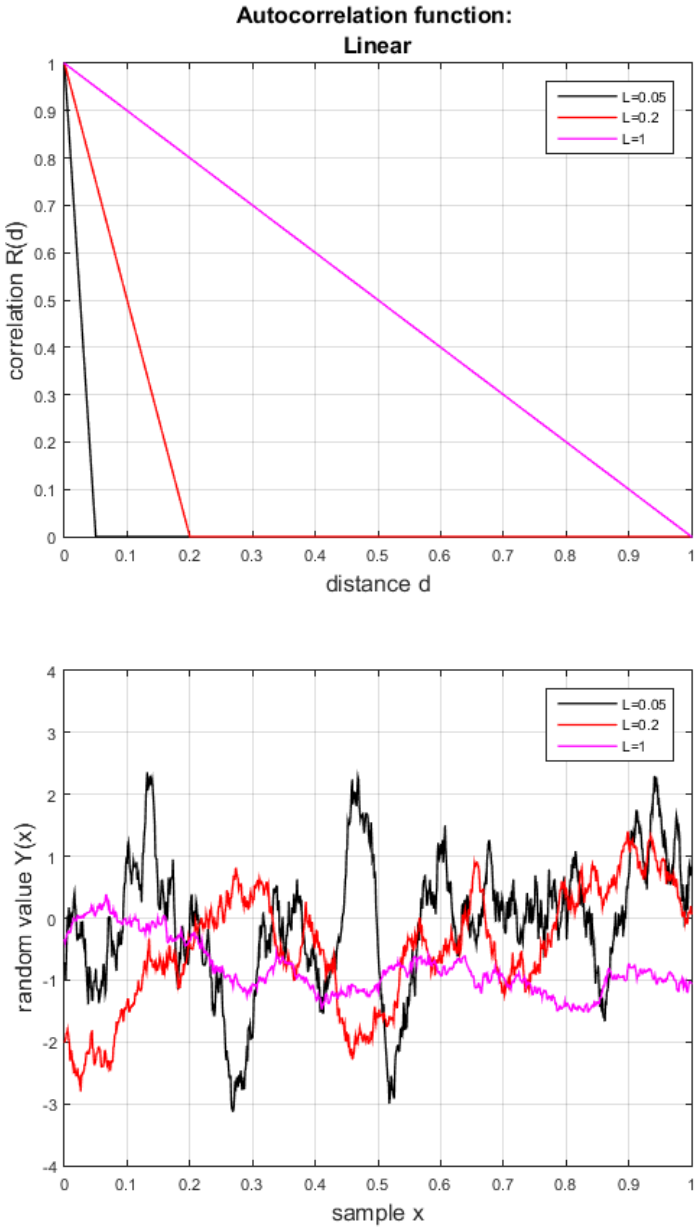
- $\mathrm{R}(0) = 1$
- R is continuous
- In all cases R is descending, and reaches zero, at a specific "range" L. Therefore, L is defined as the minimum lag distance d that $\mathrm{R}(\mathrm{d}) = 0$. In the case of correlation functions that converge to zero in $+\infty$, L is often defined such that $\mathrm{R}(\mathrm{L}) = 0.05$. The quantity L shows the distance at which any two values of the process are entirely uncorrelated. Although L can better show the physical meaning of the Gaussian process correlation, the parameters $\vartheta_\mathrm{i}, \mathrm{i} = 1,2, \dots, \mathrm{m}$ are used instead, because they are more suitably defined in an optimization procedure described in a following subchapter.

1.4.2.1 Linear autocorrelation function

$$R(\mathbf{d}) = \prod_{i=1}^{m} \max(1 - \vartheta_i \cdot |d_i|; 0), \qquad i = 1,2, \dots m \tag{1. 20}$$

$$L_i = \frac{1}{\vartheta_i}, \qquad i = 1,2, \dots m \tag{1. 21}$$

Below, an illustration of 3 random processes, with 1-dimensional input $x = [0,1]$, that differ in correlation length $L = \{0.05, 0.2, 1\}$, is shown:



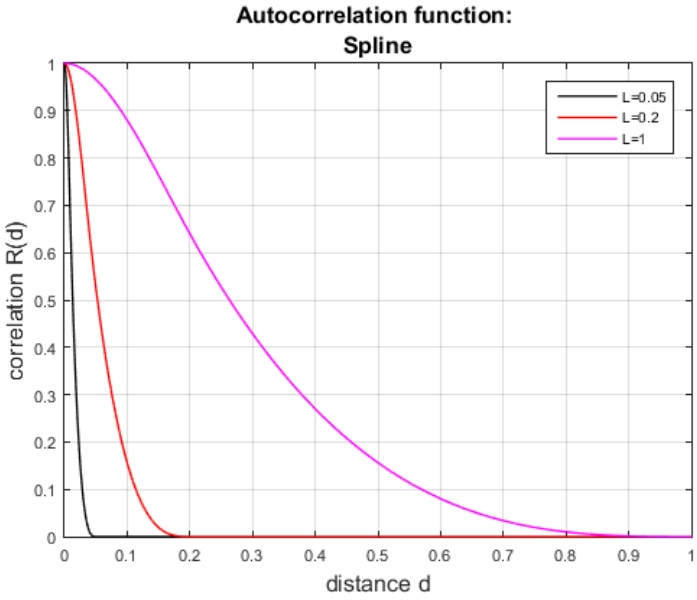
It should be noted that the resulting random process $Y(x)$ is non-differentiable.

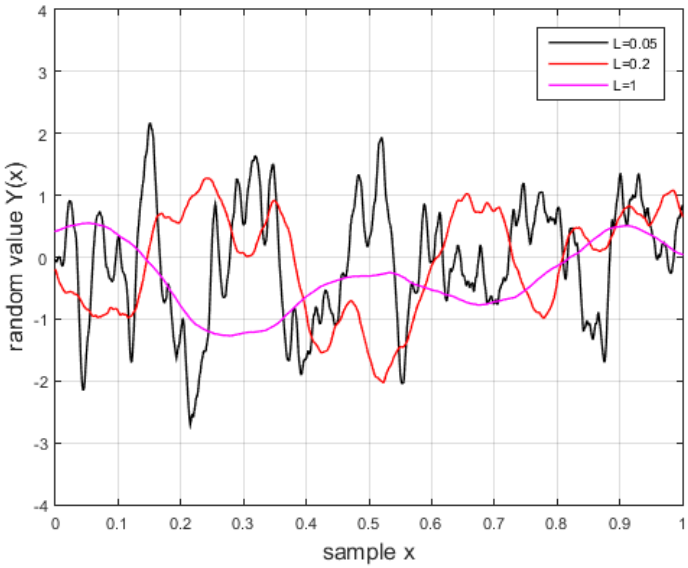
1.4.2.2 Spline autocorrelation function

$$R(\mathbf{d}) = \prod_{i=1}^{m} \begin{cases} 1 - 15(\vartheta_i|d_i|)^2 + 30(\vartheta_i|d_i|)^3 & \text{for} \quad \vartheta_i|d_i| \le 0.2 \\ 1.25(1 - \vartheta_i|d_i|)^3 & \text{for} \; 0.2 < \vartheta_i|d_i| \le 1 \\ 0 & \text{for} \quad 1 < \vartheta_i|d_i| \end{cases}, \qquad i = 1,2, \dots m \tag{1. 22}$$

$$L_i = \frac{1}{\vartheta_i}, \qquad i = 1,2, \dots m \tag{1. 23}$$

Illustration for $L = \{0.05, 0.2, 1\}$:



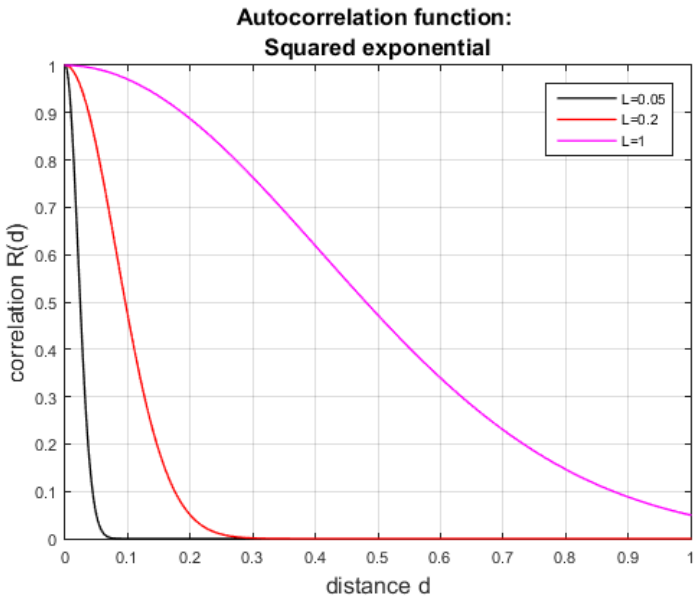


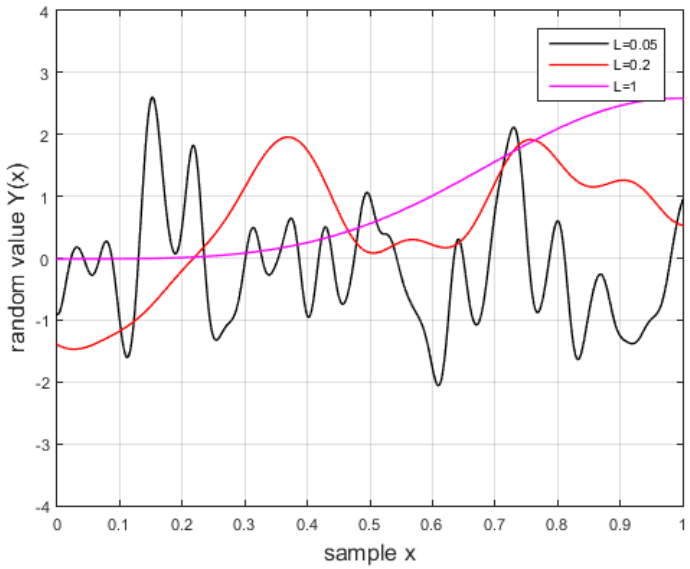
1.4.2.3 Squared Exponential autocorrelation function

$$R(\mathbf{d}) = \prod_{i=1}^{m} \exp\left(-\vartheta_i {d_i}^2\right), \qquad i = 1,2,\dots m \tag{1. 24}$$

$$L_i \approx \sqrt{\frac{3}{\vartheta_i}}, \qquad i = 1,2,\dots m \tag{1. 25}$$

Illustration for $L = \{0.05, 0.2, 1\}$:





The resulting stochastic processes, in this case, are infinitely-differentiable.

1.4.2.4 Matérn autocorrelation function

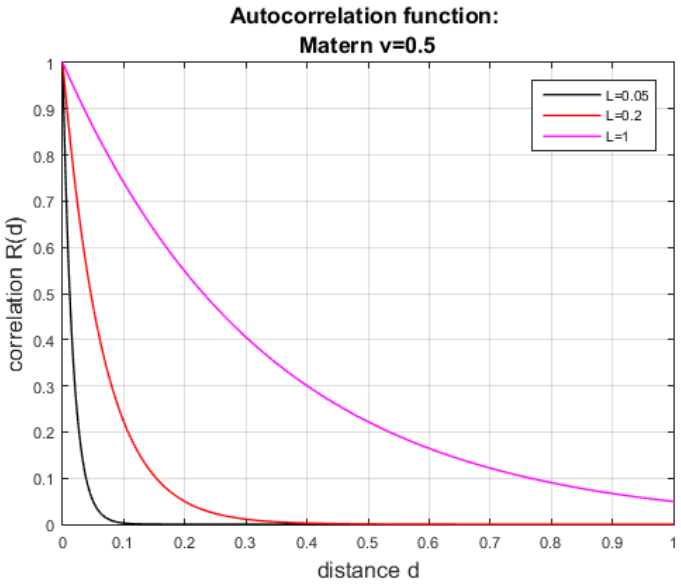
$$R(\mathbf{d}) = \prod_{i=1}^{m} \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(2\sqrt{\nu} \cdot \vartheta_i |d_i|\right)^{\nu} \cdot K_{\nu}\left(2\sqrt{\nu} \cdot \vartheta_i |d_i|\right), \qquad \nu \geq 0.5, i = 1,2, \dots m \tag{1. 26}$$

$\Gamma(\cdot)$: the Gamma function, $K_{\nu}(\cdot)$: the Bessel function of the second kind

$$L_i \approx \frac{2}{\vartheta_i}, \qquad i = 1,2, \dots m \tag{1. 27}$$

The Matérn autocorrelation function contains an extra parameter ν, which affects the differentiability of the Gaussian process. The resulting processes are $\nu - 1$ times differentiable. Controlling the differentiability is a very important advantage, which will be discussed further in a following chapter.

Illustration for $L = \{0.05, 0.2, 1\}$ and $\boldsymbol{\nu = 0.5}$:



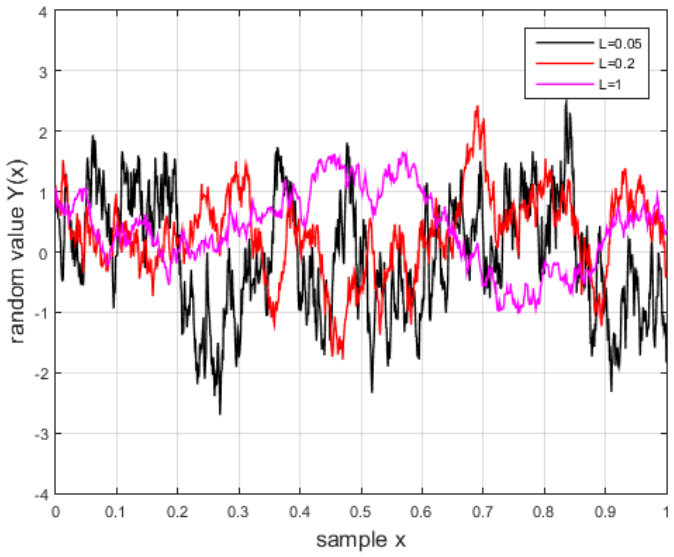
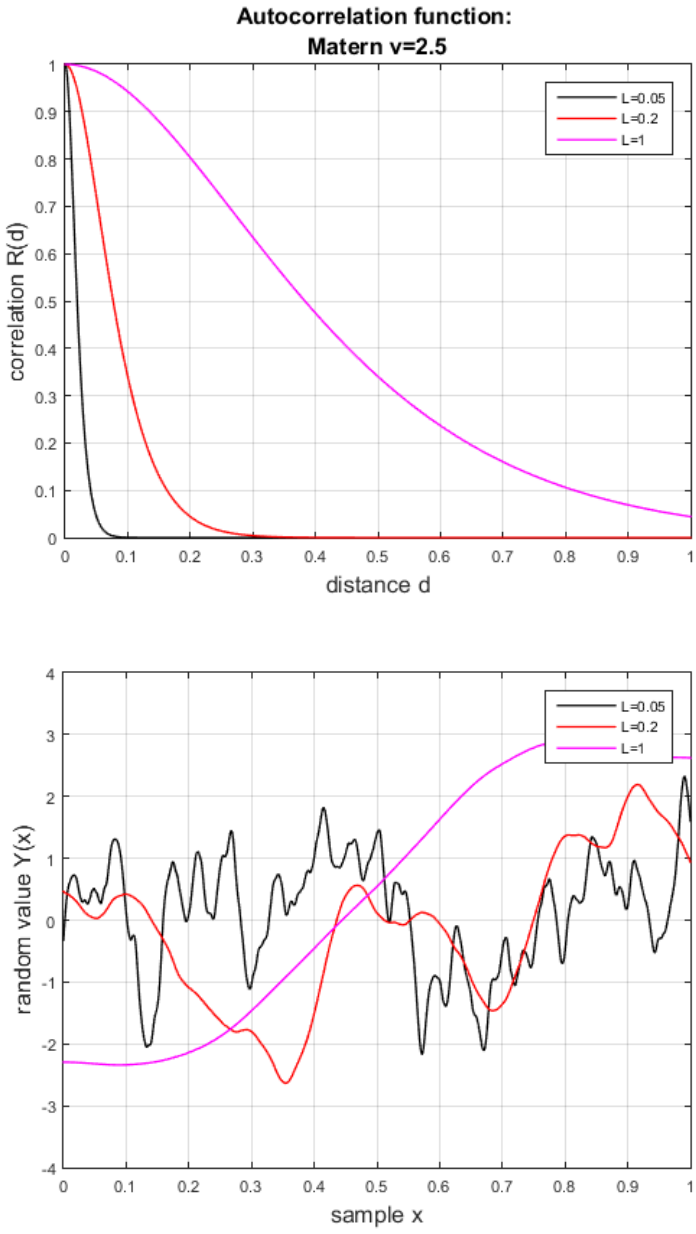


Illustration for $\mathrm{L} = \{0.05, 0.2, 1\}$ and $\boldsymbol{\nu = 2.5}$:



1.5 Kriging

1.5.1 Least-squares linear regression model

The least-square linear regression is the basis of the kriging predictor. It is defined as:

$$\mathrm{Y_i} = \mathbf{f}^{\mathrm{T}}\left(\mathbf{x}^{(\mathbf{i})}\right) \cdot \boldsymbol{\beta} + \mathrm{Z_i}, \qquad \mathrm{i} = 1,2, \ldots \mathrm{n} \tag{1. 28}$$

with:

Y: a vector with n observations in sites $\mathbf{x}^{(\mathbf{i})}$

$\mathbf{f}(\mathbf{x})$: a regression function

$\boldsymbol{\beta}$: the weights of the regression to be calculated

Z_i: the error of the regression measured at site $\mathbf{x}^{(i)}$

A regression function $\mathbf{f}(\mathbf{x})$ is usually a polynomial of order 0, 1 or 2:

- Constant

$$\mathbf{f}(\mathbf{x}) = 1 \tag{1. 29}$$

with size $p = 1$

- Linear

$$\mathbf{f}(\mathbf{x}) = [1\ x_1\ x_2 \dots x_m]^T \tag{1. 30}$$

with size $p = m + 1$

- Quadratic

$$\mathbf{f}(\mathbf{x}) = [1\ \ x_1\ \ x_2 \dots x_m\ \ x_1^2\ \ x_1x_2 \dots x_1x_m \dots x_m^2]^T \tag{1. 31}$$

with size $p = \frac{1}{2}(m + 1) \cdot (m + 2)$

The number of observations n must always be more than p. Otherwise, the problem is under-determined. Among the above polynomial regressions, the linear is generally preferred. The quadratic regression, despite being more accurate, depends on many observations and therefore is not efficient in problems with many dimensions. For example, a 50-dimensional problem requires at least 1326 observations.

The vector $\mathbf{Z}$ is a Gaussian vector with $\mathbb{E}[\mathbf{Z}] = 0$ and $\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}] = \mathbb{E}[\mathbf{Z}\mathbf{Z}^T] = \sigma^2\mathbf{R}$.

Considering the correlation matrix R known, the model is evaluated by solving the optimization problem described below, through maximizing the likelihood of the weight vector $\boldsymbol{\beta}$ and the variance σ^2, with $\mathbf{y}$ being the realization of the observations $\mathbf{Y}$:

$$L(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\left((2\pi\sigma^2)^n \det(\mathbf{R})\right)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{F}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\boldsymbol{\beta})\right] \tag{1. 32}$$

where:

$\mathbf{F} = \left[\mathbf{f}^T(\mathbf{x}^{(1)})\ \ \mathbf{f}^T(\mathbf{x}^{(2)}) \dots \mathbf{f}^T(\mathbf{x}^{(n)})\right]^T$, a matrix of size $n \times p$, with the regression functions evaluated at the observation sites $\mathbf{x}^{(i)}$.

The maximum likelihood problem can be rewritten, using the natural logarithm. Equivalently:

$$(\boldsymbol{\beta}, \sigma^2) = \arg min_{(\boldsymbol{\beta}, \sigma^2)}\left[-\log\left(L(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)\right)\right] \tag{1. 33}$$

Solving the minimization problem:

$$\left.\begin{aligned} \nabla_{\boldsymbol{\beta}} \log L(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) &= 0 \\ \frac{\partial \log L(\mathbf{y}|\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= 0 \end{aligned}\right\} \Rightarrow \begin{aligned} \boldsymbol{\beta} &= \left(\mathbf{F}^T\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{R}^{-1}\mathbf{y} \qquad (1.\ 34) \\ \sigma^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{F}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\boldsymbol{\beta}) \qquad (1.\ 35) \end{aligned}$$

1.5.2 The best linear unbiased predictor – Kriging

A two-stage Gaussian process is assumed, consisting of the regression part and a Gaussian process $\mathrm{Z}(\mathbf{x})$.

$$\mathrm{Y_i} = \sum_{\mathrm{j=1}}^{\mathrm{p}} \beta_\mathrm{j} \mathrm{f_j}\left(\mathbf{x}^{(\mathrm{i})}\right) + \mathrm{Z}\left(\mathbf{x}^{(\mathrm{i})}\right), \qquad \mathrm{i} = 0,1,2,\dots \mathrm{n} \tag{1. 36}$$

where, $\mathrm{Y_0}$ is the unobserved value to be estimated. Also the Gaussian process $\mathrm{Z}(\mathbf{x})$ is defined with zero mean:

$$\mathbb{E}[\mathrm{Z}(\mathbf{x})] = 0, \qquad \forall \boldsymbol{x} \in \mathbb{X} \tag{1. 37}$$

and stationary autocovariance:

$$\mathrm{Cov}[\mathrm{Z}(\mathbf{x}), \mathrm{Z}(\mathbf{x}')] = \sigma^2 \mathrm{R}(\mathbf{x}, \mathbf{x}'), \qquad \forall (\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times \mathbb{X} \tag{1. 38}$$

Subsequently, the following Gaussian process is defined:

$$\begin{Bmatrix} \mathrm{Y_0} \\ \mathbf{Y} \end{Bmatrix} = \mathcal{N}\left(\begin{Bmatrix} \mathbf{f}^{\mathrm{T}}\left(\mathbf{x}^{(\mathbf{0})}\right)\boldsymbol{\beta} \\ \mathbf{F}\boldsymbol{\beta} \end{Bmatrix}, \sigma^2 \begin{bmatrix} 1 & \mathbf{r_0^{\mathrm{T}}} \\ \mathbf{r_0} & \mathbf{R} \end{bmatrix} \right) \tag{1. 39}$$

where:

$$\mathrm{r_{0,i}} = \mathrm{R}\left(\mathbf{x}^{(\mathbf{0})}, \mathbf{x}^{(\mathbf{i})}\right), \qquad \mathrm{i} = 1,2,\dots \mathrm{n} \tag{1. 40}$$

which is the correlation vector that contains the correlation between the prediction $\widehat{\mathrm{Y}}_0$ and the observed $\mathrm{Y_i}$.

The matrix $\mathbf{R}$ and vector $\mathbf{r}$ are unknown, but assuming a correlation function as described in paragraph 1.4.2, they are defined as:

$$\mathrm{r_{0,i}} = \mathrm{R}\left(\mathbf{x}^{(\mathbf{0})} - \mathbf{x}^{(\mathbf{i})}, \boldsymbol{\vartheta}\right), \qquad \mathrm{R_{i,j}} = \mathrm{R}\left(\mathbf{x}^{(\mathbf{i})} - \mathbf{x}^{(\mathbf{j})}, \boldsymbol{\vartheta}\right), \qquad \mathrm{i,j} = 1,2,\dots \mathrm{n} \tag{1. 41}$$

where $\boldsymbol{\vartheta}$ is the vector with the parameters of the correlation function, so far undetermined.

Theorem: The best linear unbiased predictor

The best linear unbiased predictor of the unobserved quantity of interest $\mathrm{y_0} = \mathrm{y}\left(\mathbf{x}^{(\mathbf{0})}\right)$ is the Gaussian random variate $\widehat{\mathrm{Y}}_0$ with mean:

$$\mu_{\widehat{\mathrm{Y}}_0} = \mathbb{E}\left[\widehat{\mathrm{Y}}_0\right] = \mathbb{E}[\mathrm{Y_0}] = \mathbf{f_0}^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \mathbf{r_0^{\mathrm{T}}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{F}\widehat{\boldsymbol{\beta}}\right) \tag{1. 42}$$

and variance:

$$\sigma^2_{\widehat{\mathrm{Y}}_0} = \mathbb{E}\left[\left(\widehat{\mathrm{Y}}_0 - \mathrm{Y_0}\right)^2\right] = \sigma^2\left(1 - \mathbf{r_0^{\mathrm{T}}}\mathbf{R}^{-1}\mathbf{r_0} + \mathbf{u_0}^{\mathrm{T}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{u_0}\right) \tag{1. 43}$$

where:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}$$

$$\mathbf{u_0} = \mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r_0} - \mathbf{f_0} \tag{1. 44}$$

Proof:

Firstly, the following properties are defined:

- Any prediction $\widehat{\mathrm{Y}}_0$ is a linear expression of the vector of observations $\mathbf{Y}$ through a vector of weights $\mathbf{a}$:

$$\widehat{\mathrm{Y}}_0 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a}_{0,\mathrm{i}} \mathrm{Y}_\mathrm{i} = \mathbf{a}_0{}^{\mathrm{T}}\mathbf{Y}, \qquad \mathbf{a}_0 \equiv \mathbf{a}\left(\mathbf{x}^{(\mathbf{0})}\right) \in \mathbb{R}^{\mathrm{n}} \tag{1. 45}$$

- The predictor is unbiased:

$$\mathbb{E}\left[\widehat{\mathrm{Y}}_0 - \mathrm{Y}_0\right] = 0 \tag{1. 46}$$

- The prediction $\widehat{\mathrm{Y}}_0$ minimizes the mean squared error:

$$\widehat{\mathrm{Y}}_0 = \arg\min_{\mathrm{Y}_0^*} \mathbb{E}[(\mathrm{Y}_0^* - \mathrm{Y}_0)^2] \tag{1. 47}$$

Therefore, the optimal weight vector $\mathbf{a}_0$ is found by solving the following minimization problem:

$$\mathbf{a}_0 = \arg min_{\mathbf{a}_0} \mathbb{E}\left[\left(\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Y} - \mathrm{Y}_0\right)^2\right] \tag{1. 48}$$

$$\text{with constraint: } \mathbb{E}[\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Y} - \mathrm{Y}_0] = 0 \tag{1. 49}$$

$\mathbf{Y}$ and Y_0 are replaced using the two-stage Gaussian process expression:

$$\widehat{\mathrm{Y}}_0 - \mathrm{Y}_0 = \mathbf{a}_0{}^{\mathrm{T}}\mathbf{Y} - \mathrm{Y}_0 = \mathbf{a}_0{}^{\mathrm{T}}(\mathbf{F}\boldsymbol{\beta} + \mathbf{Z}) - \left(\mathbf{f}_0{}^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{Z}_0\right) = \mathbf{a}_0{}^{\mathrm{T}}\mathbf{Z} - \mathrm{Z}_0 + \left(\mathbf{a}_0{}^{\mathrm{T}}\mathbf{F} - \mathbf{f}_0{}^{\mathrm{T}}\right)\boldsymbol{\beta} \Longrightarrow$$

$$\mathbb{E}\left[\widehat{\mathrm{Y}}_0 - \mathrm{Y}_0\right] = \mathbb{E}\left[\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Z} - \mathrm{Z}_0 + \left(\mathbf{a}_0{}^{\mathrm{T}}\mathbf{F} - \mathbf{f}_0{}^{\mathrm{T}}\right)\boldsymbol{\beta}\right] = \mathbb{E}[\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Z} - \mathrm{Z}_0] + \left(\mathbf{a}_0{}^{\mathrm{T}}\mathbf{F} - \mathbf{f}_0{}^{\mathrm{T}}\right)\boldsymbol{\beta}$$

Considering the zero mean of $\mathrm{Z}(\mathbf{x})$ and the unbiasedness property the constraint expression is simplified:

$$\left.\begin{array}{rl} \mathbb{E}[\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Z} - \mathrm{Z}_0] & = 0 \\ \mathbb{E}\left[\widehat{\mathrm{Y}}_0 - \mathrm{Y}_0\right] & = 0 \end{array}\right\} \Longrightarrow \left(\mathbf{a}_0{}^{\mathrm{T}}\mathbf{F} - \mathbf{f}_0{}^{\mathrm{T}}\right)\boldsymbol{\beta} = \mathbf{0} \Longrightarrow \mathbf{a}_0{}^{\mathrm{T}}\mathbf{F} - \mathbf{f}_0{}^{\mathrm{T}} = 0$$

The mean squared prediction error is rewritten us such:

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{\mathrm{Y}}_0 - \mathrm{Y}_0\right)^2\right] &= \mathbb{E}\left[\left(\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Z} - \mathrm{Z}_0\right)^2\right] = \mathbb{E}\left[\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathbf{T}}\mathbf{a}_0 + \mathrm{Z}_0{}^2 - 2\mathbf{a}_0{}^{\mathrm{T}}\mathbf{Z}\mathrm{Z}_0\right] \\ &= \mathbf{a}_0{}^{\mathrm{T}}\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathbf{T}}]\mathbf{a}_0 + \mathbb{E}\left[\mathrm{Z}_0{}^2\right] - 2\mathbf{a}_0{}^{\mathrm{T}}\mathbb{E}[\mathbf{Z}\mathrm{Z}_0] = \mathbf{a}_0{}^{\mathrm{T}}\sigma^2\mathbf{R}\mathbf{a}_0 + \sigma^2 - 2\mathbf{a}_0{}^{\mathrm{T}}\sigma^2\mathbf{r}_0 \\ &\Longrightarrow \mathbb{E}\left[\left(\widehat{\mathrm{Y}}_0 - \mathrm{Y}_0\right)^2\right] = \sigma^2\left(1 + \mathbf{a}_0{}^{\mathrm{T}}(\mathbf{R}\mathbf{a}_0 - 2\mathbf{r}_0)\right) \end{aligned}$$

where, by definition:

$$\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathbf{T}}] = \sigma^2\mathbf{R}, \mathbb{E}\left[\mathrm{Z}_0{}^2\right] = \sigma^2, \mathbb{E}[\mathbf{Z}\mathrm{Z}_0] = \sigma^2\mathbf{r}_0$$

Introducing Lagrange multipliers $\boldsymbol{\lambda_0} \equiv \boldsymbol{\lambda}\left(\mathbf{x}^{(\mathbf{0})}\right)$ the minimization problem with the equality constraint is converted to the following unconstrained problem:

$$(\mathbf{a}_0, \boldsymbol{\lambda_0}) = \arg min_{(\mathbf{a}_0, \boldsymbol{\lambda_0})} \mathrm{L}(\mathbf{a}_0, \boldsymbol{\lambda_0}) \tag{1. 50}$$

$$\mathrm{L}(\mathbf{a}_0, \boldsymbol{\lambda_0}) = \sigma^2\left(1 + \mathbf{a}_0{}^{\mathrm{T}}(\mathbf{R}\mathbf{a}_0 - 2\mathbf{r}_0)\right) + \boldsymbol{\lambda_0}{}^{\mathrm{T}}\left(\mathbf{a}_0{}^{\mathrm{T}}\mathbf{F} - \mathbf{f}_0{}^{\mathrm{T}}\right) \tag{1. 51}$$

Solving the optimization problem:

$$\left.\begin{matrix}\nabla_{\mathbf{a}_0} \mathrm{L} & = 0 \\ \nabla_{\boldsymbol{\lambda}_0} \mathrm{L} & = 0\end{matrix}\right\} \Longrightarrow \left.\begin{matrix}2\sigma^2(\mathbf{R}\mathbf{a}_0 - \mathbf{r}_0) + \mathbf{F}\boldsymbol{\lambda}_0 & = 0 \\ \mathbf{F}^{\mathrm{T}}\mathbf{a}_0 - \mathbf{f}_0 & = 0\end{matrix}\right\} \Longrightarrow$$

$$\begin{bmatrix}\mathbf{R} & \mathbf{F} \\ \mathbf{F}^{\mathbf{T}} & \mathbf{0}\end{bmatrix}\begin{Bmatrix}\mathbf{a}_0 \\ \tilde{\boldsymbol{\lambda}}_0\end{Bmatrix} = \begin{Bmatrix}\mathbf{r}_0 \\ \mathbf{f}_0\end{Bmatrix}, \text{ with } \tilde{\boldsymbol{\lambda}}_0 = \frac{\boldsymbol{\lambda}_0}{2\sigma^2} \tag{1. 52}$$

Therefore the weights vector is obtained:

$$\mathbf{a}_0 = \mathbf{R}^{-1}\left(\mathbf{r}_0 - \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0 - \mathbf{f}_0\right)\right) \tag{1. 53}$$

Lastly, the mean and variance of the unobserved $\widehat{\mathrm{Y}}_0$, $\mu_{\widehat{\mathrm{Y}}_0}$ and $\sigma^2_{\widehat{\mathrm{Y}}_0}$ are obtained by substitution:

$$\begin{aligned}\mu_{\widehat{\mathrm{Y}}_0} = \mathbf{a}_0{}^{\mathrm{T}}\mathbf{y} &= \left[\mathbf{r}_0 - \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0 - \mathbf{f}_0\right)\right]^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y} \\ &= \mathbf{r}_0{}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y} - \left[\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0 + \left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{f}_0\right]^{\mathbf{T}}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y} \\ &= \mathbf{f}_0{}^{\mathrm{T}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y} + \mathbf{r}_0{}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}\right) \\ &= \mathbf{f}_0{}^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \mathbf{r}_0^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{F}\widehat{\boldsymbol{\beta}}\right)\end{aligned}$$

$$\begin{aligned}\sigma^2_{\widehat{\mathrm{Y}}_0} &= \sigma^2\left(1 + \mathbf{a}_0{}^{\mathrm{T}}(\mathbf{R}\mathbf{a}_0 - 2\mathbf{r}_0)\right) \\ &= \sigma^2\left(1 + \left(\mathbf{r}_0 - \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\left(\mathbf{r}_0 - \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right) - 2\mathbf{r}_0\right)\right) \\ &= \sigma^2\left(1 - \left(\mathbf{r}_0 - \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{r}_0 + \mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right)\right) \\ &= \sigma^2\left(1 - \left(\mathbf{r}_0^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0 - \left(\mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right)^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right)\right) \\ &= \sigma^2\left(1 - \mathbf{r}_0^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0 + \mathbf{u}_0{}^{\mathrm{T}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right) \\ &= \sigma^2\left(1 - \mathbf{r}_0^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}_0 + \mathbf{u}_0{}^{\mathrm{T}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-\mathbf{1}}\mathbf{u}_0\right)\end{aligned}$$

1.5.3 Estimation of parameters $\boldsymbol{\vartheta}$

The estimation of the parameters $\boldsymbol{\vartheta}$, is based on the least-squares maximum likelihood estimator:

$$\mathrm{L}(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}) = \frac{1}{\left((2\pi\sigma^2)^{\mathrm{n}}\det\left(\mathbf{R}(\boldsymbol{\vartheta})\right)\right)^{1/2}}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{F}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\boldsymbol{\beta})\right]$$

The vector $\boldsymbol{\beta}$ and variance σ^2 depend on $\boldsymbol{\vartheta}$:

$$\boldsymbol{\beta}(\boldsymbol{\vartheta}) = \left(\mathbf{F}^{\mathrm{T}}\mathbf{R}(\boldsymbol{\vartheta})^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{R}(\boldsymbol{\vartheta})^{-1}\mathbf{y}$$

$$\sigma^2(\boldsymbol{\vartheta}) = \frac{1}{\mathrm{n}}\left(\mathbf{y} - \mathbf{F}\boldsymbol{\beta}(\boldsymbol{\vartheta})\right)^{\mathrm{T}}\mathbf{R}(\boldsymbol{\vartheta})^{-1}\left(\mathbf{y} - \mathbf{F}\boldsymbol{\beta}(\boldsymbol{\vartheta})\right)$$

By replacing $\boldsymbol{\beta}(\boldsymbol{\vartheta})$ and $\sigma^2(\boldsymbol{\vartheta})$ on the maximum likelihood estimator, and using the natural logarithm, the following expression is obtained:

$$-\log\left(\mathrm{L}(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\vartheta})\right)$$
$$= \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{F}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{R}(\boldsymbol{\vartheta})^{-1}(\mathbf{y}-\mathbf{F}\boldsymbol{\beta}) + \frac{\mathrm{n}}{2}\log(2\pi) + \frac{\mathrm{n}}{2}\log(\sigma^2)$$
$$+ \frac{1}{2}\log([\det\mathbf{R}(\boldsymbol{\vartheta})]) = \frac{\mathrm{n}}{2}\log\left(\sigma^2(\boldsymbol{\vartheta})[\det\mathbf{R}(\boldsymbol{\vartheta})]^{\frac{1}{\mathrm{n}}}\right) + \frac{\mathrm{n}}{2}(\log(2\pi)+1)$$

Let: $\psi(\boldsymbol{\vartheta}) = \sigma^2(\boldsymbol{\vartheta})[\det\mathbf{R}(\boldsymbol{\vartheta})]^{\frac{1}{\mathrm{n}}}$

Therefore, the parameters $\boldsymbol{\vartheta}$ are obtained by solving the following optimization problem:

$$\widehat{\boldsymbol{\vartheta}} = \arg min_{\boldsymbol{\vartheta}}\, \psi(\boldsymbol{\vartheta}) \tag{1. 54}$$

which eventually is equivalent to maximizing the maximum likelihood estimator.

The optimization problem described above cannot be solved analytically, and numerical methods need to be used. Such methods can be gradient-based algorithm, or a genetic algorithm.

1.5.4 Computational difficulties

It should be noted, that in many cases, numerical difficulties arise in the computation of the inverse matrix of $\mathbf{R}$ which is generally very ill-conditioned. The ill-conditioness is treated with various techniques, such as adding in the diagonal of $\mathbf{R}$ a very small value, historically called nugget. The nugget in Applied Geostatistics, a scientific field in which kriging was at first proposed, is a value that represents the measurement error, which produces noisy observations. In this thesis, simulations don't produce any noisy data, therefore the nugget value should be, in theory, zero. The effect of using a small nugget value, for clearly computational reasons, is that the mean kriging predictor becomes more "smooth" and does not interpolate the observations, but such an error is negligibly small and is ignored.

Additionally, the selection of the autocorrelation function is important. It has been noticed by Marrel (2005, 2008), that when the support points are very dense, which is a common case in following applications, the squared exponential function is not efficient enough. That is because it is infinitely differentiable. On the other hand, the use of a differentiable autocorrelation function is desirable, because it depicts more accurately the physical phenomenon that is emulated. Consequently, the Matérn autocorrelation function is very efficient, due to the extra parameter ν that affects differentiability, and performs well in models with support points non-uniformly distributed (Vazquez 2005).

1.5.5 One-dimensional example of the kriging predictor

The following function is emulated with the kriging predictor, choosing various arbitrary values for the $\boldsymbol{\vartheta}$ parameter of the correlation function, to demonstrate the way it affects the prediction. Finally, the optimal parameter $\boldsymbol{\vartheta}$, and the probabilistic classification function $\pi(\mathbf{x})$ are calculated. For the regression stage of the model, the constant regressor is chosen. The autocorrelation functions used are: (a) the squared exponential autocorrelation function and (b) the linear autocorrelation function. The support points, or observations, are also arbitrarily chosen, for the purpose of this example.

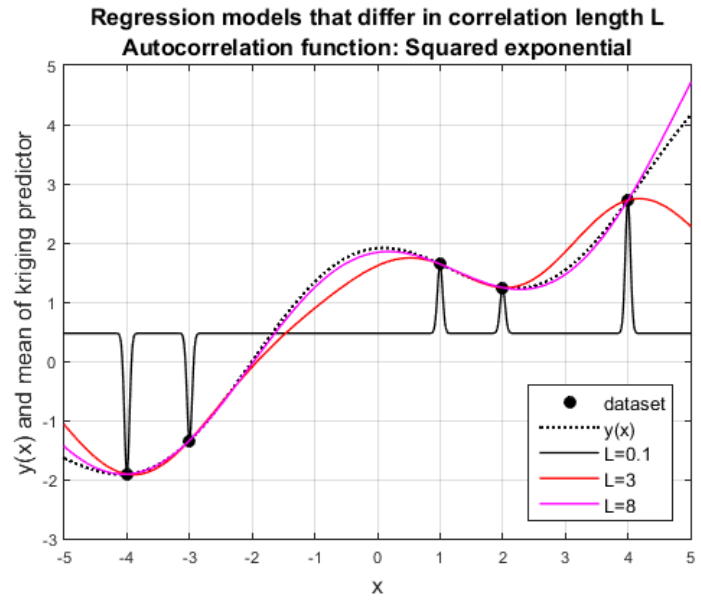
$$\mathrm{y}(\mathrm{x}) = \frac{1}{2}(\mathrm{x}+2) + \sin(\mathrm{x}+2), \qquad \mathrm{x} \in \mathbb{R}$$

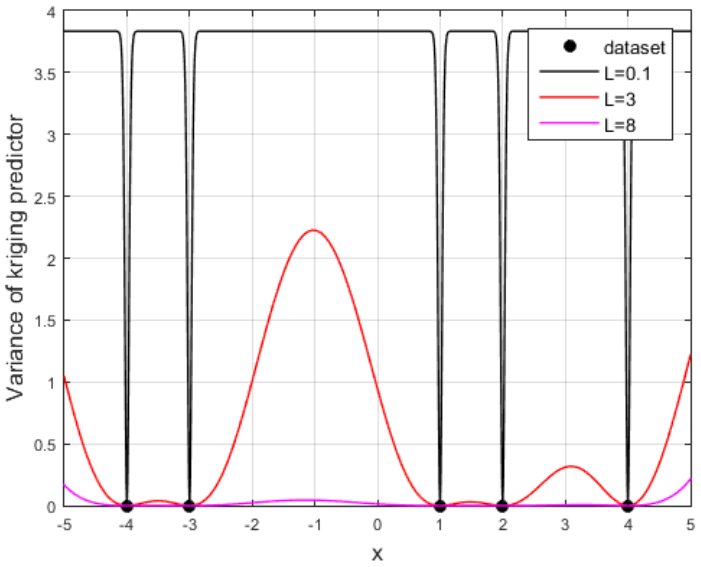
The evaluations of the function $y(x)$, which are used as support points for the kriging model are shown below:

i	x_i	y_i
1	-4	-1.909
2	-3	-1.341
3	1	1.641
4	2	1.243
5	4	2.721

(a) Squared exponential autocorrelation function (Gaussian)

The mean value $\mu_{\hat{Y}}(x)$ and variance $\sigma_{\hat{Y}}{}^2(x)$ of the kriging predictor $\hat{Y}$ are evaluated in the area of interest $x \in [-5,5]$, choosing 3 different correlation lengths. For $L = \{0.1\,;3\,;8\}$ ϑ is obtained: $\vartheta = \{300\,;1/3\,;0.0469\}$, since $\vartheta = 3/L^2$.





The dotted line in the figures above, is the function $\mathrm{y}(\mathrm{x})$ and is shown for reference. It should be noted that in all cases the mean kriging predictor interpolates the data, and the variance is zero at the same points. Also, it is clear that the correlation length $\mathrm{L} = 0.1$ is a very low value, since the kriging model gives hardly any useful information in the area of interest, apart from the areas very near to the observations. For the most part, the mean kriging predictor is equal to: $\mu_{\hat{Y}}(\mathrm{x}) \approx \mathrm{f}(\mathrm{x}) \cdot \beta = 1 \cdot 0.4708 = 0.4708$

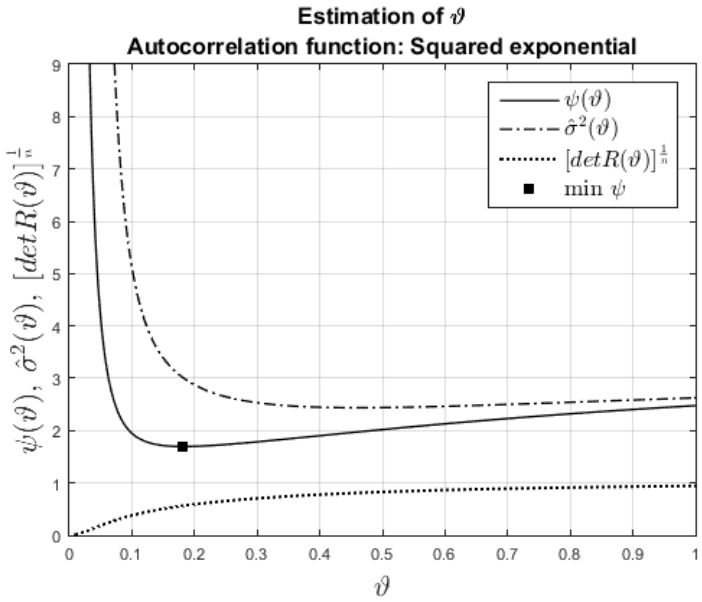
For this example, the weight vector $\mathbf{a}_0$ is also evaluated for $\mathrm{x}_0 = -1$:

	Observations				
	1	2	3	4	5
$\mathbf{a}_0(\mathrm{L} = 0.1)$	0.200	0.200	0.200	0.200	0.200
$\mathbf{a}_0(\mathrm{L} = 3)$	-0.170	0.574	0.642	-0.280	0.234
$\mathbf{a}_0(\mathrm{L} = 8)$	-0.435	0.865	1.492	-1.108	0.186

It is reminded, that: $\mu_{\hat{Y}}(\mathrm{x}_0) = \mathbf{a_0}^{\mathrm{T}}\mathbf{y}$, with observations $\mathbf{y}$ defined in previous table.

Finally, solving the optimization problem of paragraph **1.5.3**, the parameter ϑ is obtained. Due to the simplicity of this example, the objective function is also plotted.

It is reminded, that: $\psi(\boldsymbol{\vartheta}) = \sigma^2(\boldsymbol{\vartheta})[\det\mathbf{R}(\boldsymbol{\vartheta})]^{\frac{1}{\mathrm{n}}}$



The estimation of ϑ is:

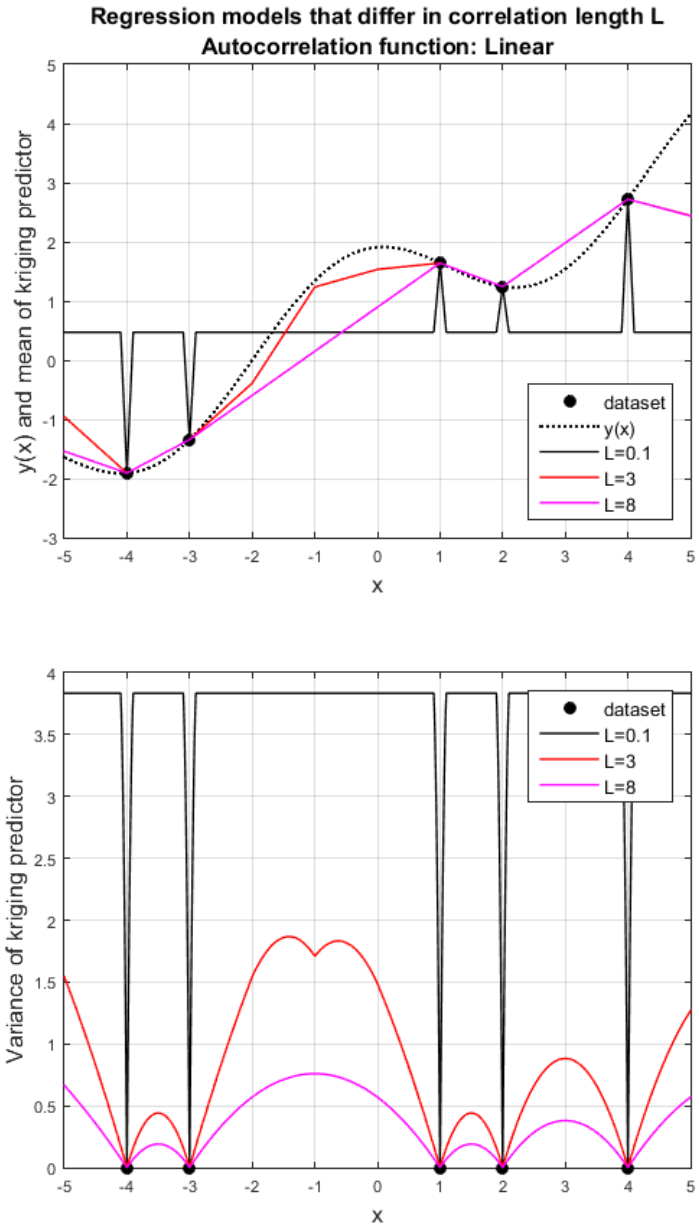
$$\vartheta = 0.182 \Longrightarrow \mathrm{L} = 4.06$$

In the figure above, it is clear that for $\vartheta \to 0^+$ the quantities $\sigma^2(\boldsymbol{\vartheta})$ and $[\det\mathbf{R}(\boldsymbol{\vartheta})]^{\frac{1}{\mathrm{n}}}$ approach $+\infty$ and zero respectively. Therefore, the objective function of the optimization problem is very sensitive to numerical errors in the calculation of these quantities.

(b) Linear autocorrelation function

The above procedure is repeated with the use of the linear autocorrelation function. Correlation lengths are the same:

$\mathrm{L} = \{0.1\,;3\,;8\} \Longrightarrow \vartheta = \{10\,;1/3\;;0.125\}$, since $\vartheta = 1/\mathrm{L}$ in this case.



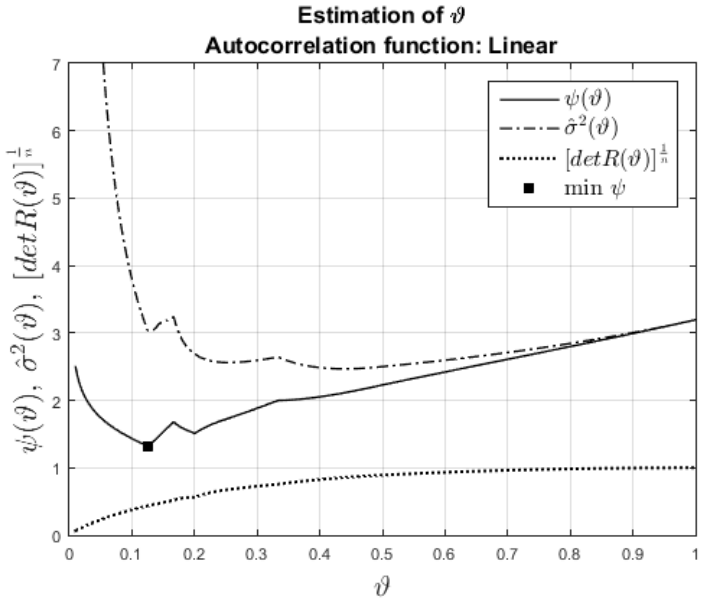
The main difference with (a), is that the mean kriging predictor $\mu_{\hat{Y}}(x)$ is linear in parts, for any $x \in \mathbb{R}$. Also the points, at which $\mu_{\hat{Y}}(x)$ is non-differentiable ("pointy"), differ by distance L from a support point. This is because the linear autocorrelation function reaches zero in a non-smooth way, compared to other autocorrelation functions, such as the squared-exponential, or the spline.

The weight vector $\mathbf{a}_0$, evaluated for $x_0 = -1$ is the following:

	Observations				
	1	2	3	4	5
$\mathbf{a}_0(\mathrm{L} = 0.1)$	0.200	0.200	0.200	0.200	0.200
$\mathbf{a}_0(\mathrm{L} = 3)$	-0.313	0.688	0.813	-0.500	0.313
$\mathbf{a}_0(\mathrm{L} = 8)$	0.000	0.500	0.500	0.000	0.000

In the case that $\mathrm{L} = 8$, the mean kriging predictor $\mu_{\hat{\mathrm{Y}}}(\mathrm{x})$ is actually a simple linear interpolation between the observations, therefore the weights of any unobserved value are dependent only on the distance of the nearest support point, of each side on the x axis. Thus, the weights $\mathbf{a}_0$ for $\mathrm{x}_0 = -1$, depend only on $|\mathrm{x}_0 - \mathrm{x}_2| = 2$ and $|\mathrm{x}_0 - \mathrm{x}_3| = 2$.

Solving the optimization problem, parameter ϑ is obtained:



$$\vartheta = 0.125 \Longrightarrow \mathrm{L} = 8$$

1.5.6 Probabilistic classification function $\boldsymbol{\pi}(\mathbf{x})$

In reliability analysis, a limit-state function $\mathrm{g}(\mathbf{x})$ provides the information to classify a state $\mathbf{x}$ in the failure domain:

$$\mathbb{F} = \{\mathbf{x} \in \mathbb{X} : \mathrm{g}(\mathbf{x}) \leq 0\} \tag{1. 55}$$

or the safe domain:

$$\mathbb{F}^{\mathrm{c}} = \{\mathbf{x} \in \mathbb{X} : \mathrm{g}(\mathbf{x}) > 0\} \tag{1. 56}$$

The following indicator function $\mathrm{I}_{\mathrm{F}}(\mathbf{x})$ with binary output is defined:

$$\mathrm{I}_{\mathrm{F}}(\mathbf{x}) = \begin{cases} 1 & \text{for } \mathrm{g}(\mathbf{x}) \leq 0 \\ 0 & \text{for } \mathrm{g}(\mathbf{x}) > 0 \end{cases} \tag{1. 57}$$

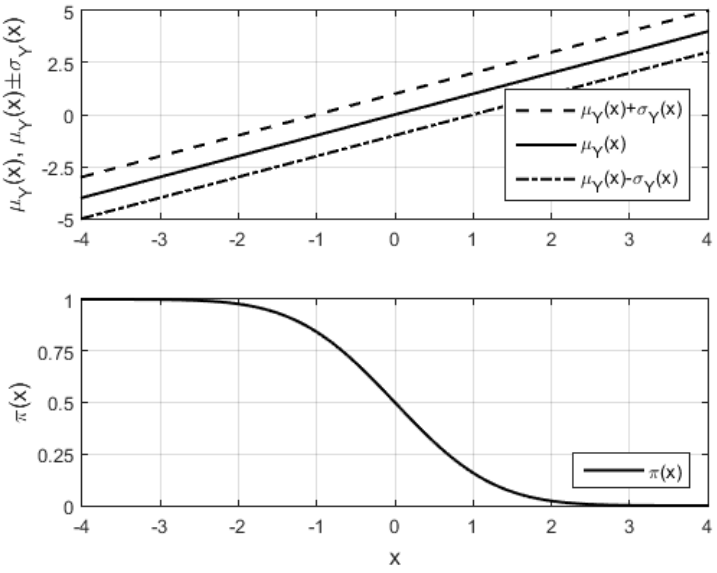
Respectively, when a limit-state function is emulated through a Gaussian process, such as the kriging predictor $\hat{\mathrm{Y}}(\mathbf{x})$, the classification is approached in a probabilistic way, rather than deterministically. The probabilistic classification function $\pi(\mathbf{x})$ describes the probability that a state $\mathbf{x}$ belongs in the failure domain, and is the equivalent of the indicator function $\mathrm{I}_{\mathrm{F}}(\mathbf{x})$ in a probabilistic way. Therefore $\pi(\mathbf{x})$ is defined as:

$$\pi(\mathbf{x}) = \mathcal{P}\big(\hat{\mathrm{Y}}(\mathbf{x}) \leq 0\big) = \Phi\left(\frac{-\mu_{\hat{\mathrm{Y}}}(\mathbf{x})}{\sigma_{\hat{\mathrm{Y}}}(\mathbf{x})}\right) \tag{1. 58}$$

where Φ the standard normal cumulative distribution function. It is clear that $\pi(\mathbf{x})$ takes values from 0 to 1, and that for $\sigma_{\hat{Y}}(\mathbf{x}) = 0$ then $\pi(\mathbf{x}) = I_F(\mathbf{x})$.

In the following example, it is assumed that $\mu_{\hat{Y}}(x) = x$ and $\sigma_{\hat{Y}}(x) = 1$ for $x \in [-4,4]$.

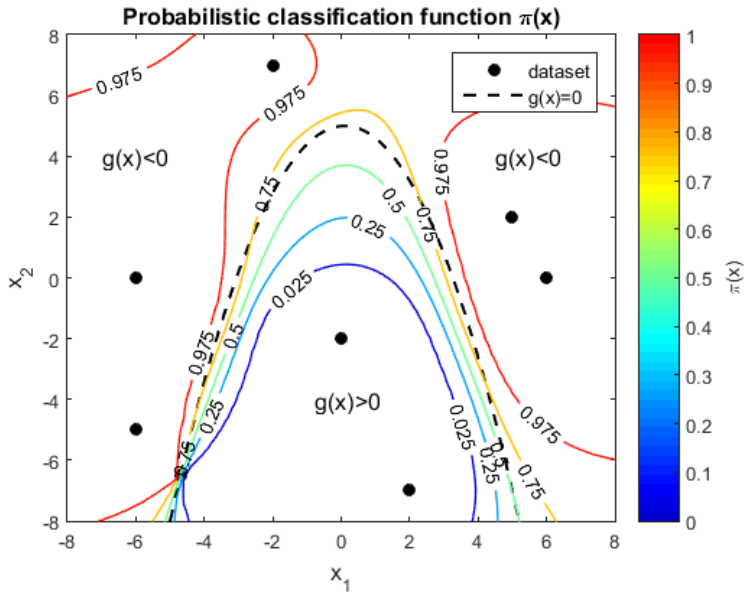
$$\pi(x) = \Phi\left(\frac{-\mu_{\hat{Y}}(x)}{\sigma_{\hat{Y}}(x)}\right) = \Phi(-x)$$



In the second example, a kriging model is built with 8 support points, from the following limit-state function, by Der Kiureghian and Dakessian (1998):

$$g(\mathbf{x}) = b - x_2 - \kappa(x_1 - e)^2, \qquad \mathbf{x} \in \mathbb{X} = [-8,8] \times [-8,8]$$

where $b = 5$, $\kappa = 0.5$ and $e = 0.1$. The kriging parameters $\boldsymbol{\vartheta}$ are set to $\boldsymbol{\vartheta} = (0.25, 0.25)$, the correlation function is the squared exponential and the regressor is constant.



The dashed line represents the limit $g(\mathbf{x}) = 0$ and the cyan-green represents $\pi(\mathbf{x}) = 0.5 \Leftrightarrow \mu_{\hat{Y}}(\mathbf{x}) = 0$. Choosing a confidence-level of 95%, three areas are distinguished:

- $\mathbf{x} \in \mathbb{X} : \pi(\mathbf{x}) \leq 2.5\%$, in which safety is almost certain
- $\mathbf{x} \in \mathbb{X} : \pi(\mathbf{x}) > 97.5\%$, in which failure is almost certain, and finally
- $\mathbf{x} \in \mathbb{X} : 2.5\% < \pi(\mathbf{x}) \leq 97.5\%$, the "grey area" in which the model fails to provide confident classification. In later chapter, the database that is used to build the kriging model, is refreshed with new data, with purpose to eliminate completely this grey area, and bring the $\pi(\mathbf{x})$ function closer to $\mathrm{I}_{\mathrm{F}}(\mathbf{x})$.

2 Reliability analysis

2.1 Introduction

In structural design, many types of uncertainties are encountered. Those uncertainties derive mostly from lack of information, of ignorance of the physical laws. Material properties, environmental loads, and other quantities, are very hard or impossible to determine, therefore these quantities must be dealt us random variables. As a consequence, considering all those uncertainties is necessary, in order to evaluate the probability a structure to fail its purpose, in terms of strength of serviceability.

Firstly, the deterministic and random quantities are separated and defined. The deterministic quantities are those that are known and can be defined in advance. Also random variables with relatively small variance, can be assumed as deterministic. Otherwise, random variables describe quantities, that are unknown beforehand, such us the strength of concrete, or the maximum snow load in the next 50 years. In almost all cases though, such uncertainties can be described in a probabilistic way, through mathematically defined distributions that derive from statistical data.

Reliability analysis aims to calculate the failure probability, which depends on all present uncertainties. In other words, all uncertainties are reduced to a single one, which can describe the safety level of a structure. Nevertheless, evaluating such a failure probability, analytical solutions are proven to be very restricted. One should rely on numerical methods, and more specifically, on computer experiments.

This thesis focuses on the application of structural reliability, in problems with difficulties, such as very low failure probability ($< 10^{-4}$), high number of random variables, and computationally costly performance functions. In this chapter, the most efficient and advanced simulation techniques are presented, incorporating the probabilistic classification, provided by the kriging meta-modelling.

2.2 Problem definition

Let vector $\mathbf{X}$ a collection of m random variables, described by the multivariate probability density function $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$, $\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^{\mathbf{m}}$. The performance function $\mathrm{G} = \mathrm{g}(\mathbf{x})$: $\mathbb{X} \to \mathbb{G} \subset \mathbb{R}$, which is also random since it depends on $\mathbf{X}$, measures a specific quantity, that is adequate to classify a state $\mathbf{x}$ to either:

the failure domain:

$$\mathbb{F} = \{\mathbf{x} \in \mathbb{X} : \mathrm{g}(\mathbf{x}) \leq 0\} \tag{2. 1}$$

or the safe domain:

$$\mathbb{F}^{\mathrm{c}} = \{\mathbf{x} \in \mathbb{X} : \mathrm{g}(\mathbf{x}) > 0\} \tag{2. 2}$$

Therefore, the failure probability is defined as:

$$p_f = \mathcal{P}(G \le 0) = \mathcal{P}(g(\mathbf{X}) \le 0) = \int_{\mathbb{F}} f_{\mathbf{X}}(\mathbf{x})\, dx \tag{2. 3}$$

The indicator function $I_F(\mathbf{x})$ is defined as:

$$I_F(\mathbf{x}) = \begin{cases} 1 & \text{for } g(\mathbf{x}) \le 0 \\ 0 & \text{for } g(\mathbf{x}) > 0 \end{cases} \tag{2. 4}$$

Equivalently, the failure probability can be defined using the real performance function $g(\mathbf{x})$ or the probabilistic classification function $\pi(\mathbf{x})$ from the kriging predictor $\widehat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), \hat{\sigma}_G(\mathbf{x})\right)$:

$g(\mathbf{x})$ $\qquad\qquad$ $\widehat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), \hat{\sigma}_G{}^2(\mathbf{x})\right)$

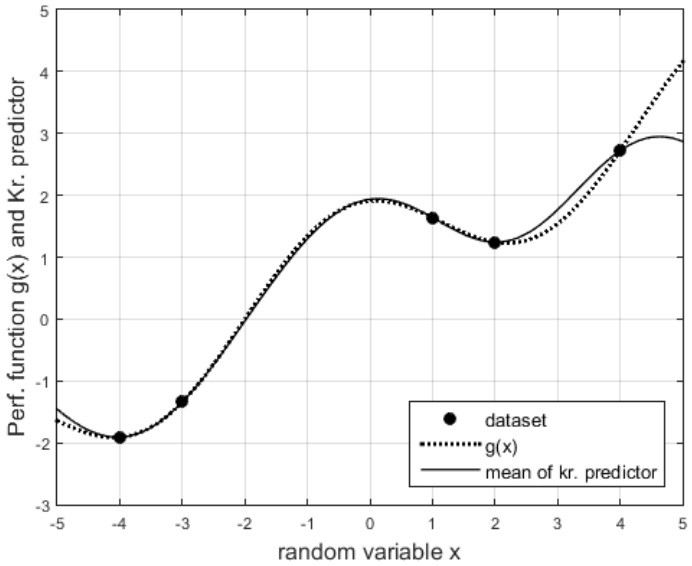
$$p_f = \int_{\mathbb{X}} I_F(\mathbf{x}) \cdot f_{\mathbf{X}}(\mathbf{x})\, dx \tag{2. 5}$$

$$p_f \approx p_{f\varepsilon} = \int_{\mathbb{X}} \pi(\mathbf{x}) \cdot f_{\mathbf{X}}(\mathbf{x})\, dx \tag{2. 6}$$

The failure probability, when evaluated from a kriging model, is only an estimation, because uncertainties of the model are present. In the rest of the chapter, all definitions and methodology, are presented in two ways, in parallel. Firstly, using the real performance function and the indicator function $I_F(\mathbf{x})$, and secondly, through the model estimation and the probabilistic classification function $\pi(\mathbf{x})$.

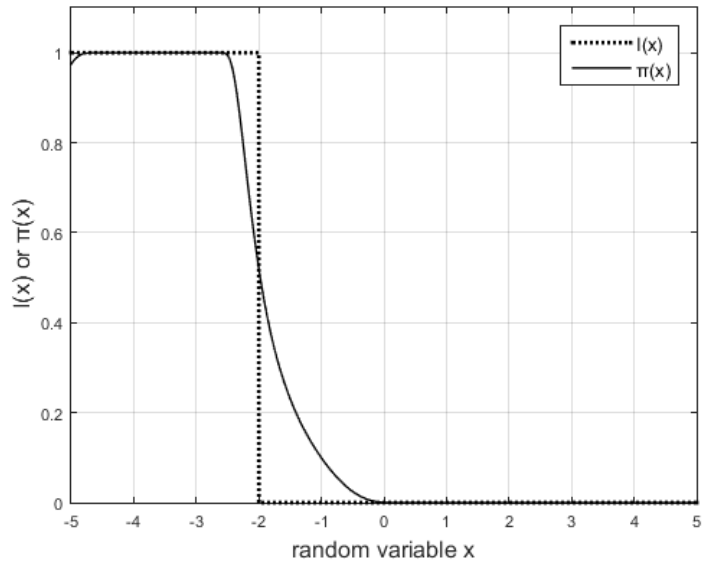
In the following example, the functions $I_F(\mathbf{x})$ and $\pi(\mathbf{x})$ are presented, considering the function of 1.5.4. The random variable $\mathbf{X}$, is normally distributed with zero mean and unit standard deviation. The kriging model is built, using constant regression function, and the squared exponential autocorrelation function with optimal parameter ϑ.

$$g(x) = \frac{1}{2}(x + 2) + \sin(x + 2), \qquad x \in \mathbb{R}$$

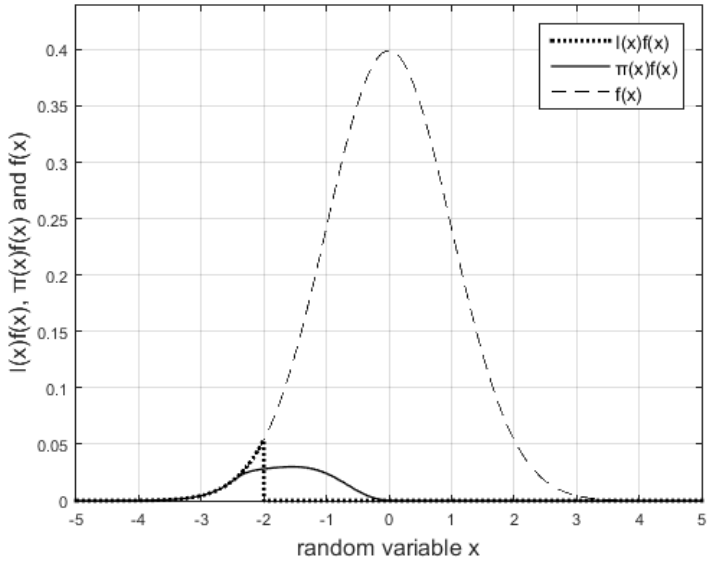


It is worth noticing that, the mean kriging predictor $\mu_{\widehat{G}}(x)$ is almost identical to the real function $g(x)$, in the area of interest. Nevertheless, in the interval $x \in [-3,1]$ there is little

information (no observations) and the model variance is large. This uncertainty is depicted in the following figure:



It is clear that the probabilistic classification function $\pi(\mathbf{x})$, is actually an estimation of the indicator function. Although the estimation, may seem to be relatively good, the estimated failure probability may differ by a large margin to the actual one.



The error of the estimation, is clearer above. The area of the two functions depicted, is equal to the failure probability as defined in the above expressions. In this specific case, the PDF enlargers the error of the estimator.

By numerical integration, the values of $\mathrm{p_f}$ and $\mathrm{p_{f\varepsilon}}$ are obtained:

$$\mathrm{g}(\mathbf{x}) \qquad\qquad \widehat{\mathrm{G}}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_{\mathrm{G}}(\mathbf{x}), \hat{\sigma}_{\mathrm{G}}{}^{2}(\mathbf{x})\right)$$

$$\mathrm{p_f} = 2.275 \cdot 10^{-2} \qquad\qquad \mathrm{p_{f\varepsilon}} = 5.704 \cdot 10^{-2}$$

The calculation of the definite multiple integral in the equations above, is generally impossible to compute analytically, because the bounds are dependent on $g(\mathbf{x})$. Also numerical integration is just as inefficient, because the computational effort is rising exponentially with higher dimensionality. Therefore, the following simulation techniques need to be implemented, to calculate low failure probabilities in many dimensions.

2.3 Monte Carlo Simulation

The Monte Carlo Simulation (MCS), is a method to calculate numerically the integral that defines the failure probability p_f. It is a necessary condition, that generating samples of any size from the random vector $\mathbf{X}$ according to the predetermined PDF $f_X(x)$ is possible. This can be achieved by sampling from a random variable $U \sim \mathcal{U}(0,1)$ using a pseudorandom number generator, and then compute realizations of X through an isoprobabilistic transformation T.

$$x^{(i)} = T\left(u^{(i)}\right), \qquad i = 1, 2, \dots N \tag{2. 7}$$

where:

$x^{(i)}$ realization i of random variable X
$u^{(i)}$ realization i of random variable $U \sim \mathcal{U}(0,1)$
N the sample size, or number of simulations
$T(u)$ isoprobabilistic transformation $U \rightarrow X$

$$T(u) = {F_X}^{-1}\left(F_U(u)\right) = {F_X}^{-1}(u) \tag{2. 8}$$

where: $F_X(x)$ the CDF of X and $F_U(u) = u$ the CDF of U

The calculation of the failure probability is estimated from the sample $x^{(i)},\ i = 1, 2, \dots N$ as follows:

$g(\mathbf{x})$ $\qquad\qquad$ $\hat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), {\hat{\sigma}_G}^2(\mathbf{x})\right)$

$$\hat{p}_f = \frac{1}{N}\sum_{i=1}^{N} I_F\left(x^{(i)}\right) \tag{2. 9}$$

$$\hat{p}_f \approx \hat{p}_{f\varepsilon} = \frac{1}{N}\sum_{i=1}^{N} \pi\left(x^{(i)}\right) \tag{2. 10}$$

The above estimation is dependent on the size sample N, and the coefficient of variation is given below:

$$\delta = \sqrt{\frac{1 - p_f}{N \cdot p_f}} \tag{2. 11}$$

$$\delta_\varepsilon = \frac{1}{p_{f\varepsilon}} \sqrt{\frac{1}{N}\left(\frac{1}{N}\sum_{\iota=1}^{N} \pi(x^{(i)})^2 - {p_{f\varepsilon}}^2\right)} \tag{2. 12}$$

It should be reminded, that $\hat{p}_{f\varepsilon}$ is estimated with uncertainties that originate from the kriging model and uncertainties from the MCS as well. The coefficient of variation δ_ε expresses only the probable error from the MCS, which is also called statistical error.

Usually, the coefficient of variation of the failure probability δ (or δ_ε), defines the number of simulations N, and is set to one of the following values: $\delta = \{1\%, 2\%, 5\%, 10\%\}$.

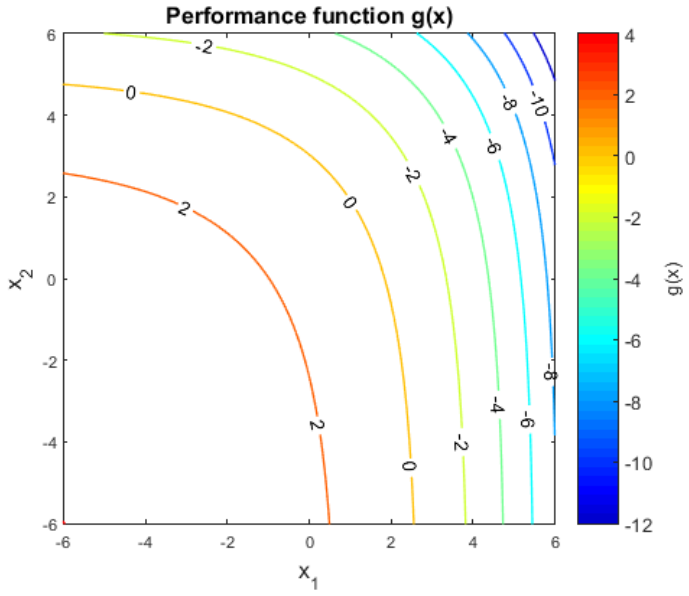
Monte Carlo Simulation is a powerful and robust method to calculate the failure probability. It is independent of dimensionality and relies only on $I_F(\mathbf{x})$, meaning that the differentiability, the linearity or the convexity of $g(\mathbf{x})$ are irrelevant. On the other hand, MCS may prove to be computationally costly, because achieving low δ requires many simulations when estimating low failure probabilities. This disadvantage may often be amplified, when each evaluation of $g(\mathbf{x})$ is also computationally time consuming.

In the following example, the Monte Carlo Simulation is illustrated, using a performance function from Rackwitz (2001):

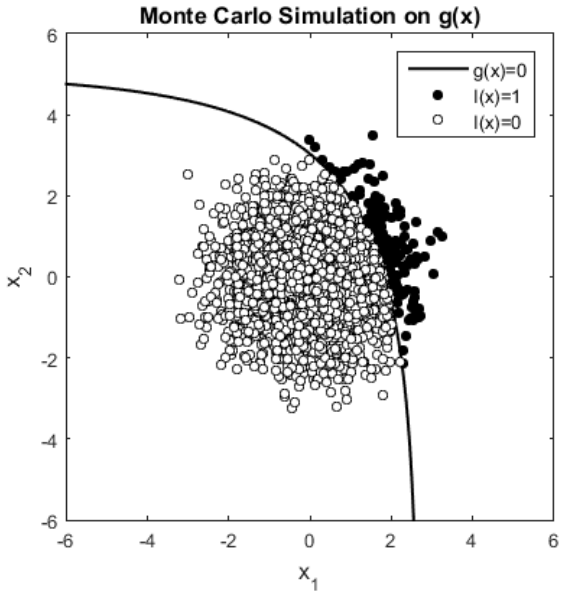
$$g(\mathbf{x}) = 4.5 - e^{0.3 \cdot x_1 + 0.7} - e^{0.3 \cdot x_2}, \qquad x_1, x_2 \in \mathbb{R}^2$$

where:

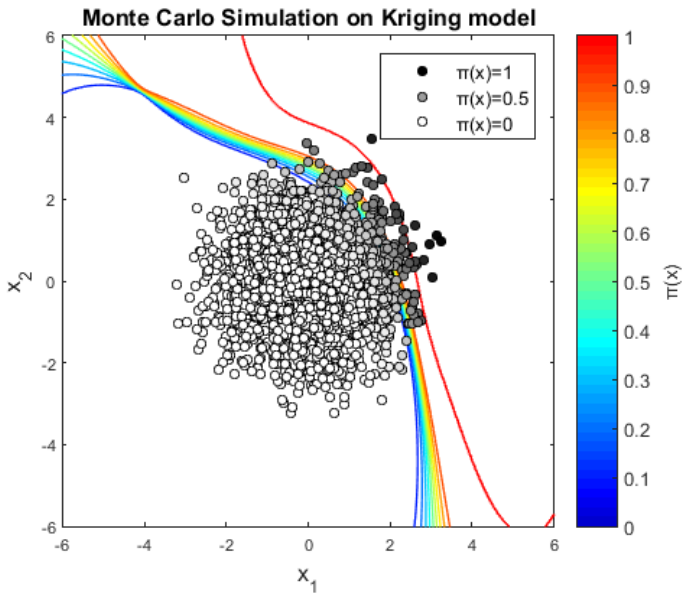
x_1, x_2 normally distributed random variables, with zero mean and unit variance



The kriging model is built with 10 observations. The MCS is run on sample with size $N = 3000$.



The samples in the failure domain ($I_F(\mathbf{x}) = 1$), are illustrated with filled black dots.



The colors in gray scale, correspond to values of $\pi(\mathbf{x})$ from 0 to 1. The failure probability and its coefficient of variation, are calculated from the real $g(\mathbf{x})$ function and from the kriging model:

$g(\mathbf{x})$	$\hat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), \hat{\sigma}_G{}^2(\mathbf{x})\right)$
$\hat{p}_f = 4.47 \cdot 10^{-2}$	$\hat{p}_{f\varepsilon} = 2.89 \cdot 10^{-2}$
$\delta = 8.44\%$	$\delta_\varepsilon = 9.20\%$

The difference between $\hat{p}_f$ and $\hat{p}_{f\varepsilon}$ is attributed to the inaccuracy of the kriging model, rather than the statistical error, because δ and δ_ε are too small to justify it. It is worth noticing, that in the first case $N = 3000$ evaluations of $g(x)$ are required, while in the second only 10 evaluations of $g(x)$ are required and N evaluations of the kriging model.

2.4 Importance Sampling

Importance sampling allows the use of an arbitrary PDF $h(\mathbf{x})$ in sampling generation in simulation techniques, unlike MCS which requires the use of the PDF $f_{\mathbf{X}}(\mathbf{x})$ that defines the random vector $\mathbf{X}$. Although it is proven that any $h(\mathbf{x})$ is utilizable, in a simulation context the proper choice of $h(\mathbf{x})$ can drastically decrease the number of simulations N. The optimal $h(\mathbf{x})$ focuses entirely on the failure domain, thus, the sampling is more efficient. However, defining the optimal $h(\mathbf{x})$ is not possible, but suboptimal PDFs can be almost as efficient.

Importance sampling is presented below, at a more general framework. The purpose is to evaluate $\bar{\phi}$ which is defined by the following integral:

$$\bar{\phi} = \int_{\mathbb{X}} \phi(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}_{\mathbf{X}}[\phi(\mathbf{X})] \tag{2. 13}$$

where $\mathbf{X} \in \mathbb{X} \subset \mathbb{R}^m$ the random vector with PDF $f_{\mathbf{X}}(\mathbf{x})$ and $\phi(\mathbf{x})$ a real-valued function. It should be noted that $\bar{\phi}$ is actually the failure probability p_f, when $\phi(\mathbf{x})$ is substituted by $I_F(\mathbf{x})$

or $\pi(\mathbf{x})$.

Let $h(\mathbf{x})$ a PDF, such that $\phi(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) \neq 0 \Longrightarrow h(\mathbf{x}) \neq 0$. The expression of $\bar{\phi}$ is rewritten as follows:

$$\bar{\phi} = \int_{\mathbb{X}} \phi(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}) \frac{h(\mathbf{x})}{h(\mathbf{x})} d\mathbf{x} = \int_{\mathbb{X}} \frac{\phi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{h(\mathbf{x})} h(\mathbf{x})\, d\mathbf{x} = \mathbb{E}_{Y}\left[\frac{\phi(\mathbf{Y}) f_{\mathbf{X}}(\mathbf{Y})}{h(\mathbf{Y})}\right] \tag{2. 14}$$

where $\mathbf{Y} \in \mathbb{X} \subset \mathbb{R}^{m}$ a random vector distributed according to $h(\mathbf{y})$

Resorting to the Monte Carlo Simulation, the estimated value of $\bar{\phi}$ is the following:

$$\hat{\bar{\phi}} = \frac{1}{N} \sum_{i=1}^{N} \frac{\phi(\mathbf{y^{(i)}}) f_{\mathbf{X}}(\mathbf{y^{(i)}})}{h(\mathbf{y^{(i)}})} \tag{2. 15}$$

where:

N the sample size, or number of simulations
$\mathbf{y^{(i)}}, i = 1,2, \ldots N$ the sample generated according to PDF $h(\mathbf{y})$

Next, the variance of $\hat{\bar{\phi}}$ is evaluated:

$$\begin{aligned} \sigma_{\hat{\bar{\phi}}}{}^{2} = \mathrm{Var}_{Y}\left[\hat{\bar{\phi}}\right] &= N \cdot \mathbb{E}_{Y}\left[\frac{1}{N^{2}} \frac{\phi(\mathbf{Y})^{2} f_{\mathbf{X}}(\mathbf{Y})^{2}}{h(\mathbf{Y})^{2}}\right] - N \cdot \mathbb{E}_{Y}\left[\frac{1}{N} \frac{\phi(\mathbf{Y}) f_{\mathbf{X}}(\mathbf{Y})}{h(\mathbf{Y})}\right]^{2} \\ &= \frac{N}{N^{2}}\left(\int_{\mathbb{X}} \frac{\phi(\mathbf{y})^{2} f_{\mathbf{X}}(\mathbf{y})^{2}}{h(\mathbf{y})^{2}} h(\mathbf{y}) d\mathbf{y} - \hat{\bar{\phi}}^{2}\right) = \frac{1}{N}\left(\frac{1}{N} \sum_{i=1}^{N} \frac{\phi(\mathbf{y^{(i)}})^{2} f_{\mathbf{X}}(\mathbf{y^{(i)}})^{2}}{h(\mathbf{y^{(i)}})^{2}} - \hat{\bar{\phi}}^{2}\right) \end{aligned} \tag{2. 16}$$

And the coefficient of variation of $\hat{\bar{\phi}}$ is:

$$\delta_{\hat{\bar{\phi}}} = \frac{1}{\hat{\bar{\phi}}} \sqrt{\frac{1}{N}\left(\frac{1}{N} \sum_{i=1}^{N} \frac{\phi(\mathbf{y^{(i)}})^{2} f_{\mathbf{X}}(\mathbf{y^{(i)}})^{2}}{h(\mathbf{y^{(i)}})^{2}} - \hat{\bar{\phi}}^{2}\right)} \tag{2. 17}$$

Assuming $\sigma_{\hat{\bar{\phi}}}{}^{2} = 0$ and $\phi(\mathbf{y}) \geq 0$, the optimal $h(\mathbf{y})$ is defined (Rubinstein and Kroese 2008):

$$\begin{gathered} \sigma_{\hat{\bar{\phi}}}{}^{2} = 0 \Longrightarrow \int_{\mathbb{X}} \frac{\phi(\mathbf{y})^{2} f_{\mathbf{X}}(\mathbf{y})^{2}}{h(\mathbf{y})^{2}} h(\mathbf{y}) d\mathbf{y} - \hat{\bar{\phi}}^{2} = 0 \Longrightarrow \\ \int_{\mathbb{X}} \frac{\phi(\mathbf{y})^{2} f_{\mathbf{X}}(\mathbf{y})^{2}}{\hat{\bar{\phi}} \cdot h(\mathbf{y})} d\mathbf{y} = \int_{\mathbb{X}} \phi(\mathbf{y})\, f_{\mathbf{X}}(\mathbf{y}) d\mathbf{y} \Longrightarrow \frac{\phi(\mathbf{y}) f_{\mathbf{X}}(\mathbf{y})}{\hat{\bar{\phi}} \cdot h(\mathbf{y})} = 1 \Longrightarrow \\ h(\mathbf{y}) = \frac{\phi(\mathbf{y}) f_{\mathbf{X}}(\mathbf{y})}{\hat{\bar{\phi}}} \end{gathered} \tag{2. 18}$$

Clearly, the optimal instrumental PDF $h(\mathbf{y})$ cannot be defined because the value $\hat{\bar{\phi}}$ is not known. Nevertheless, an estimation of the optimal $h(\mathbf{y})$, or else a suboptimal $h(\mathbf{y})$, may also grant satisfactory results. In Meta-Modeling Importance Sampling the suboptimal PDF is estimated from the kriging model. More on that will be discussed in the respective chapter.

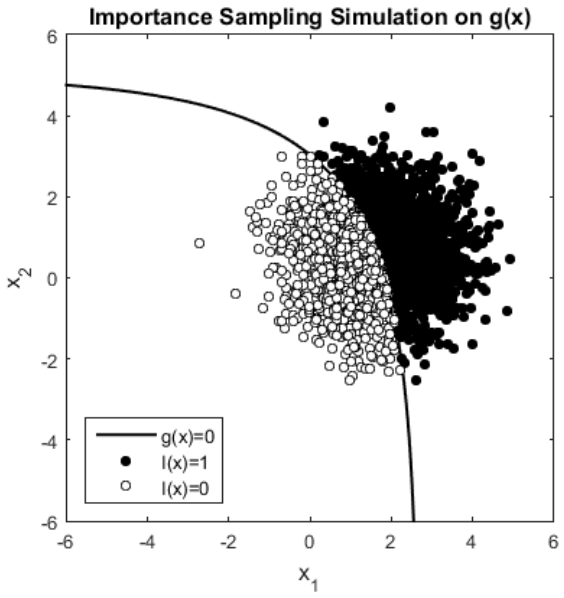
In the following example, the failure probability is estimated with the Importance Sampling technique, considering the performance function $\mathrm{g}(\mathbf{x})$ of 1.3.

The function $\mathrm{h}(\mathbf{x})$ is arbitrarily chosen for the purpose of this example:

$$\mathrm{h}(\mathbf{x}) = \mathrm{f}_{\mathbf{X}}(\mathbf{x} - \mathbf{d})$$

where $\mathbf{d}$ is the vector that maximizes $\mathrm{f}_{\mathbf{X}}(\mathbf{d})$ with the constraint that $\mathbf{d} \in \mathbb{F} \subset \mathbb{X}$. This vector is also called the design point, because it is the most probable case of failure.

In other words, $\mathrm{h}(\mathbf{x})$ defines a normal distribution with the same variance as $\mathbf{X}$, $\boldsymbol{\sigma}^2 = [1,1]$, and mean $\mathbf{m} = \mathbf{d} = \{1.646,0.609\}$. The sample size is the same as in the MCS example, $\mathrm{N} = 3000$.



As shown in the figure, about half of the samples reside in the failure domain. The failure probability and its coefficient of variation are estimated below:

$$\hat{\mathrm{p}}_{\mathrm{f}} = 4.84 \cdot 10^{-2}, \qquad \delta = 2.62\%$$

Comparing to the MCS, δ is now about 3.2 times smaller, and such a value would be achieved with about 10 times larger sample size. It is worth noticing though, that choosing a proper $\mathrm{h}(\mathbf{x})$ function is very crucial. An improper $\mathrm{h}(\mathbf{x})$ choice, would result to lower efficiency compared to the robust MCS.

Importance sampling provides a property that is fundamental for Subset Simulation and Meta-Model Based Importance Sampling, described in the following subchapters.

2.5 Subset Simulation

2.5.1 Principle

Subset Simulation is a reliability method, by Au and Beck (2001), that splits the domain in successive subsets, with purpose to detect the failure domain and estimate very low failure probabilities. Starting from the initial domain, $\mathbb{X} \subset \mathbb{R}^{\mathrm{m}}$ the limit $\mathrm{q}_1 \in \mathbb{G} \subset \mathbb{R}$ is evaluated so that:

$$\mathcal{P}(\mathbf{X} \in \mathbb{F}_1) = \mathcal{P}(g(\mathbf{X}) \leq q_1, \mathbf{x} \in \mathbb{X}) = p_1 \tag{2. 19}$$

where:

$\mathbb{F}_1$ a subset of $\mathbb{X}$ such that $g(\mathbf{x}) \leq q_1, \mathbf{x} \in \mathbb{X}$
$g(\mathbf{x})$ the performance function $g(\mathbf{x})$: $\mathbb{X} \to \mathbb{G} \subset \mathbb{R}$
p_1 an arbitrarily chosen probability referred to as an intermediate probability

Subsequently, the domain $\mathbb{F}_1$ is split into subsets $\mathbb{F}_s \subset \mathbb{F}_{s-1} \subset \cdots \subset \mathbb{F}_2 \subset \mathbb{F}_1$ successively:

$$\mathcal{P}(\mathbf{X} \in \mathbb{F}_i | \mathbf{X} \in \mathbb{F}_{i-1}) = \mathcal{P}(g(\mathbf{X}) \leq q_i, \mathbf{x} \in \mathbb{F}_{i-1}) = p_{i|i-1}, \qquad i = 2,3, \dots s \tag{2. 20}$$

where: $q_s < q_{s-1} < \cdots q_2 < q_1$ the limits that define subsets $\mathbb{F}_i = \{\mathbf{x} \in \mathbb{X} : g(\mathbf{x}) \leq q_i\}$
$p_{i|i-1}, i = 1,2, \dots s-1$ are the intermediate probabilities, arbitrarily chosen
$p_{s|s-1}$ the conditional probability of the last subset that corresponds to q_s

The failure domain is defined as $\mathbb{F} = \{\mathbf{x} \in \mathbb{X} : g(\mathbf{x}) \leq q_s = 0\}$, therefore: $\mathbb{F} = \mathbb{F}_s$

And the failure probability p_F is given from the Bayes' Theorem:

$$p_f = \mathcal{P}(\mathbf{X} \in \mathbb{F}) = \mathcal{P}(\mathbf{X} \in \mathbb{F}_s | \mathbf{X} \in \mathbb{F}_{s-1}) \cdot \dots \cdot \mathcal{P}(\mathbf{X} \in \mathbb{F}_2 | \mathbf{X} \in \mathbb{F}_1) \cdot \mathcal{P}(\mathbf{X} \in \mathbb{F}_1) = p_1 \prod_{i=2}^{s} p_{i|i-1} \tag{2. 21}$$

In the following equations, the intermediate probabilities are additionally defined, from the probabilistic classification function $\pi(\mathbf{x})$.

$g(\mathbf{x})$

$$p_1 = \int_{\mathbb{X}} I_{F1}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{2. 22}$$

$$p_{i|i-1} = p_i / p_{i-1} \Longrightarrow$$
$$p_{i|i-1} = \int_{\mathbb{F}_{i-1}} \frac{I_{Fi}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{i-1}} d\mathbf{x} \Longrightarrow$$
$$p_{i|i-1} = \int_{\mathbb{F}_{i-1}} \frac{I_{Fi}(\mathbf{x})}{I_{Fi-1}(\mathbf{x})} \frac{I_{Fi-1}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{i-1}} d\mathbf{x} \Longrightarrow$$
$$p_{i|i-1} = \int_{\mathbb{F}_{i-1}} \frac{I_{Fi}(\mathbf{x})}{I_{Fi-1}(\mathbf{x})} h_i(\mathbf{x}) d\mathbf{x} \Longrightarrow$$
$$p_{i|i-1} = \int_{\mathbb{F}_{i-1}} I_{Fi}(\mathbf{x}) h_i(\mathbf{x}) d\mathbf{x} \tag{2. 24}$$

$$I_{Fi}(\mathbf{x}) = \begin{cases} 1 & \text{for } g(\mathbf{x}) \leq q_i \\ 0 & \text{for } g(\mathbf{x}) > q_i \end{cases} \tag{2. 26}$$

$$h_i(\mathbf{x}) = \frac{I_{Fi-1}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{i-1}} \tag{2. 28}$$

$\widehat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), \hat{\sigma}_G{}^2(\mathbf{x})\right)$

$$p_{\varepsilon,1} = \int_{\mathbb{X}} \pi_1(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{2. 23}$$

$$p_{\varepsilon,i|i-1} = p_i / p_{i-1} \Longrightarrow$$
$$p_{\varepsilon,i|i-1} = \int_{\mathbb{F}_{i-1}} \frac{\pi_i(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{i-1}} d\mathbf{x} \Longrightarrow$$
$$p_{\varepsilon,i|i-1} = \int_{\mathbb{F}_{i-1}} \frac{\pi_i(\mathbf{x})}{\pi_{i-1}(\mathbf{x})} \frac{\pi_{i-1}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{i-1}} d\mathbf{x} \Longrightarrow$$
$$p_{\varepsilon,i|i-1} = \int_{\mathbb{F}_{i-1}} \frac{\pi_i(\mathbf{x})}{\pi_{i-1}(\mathbf{x})} h_i(\mathbf{x}) d\mathbf{x} \tag{2. 25}$$

$$\pi_i(\mathbf{x}) = \Phi\left(\frac{q_i - \mu_{\widehat{G}}(\mathbf{x})}{\sigma_{\widehat{G}}(\mathbf{x})}\right) \tag{2. 27}$$

$$h_i(\mathbf{x}) = \frac{\pi_{i-1}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{\varepsilon,i-1}} \tag{2. 29}$$

In Subset Simulation, the conditional intermediate probabilities $p_{i|i-1}$ up to $i = s - 1$ as well as $q_s = 0$ are known. The last conditional probability $p_{s|s-1}$ and the quantiles of $g(\mathbf{x})$ $q_1 > q_2 > \cdots > q_{s-1} > 0$, are derived successively from the definitions above. The function $h_i(\mathbf{x})$ is the PDF that describes the distribution of $\mathbf{X}$ restricted in subset $\mathbb{F}_{i-1}$, and is referred to as the conditional PDF. In the next subchapter, an efficient procedure for the estimation of the failure probability is presented.

2.5.2 Algorithm

The main advantage of Subset Simulation (SS), is that the intermediate probabilities are arbitrary, and can be relatively high, such as $p_{i|i-1} = 10\%\sim20\%$. When resorting to simulation methods, high probabilities are very costly to be estimated with low variance. Therefore, SS alters the problem. Instead of calculating a low failure probability, high probabilities are calculated, but multiple times.

The computational cost of SS, in terms of δ, is (by simplification) proportionate to k, if the failure probability is $p_f = 10^{-k}, k > 0$, whereas the computational cost of MCS is proportionate to 10^k.

On the other hand, SS depends on more advanced techniques in sampling generation, because an isoprobabilistic transformation of $h_i(\mathbf{x})$ is inevitable or impractical. Such techniques are the Markov chain Monte Carlo samplers (MCMC). In this thesis the modified Metropolis-Hastings sampler is preferred, as discussed in following chapter.

Step 0: Set values for the intermediate probabilities, and sample size per step N

The intermediate probabilities $p_{i|i-1}, i = 1,2, \dots s - 1$ are all set to a certain value p_0. A small value of p_0 means that less steps will be required, but a targeted probability variance will be harder to reach. By choosing a large value, smaller samples need to be generated but more steps are required. An optimal choice is $0.1\sim0.2$ and most commonly $p_0 = 0.1$. Therefore:

$$p_{i|i-1} = p_0, \qquad i = 1,2, \dots s - 1$$

The choice of sample size per step N depends on the target coefficient of variation δ of p_f. If the failure probability p_f can be estimated, then from the upper bound of δ:

$$N > \frac{1 - p_0}{\delta^2 \cdot p_0} \cdot \left(1 + \sqrt{1 + \gamma} \cdot (k - 1)\right)^2, \qquad k = \frac{\log(p_f)}{\log(p_0)} \tag{2. 30}$$

or less conservatively, from the lower bound:

$$N > \frac{1 - p_0}{\delta^2 \cdot p_0} \cdot \left(1 + (1 + \gamma) \cdot (k - 1)\right) \tag{2. 31}$$

For values $p_0 = 0.1$, $\delta = 10\%$, $\gamma = 3$, $p_f = 10^{-7} \Longrightarrow N > 152100$ per step. The coefficient γ takes into account the correlation in samples generated from an MCMC sampler. More are discussed in following chapter.

Step 1: Generate sample from $f_{\mathbf{X}}(\mathbf{x})$, $i = 1$

The first step is identical to MCS. The samples $\mathbf{x}^{[i=1](k)}, k = 1,2, \dots N$ is generated according to $f_{\mathbf{X}}(\mathbf{x})$.

Step 2: Estimate q_i

Estimating the quantile, q_i can be achieved from $p_{i|i-1} = p_0$. The procedure is different in the case of the kriging model:

$g(\mathbf{x})$

$$p_{i|i-1} = \frac{1}{N}\sum_{k=1}^{N} I_{Fi}\left(\mathbf{x}^{[i](k)}\right) \quad (2.32)$$

Find q_i:

1. Calculate $g\left(\mathbf{x}^{[i](k)}\right), k = 1,2, \dots N$
2. Sort $\mathbf{x}^{[i](k)}$ by ascending order with respect to $g\left(\mathbf{x}^{[i](k)}\right)$
3. $q_i = g\left(\mathbf{x}^{[i](p_0 \cdot N)}\right)$

$$\delta_i = \sqrt{\frac{1 - p_{i|i-1}}{N \cdot p_{i|i-1}}(1 + \gamma_i)} \quad (2.34)$$

$\hat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), \hat{\sigma}_G{}^2(\mathbf{x})\right)$

$$p_{\varepsilon,i|i-1} = \frac{1}{N}\sum_{k=1}^{N} \frac{\pi_i\left(\mathbf{x}^{[i](k)}\right)}{\pi_{i-1}\left(\mathbf{x}^{[i](k)}\right)} \quad (2.33)$$

Find q_i:

1. Set acceptable error $e = 0.01$
2. Calculate $q_l = \min \hat{\mu}_G\left(\mathbf{x}^{[i](k)}\right)$ and $q_u = \max \hat{\mu}_G\left(\mathbf{x}^{[i](k)}\right)$
3. Set $q^* = (q_l + q_u)/2$
4. Estimate $p^* = \frac{1}{N}\sum_{k=1}^{N} \frac{\pi^*\left(\mathbf{x}^{[i](k)}\right)}{\pi_{i-1}\left(\mathbf{x}^{[i](k)}\right)}$ where $\pi^*(\mathbf{x}) = \Phi\left(\frac{q^* - \hat{\mu}_G(\mathbf{x})}{\hat{\sigma}_G(\mathbf{x})}\right)$
5. If $p^* > p_0$ set $q_u = q^*$, else set $q_l = q^*$
6. Calculate $e^* = |p^* - p_0|/p_0$
7. If $e^* > e$ go to 3, else $p_{i|i-1} = p^*$ and $q_i = q^*$

$$\delta_{\varepsilon,i} = \cdots$$

The coefficient of variation of $p_{\varepsilon,i|i-1}$ is the following:

$$\delta_{\varepsilon,i} = \frac{1}{p_{\varepsilon,i|i-1}}\sqrt{\frac{1}{N}\left[\frac{1}{N}\sum_{k=1}^{N} \frac{\pi_i(\mathbf{x}^{[i](k)})}{\pi_{i-1}(\mathbf{x}^{[i](k)})} - p_{\varepsilon,i|i-1}{}^2\right](1 + \gamma_{\varepsilon,i})} \quad (2.35)$$

In step $i = 1$ the $\mathbb{F}_0$ "subset" $(i - 1 = 0)$ is by definition the $\mathbb{X}$ space ($\mathbb{F}_0 = \mathbb{X}$). Also, $\gamma_1 = 0$. As shown in the algorithm above, a sorting of the $g(\mathbf{x})$ values is required to find q_i. In the case of the kriging model though, such sorting is not possible. A bisection method can be utilized, to find q_i. It should be noted, that the intermediate probability p_0 is targeted, but is not achieved. Consequently, the final intermediate probabilities $p_{i|i-1}$ are close and not equal to p_0, but that has no effect to the process of SS.

Eventually, q_i will reach a negative value. At this point q_i is instead, set to zero:

$$q_i = q_s = 0$$

and the algorithm proceeds to step 4. Else, i is set to $i = i + 1$, and the algorithm continues to step 3.

Step 3: Generate sample $\mathbf{x}^{[i](k)}, k = 1,2, \dots N$ from conditional PDF $h_i(\mathbf{x})$

As mentioned before, sampling from $h_i(\mathbf{x})$ can be achieved by Markov Chain Monte Carlo samplers. Here, the Modified Metropolis-Hastings algorithm (MMHA) is used. Firstly an initial seed is defined, and secondly the sample $\mathbf{x}^{[i](k)}, k = 1,2, \dots N$ is generated:

$g(\mathbf{x})$	$\widehat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), \hat{\sigma}_G{}^2(\mathbf{x})\right)$
$\mathbf{x}_{seed}{}^{(k)} = \mathbf{x}^{[i-1](k)}, k = 1,2, \dots p_0 N$ where $\mathbf{x}^{[i-1](k)}$ is the sorted sample from the previous step. Therefore, $\mathbf{x}^{[i-1](k)} \in \mathbb{F}_{i-1}$ and $h_i(\mathbf{x}^{[i-1](k)}) > 0, \forall\, k = 1,2, \dots p_0 N$.	Find $\mathbf{x}_{seed}$ 1. $j = 1$ 2. Choose a random $k^* \in \{1,2, \dots N\}$ with probability: $r_k = \frac{\pi_{i-1}(\mathbf{x}^{[i-1](k)})}{\pi_{i-2}(\mathbf{x}^{[i-1](k)}) \cdot N p_{i-1\|i-2}}$ 3. $\mathbf{x}_{seed}{}^{(j)} = \mathbf{x}^{[i-1](k^*)}$ 4. $j = j + 1$ 5. If $j \leq p_0 N$ go to 2, else end
The sample is generated starting from $p_0 N$ independent chains, with $1/p_0 - 1$ steps, and finally sample $\mathbf{x}^{[i](k)}, k = 1,2, \dots N$ is obtained. Also, the coefficient γ_i is estimated.	The sample is generated starting from $p_0 N$ independent chains with $b + t/p_0$ steps, where b is the burn-in parameter, and t the thinning parameter ($b = 20, t = 10$). Burn-in is very important, because it corrects double samples in the initial seed, and thinning, can greatly reduce the γ_i coefficient. Those two techniques are discussed further in following chapter. Finally, the sample $\mathbf{x}^{[i](k)}, k = 1,2, \dots N$ is obtained.

Consequently, the initial seed, in the kriging model case, likely contains multiple identical samples. This is corrected with the burn-in procedure, meaning that the first b steps in the MMHA are discarded. Go to step 2.

Step 4: Final step. Estimation of p_f and δ.

At this point the intermediate probabilities $p_{i|i-1}, i = 1,2, \dots, s-1$, the coefficients of variation $\delta_i, i = 1,2, \dots, s-1$ as well as the quantiles $q_i, i = 1,2, \dots, s$ are estimated. The last conditional probability is:

$$p_{s|s-1} = \frac{1}{N} \sum_{k=1}^{N} I_{Fs}(\mathbf{x}^{[s](k)}) \qquad p_{\varepsilon,s|s-1} = \frac{1}{N} \sum_{k=1}^{N} \frac{\pi_s(\mathbf{x}^{[s](k)})}{\pi_{s-1}(\mathbf{x}^{[s](k)})}$$

where, $I_{Fs}(\mathbf{x}) = I_F(\mathbf{x})$ and $\pi_s(\mathbf{x}) = \pi(\mathbf{x})$ by definition, because $q_s = 0$.

Therefore:

$$\hat{p}_f = p_1 \prod_{i=2}^{s} p_{i|i-1}$$

The coefficient of variation of $\hat{p}_f$ cannot be determined. Instead, a lower and upper bound are defined (by Au and Beck 2001):

$$\sqrt{\sum_{i=1}^{s} \delta_i^2} \leq \delta \leq \sqrt{\sum_{i=1}^{s} \sum_{j=1}^{s} \delta_i \delta_j} \qquad (2.\ 36)$$

In many examples (Au et al. 2007), it is shown that δ is closer to the lower bound. Also, it should be noted that the minimum size sample per step N, calculated in Step 0, is based on the upper and the lower bound respectively.

2.5.3 Example

In the following example, the failure probability is estimated from Subset Simulation. The performance function is the following:

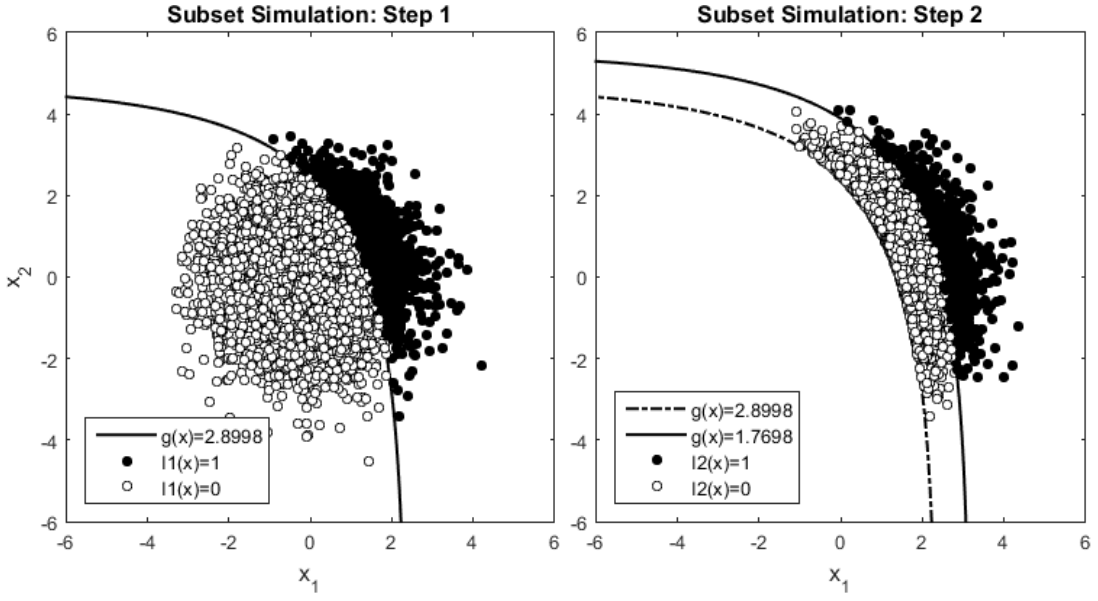
$$g(\mathbf{x}) = 7 - e^{0.3 \cdot x_1 + 0.7} - e^{0.3 \cdot x_2}, \qquad x_1, x_2 \in \mathbb{R}^2$$

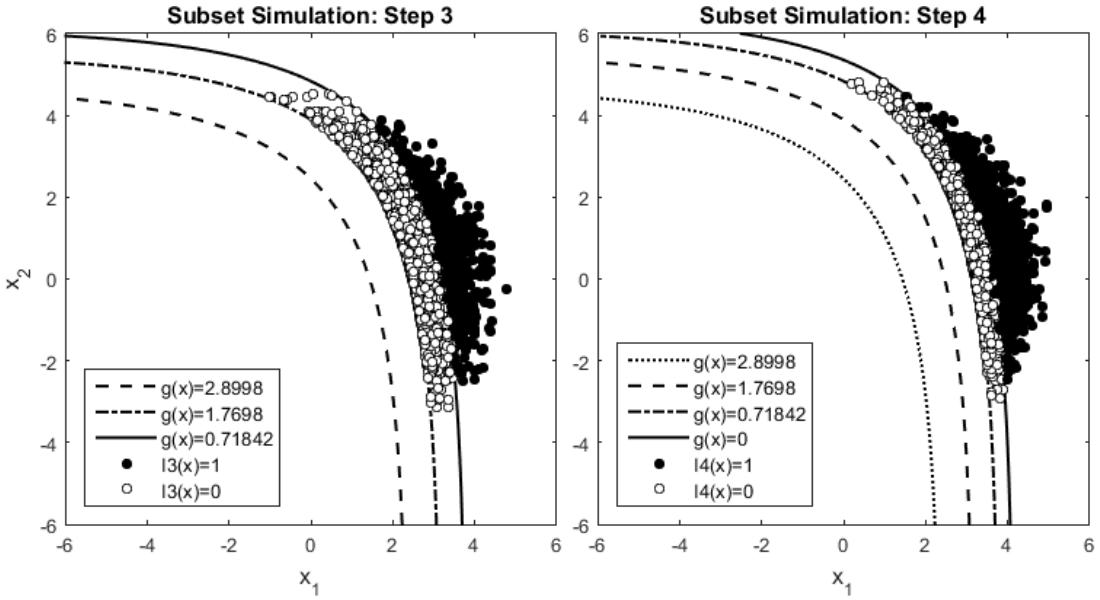
where x_1, x_2 normally distributed random variables, with zero mean and unit variance

Firstly, the failure probability is estimated from Monte Carlo Simulation, for reference:

$$p_f = 2.137 \cdot 10^{-4}, \delta = 3.66\% \,, N = 3.5 \cdot 10^6$$

Starting the Subset Simulation algorithm, the size sample per step is set to $N = 10^5$, and the intermediate probability to $p_0 = 0.1$. The results are shown in the following table. Also the samples of each step, are illustrated. It should be noted, that only $1/10$ of the samples is plotted for the purpose of the illustration:





The black dots in each step, represent the samples that reside in the lowest 10% of $g(\mathbf{x})$ values, and the highest of those defines the quantile q_i. These samples, are also the seed used to generate the samples of the next step.

i	$p_{i\|i-1}$	q_i	δ_i	γ_i
1	0.1	2.900	0.95%	0
2	0.1	1.770	1.84%	2.757
3	0.1	0.718	1.79%	2.560
4	0.215	0	1.23%	3.137

In the fourth step, the quantile q_4 is found $q_4 = -0.351$. Therefore, it is set to $q_4 = 0$, and the final intermediate probability is estimated: $p_{4|3} = 0.215$. The failure probability is:

$$\hat{p}_f = p_1 \cdot \prod_{i=2}^{4} p_{i|i-1} = 0.1^3 \cdot 0.215 = 2.15 \cdot 10^{-4}$$

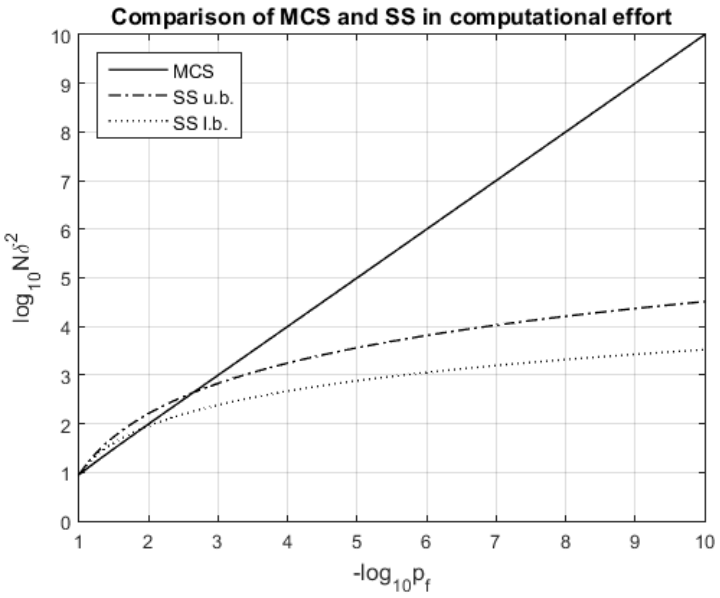
The coefficient γ_i accounts for the correlation of the generated sample, which is inevitable in Markov chains. Its value could be reduced, by increasing the number of chains, i.e. the size of the seed. In all three cases, in which sample was generated from the MMHA, $\gamma_i \approx 3$, while the samples in the first step are completely independent and uncorrelated, therefore $\gamma_1 = 0$. In case a kriging model is built, γ_i can be reduced to zero, due to thinning. The coefficient of variation of p_f cannot be estimated, but is bounded from the following values:

$$3.00\% \leq \delta \leq 5.81\%$$

2.5.4 Comparison of the Subset Simulation and the Monte Carlo Simulation

Comparing the SS to the MCS, it is clear that, in the example above, the same variance of probability is achieved, with much less computational effort. In terms of $g(\mathbf{x})$ evaluations, the MCS is completed with $N_{MC}/N_{SS} = (3.5 \cdot 10^6)/(4 \cdot 10^5) = 8.75$ times more evaluations. Generally, the ratio N_{MC}/N_{SS} increases rapidly, while the failure probability decreases, and the SS becomes more efficient.

In the graph below, the required total sample size N is shown, for a target c.o.v. δ, and a failure probability p_f. It is considered that $p_0 = 0.1$ and $\gamma_i = 3$, which are common values. Both bounds of δ of SS are shown.



As shown in the graph, Subset Simulation is preferred to MCS, for low failure probabilities, such as $p_f < 10^{-4}$.

2.6 Meta-Model Based Importance Sampling

2.6.1 Introduction

Meta-Model Based Importance Sampling (MIS) (Dubourg 2011), is a hybrid method that combines kriging modeling with importance sampling. The reliability analysis methods described in this chapter, are based on numerous evaluations of the performance function $g(\mathbf{x})$ to generate sample, or very few evaluations, with which the kriging model is built. The kriging model approach can provide the same accuracy in estimating the failure probability, with low computational effort, provided it is refined with additional observations, carefully selected as presented in the next chapter. However, it is noticed, that in certain cases, a kriging model cannot be sufficiently refined. Most frequently, in problems with high dimensionality, a kriging model has to be built with too many support points, in order to achieve acceptable accuracy.

Such an insufficiently refined kriging model though, can provide an approximation of the optimal instrumental probability density function, as described in chapter 1.4. The Meta-Model Based Importance Sampling utilizes the sub-optimal PDF, provided by the kriging model, to sample values of $g(\mathbf{x})$ and reach low variance of the failure probability, with very few model evaluations. Therefore, this hybrid reliability analysis method, consists of two parts: Firstly, the construction of a kriging surrogate model, and secondly an importance sampling application on the sub-optimal instrumental PDF.

In this subchapter, the Meta-Model Based Importance Sampling method is presented, and illustrated on the example of subchapter 1.3.

2.6.2 Implementation

As mentioned in the 1.4 Importance Sampling subchapter, the optimal instrumental PDF that achieves $\delta = 0$, independently of sample size N, is the following:

$$h^*(\mathbf{x}) = \frac{I_F(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_f}$$

which depends on the unknown probability p_f. In meta-model based importance sampling, the optimal PDF $h^*(\mathbf{x})$ is approximated by the sub-optimal PDF $\hat{h}^*(\mathbf{x})$:

$$\hat{h}^*(\mathbf{x}) = \frac{\pi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{f\varepsilon}} \tag{2. 37}$$

where $p_{f\varepsilon}$ is the failure probability, estimated based on a kriging surrogate model. Therefore:

$$p_f = \int_{\mathbb{X}} \frac{I_F(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{\hat{h}^*(\mathbf{x})} \hat{h}^*(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{X}} \frac{I_F(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{\pi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})} p_{f\varepsilon} \hat{h}^*(\mathbf{x}) d\mathbf{x} = p_{f\varepsilon} \int_{\mathbb{X}} \frac{I_F(\mathbf{x})}{\pi(\mathbf{x})} \hat{h}^*(\mathbf{x}) d\mathbf{x}$$

and the correction factor is defined as:

$$a_{corr} = \int_{\mathbb{X}} \frac{I_F(\mathbf{x})}{\pi(\mathbf{x})} \hat{h}^*(\mathbf{x}) d\mathbf{x} \tag{2. 38}$$

which corrects the first estimation $p_{f\varepsilon}$. The closer $\pi(\mathbf{x})$ is to $I_F(\mathbf{x})$, the closer a_{corr} is to unity. The coefficient a_{corr} is estimated from Monte Carlo sampling, with sample generated according to $\hat{h}^*(\mathbf{x})$ with a Markov Chain Monte Carlo sampler, such as the Modified Metropolis Hasting Algorithm:

Step 0: Set target δ, δ_ε

The coefficient of variation of the failure probability in meta-model based importance sampling is dependent on δ_ε of the simulation technique which is based on the kriging model, and δ_{corr} which is the coefficient of variation of a_{corr}:

$$\delta = \sqrt{{\delta_\varepsilon}^2 + {\delta_{corr}}^2 + {\delta_\varepsilon}^2 {\delta_{corr}}^2} \approx \sqrt{{\delta_\varepsilon}^2 + {\delta_{corr}}^2} \tag{2. 39}$$

Since δ_ε does not depend on evaluations of $g(\mathbf{x})$, it should be significantly low, e.g. $\delta_\varepsilon = 1\%$. Therefore, the targeted value of δ_{corr} is $\delta_{corr} = \sqrt{\delta^2 - {\delta_\varepsilon}^2}$

Step 1: Implementation of a simulation technique to obtain $p_{f\varepsilon}$

At this step, a simulation technique, such as the Monte Carlo Simulation or the Subset Simulation is implemented with a low targeted δ_ε, to estimate $p_{f\varepsilon}$. Also the samples generated are used to find initial seed for the sampling from PDF $\hat{h}^*(\mathbf{x})$ in the next step.

Step 2: Sampling from $\hat{h}^*(\mathbf{x})$, and estimation of a_{corr}

The MCS provides the sample $\{\mathbf{x}^{(k)}, k = 1,2, \dots N\}$, while the SS provides the sample $\{\mathbf{x}^{[s](k)}, k = 1,2, \dots N\}$ where s the last subset. In the respective case, an initial seed $\mathbf{x}_{seed}{}^{(k)}, k = 1,2, \dots K$, where K is the number of chains that generate samples. K should be

set to a relatively small value, $K = \{50, 100, 200\}$. The sample is generated from the following sub-algorithm:

1. Set $j = 1$
2. Choose randomly a $k^* \in \{1,2, \dots N\}$ with probability:
$$r_k = \frac{\pi_s\left(\mathbf{x}^{[s](k)}\right)}{\pi_{s-1}(\mathbf{x}^{[s](k)}) \cdot Np_{s|s-1}}$$
where $\pi_{s-1}\left(\mathbf{x}^{[s](k)}\right) = 1$, $p_{s|s-1} = p_{1|0} = p_{f\varepsilon}$ by definition, if $s = 1$
3. $\mathbf{x}_{seed}{}^{(j)} = \mathbf{x}^{(k^*)}$, set $j = j + 1$
4. If $j \leq K$ go to 2, else go to 5
5. Choose burn-in parameter $b = 20$ and thinning parameter $t = 10$
6. Perform b MMHA steps to acquire new corrected $\mathbf{x}_{seed}$, set $i = 1$
7. Perform t MMHA steps to acquire sample $\left\{\mathbf{x}^{((i-1)K+k)}, k = 1,2, \dots K\right\}$ from the last state of each chain
8. Calculate δ_{corr} and $N_{corr} = i \cdot K$. If δ_{corr} is not low enough, set $i = i + 1$ and go to 7. Else end.

The correction factor a_{corr} is estimated from:

$$a_{corr} = \frac{1}{N_{corr}} \sum_{i=1}^{N_{corr}} \frac{I_F\left(\mathbf{x}^{(i)}\right)}{\pi(\mathbf{x}^{(i)})} \tag{2. 40}$$

and the coefficient δ_{corr} is calculated from the following expression:

$$\delta_{corr} = \frac{1}{a_{corr}} \sqrt{\frac{1}{N_{corr}} \left[\frac{1}{N_{corr}} \sum_{i=1}^{N_{corr}} \frac{I_F(\mathbf{x}^{(i)})^2}{\pi(\mathbf{x}^{(i)})^2} - {a_{corr}}^2 \right] (1 + \gamma_{corr})} \tag{2. 41}$$

Burn-in is very important because identical samples are dealt with. Thinning is also crucial, because it decreases the γ_{corr} coefficient to very low values. That way, the correlation that is innate in Markov Chains is eliminated. The last state of the thinning steps, is the sample added to the whole set of samples. Therefore, every t MMHA steps, K new samples are obtained, and δ_{corr} is revaluated. If the new δ_{corr} is low enough, the MIS ends.

It is worth noticing that sampling from $\hat{h}^*(\mathbf{x})$, evaluations of $g(\mathbf{x})$ are not required. Therefore, the thinning procedure is computationally efficient, and the generated sample of size N_{corr} is small and efficient.

2.6.3 Example

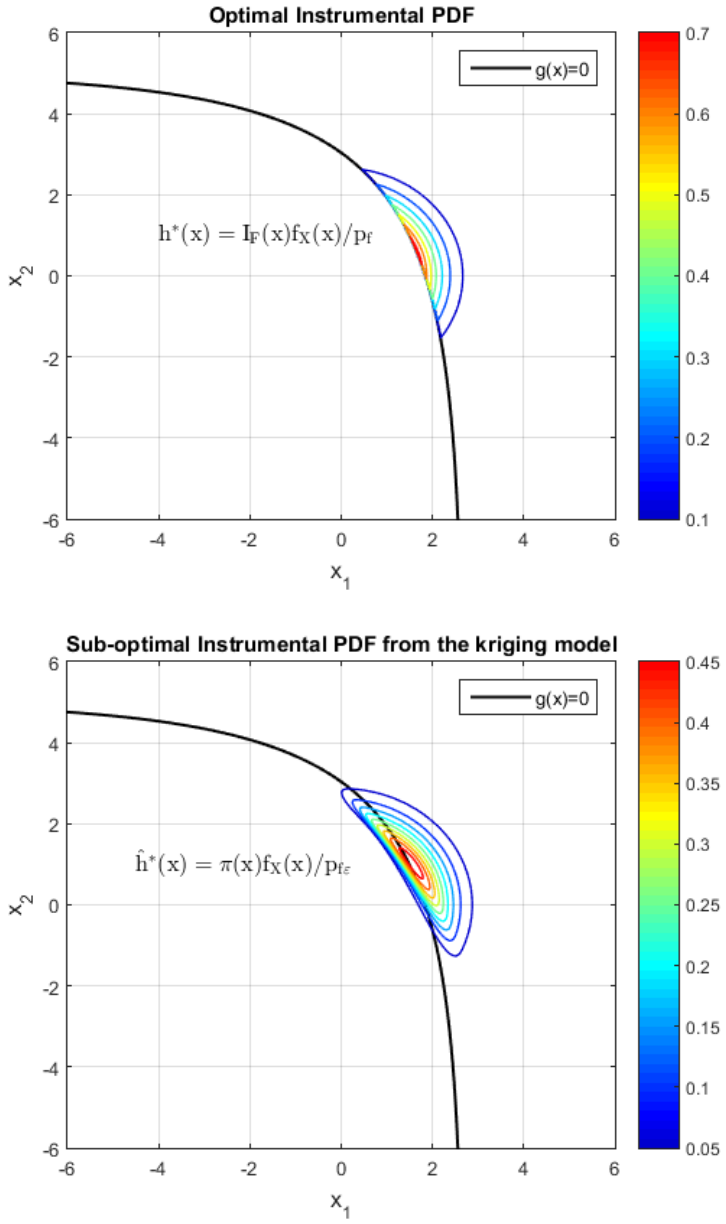
In the following example of subchapter 1.3, the kriging model is deliberately built insufficiently, on only 11 observations (new ones), to illustrate the application of MIS.

Firstly, the failure probability $p_{f\varepsilon}$ is estimated, with Monte Carlo Simulation on the kriging model. The targeted δ_ε is 0.9% and the size sample $N = 200000$.

$$p_{f\varepsilon} = 5.17 \cdot 10^{-2}$$

The above estimation is sufficient in terms of Monte Carlo Sampling, but the kriging model is not accurate. The actual failure probability is $p_f = 4.62 \cdot 10^{-2}$, for reference. The burn-in and the thinning parameters are set to $b = 20$ and $t = 20$. The initial seed is chosen $K = 100$.

In the figures below, the optimal PDF and the sub-optimal approximated PDF from the kriging model are illustrated:

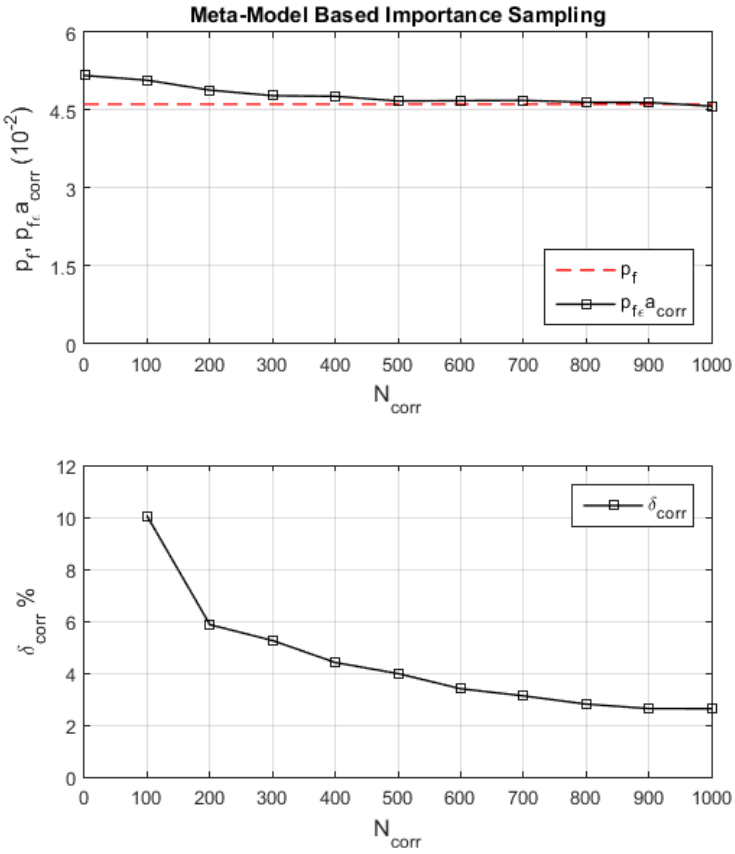


As shown above, the sub-optimal PDF $\hat{h}^*(\mathbf{x})$ is a relatively good approximation, of the optimal, yet impossible to use, $h^*(\mathbf{x})$. Considering that defining $\hat{h}^*(\mathbf{x})$ required only 11 observations, i.e. evaluations of $g(\mathbf{x})$, the MIS is proven to be very efficient. In the following graphs, the estimation of $p_{f\varepsilon} \cdot a_{corr}$ is displayed at each i step, as described in the algorithm of 1.6.2. In the horizontal axis, the total sample size N_{corr} is shown up to each i step. Assuming a targeted $\delta_{corr} = 4\%$, the MIS should stop at $i = 5 \Longrightarrow N_{corr} = 500$. The estimated failure probability is corrected:

$$p_{f\varepsilon} \cdot a_{corr} = 5.17 \cdot 10^{-2} \cdot 0.906 = 4.684 \cdot 10^{-2}$$

with coefficient of variation:

$$\delta = \sqrt{{\delta_{corr}}^2 + {\delta_{\varepsilon}}^2} = \sqrt{0.04^2 + 0.009^2} = 4.1\%$$



It is worth noticing, that the MIS required only $N_{MIS} = 11 + 500$ evaluations, while the Monte Carlo Simulation based on $g(\mathbf{x})$, would require $N_{MCS} = 13000$ for the same δ coefficient. Also, it is important to consider, that the sample size N_{corr} does not depend on the failure probability, in contrast to the MCS, but on the accuracy of the kriging model. For the purpose of this example, the particular kriging model is not refined. The refinement procedure, which is presented in the following chapter, can achieve such close approximations of the failure probability, in low dimensional problems, that estimating $p_{f\varepsilon}$ would be sufficient ($a_{corr} = 1$). Consequently, the MIS is necessary, only in cases that the refinement procedure requires a large number of observations. In such cases, which are usually high-dimensional problems, the kriging model is costly to build and the refinement's additional observations are more than the number of evaluations N_{corr}, required by the MIS.

3 Adaptive Refinement

3.1 Introduction

A kriging surrogate model is built from randomly chosen observations in the area of interest. The area of interest is determined from the reliability problem. For example, assuming a random vector $\mathbf{X} \in \mathbb{X} \equiv \mathbb{R}^{\mathrm{m}}$ with m independent components with normal distribution, the model is restricted in $\mathbb{X}^*$:

$$\mathbb{X}^* \equiv [-\mathrm{a} \cdot \sigma_1, \mathrm{a} \cdot \sigma_1] \times [-\mathrm{a} \cdot \sigma_2, \mathrm{a} \cdot \sigma_2] \times \dots \times [-\mathrm{a} \cdot \sigma_{\mathrm{m}}, \mathrm{a} \cdot \sigma_{\mathrm{m}}] \tag{3. 1}$$

where, $\sigma_{\mathrm{i}}, \mathrm{i} = 1,2, \dots \mathrm{m}$ the standard deviation of each component, and a the range in normalized space that affects greatly the measurement of a probability. The range of $\mathrm{a} = 8$, means that the kriging model won't provide reliable classification for any point that is further than 8 standard deviations from the mean value. That error is negligible and easily measurable, therefore it is ignored.

After defining the area $\mathbb{X}^*$, the kriging model is built only from observations than reside in the area of interest. The number of initial observations, n_0, depends partly on the dimensionality of the problem. It is proposed to set $\mathrm{n}_0 = \min(10; 2\mathrm{m})$. The initial observations, or dataset, $\mathfrak{D} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(\mathrm{n}_0)}\}$, are a Latin Hypercube sample in $\mathbb{X}^*$. Latin Hypercube Sampling (LHS) provides a uniform distribution of $\mathfrak{D}$, and large areas in $\mathbb{X}^*$ without observations are less likely to occur, compared to the Monte Carlo Sampling.

Depending only on randomly selected observations though, it is certain that the kriging model will be insufficient. Consequently, the dataset, should be enriched with new points, which should be chosen strategically. The classification of any point in $\mathbb{X}$, is the only information required in reliability analysis. In other words, a reliability simulation method relies in distinguishing the failure domain from the safe domain. Therefore a kriging model has to be enriched with new observations, in areas where the classification is not confident. Such "confidence" in classifying a point in $\mathbb{X}$**,** is expressed from the probabilistic classification function $\pi(\mathbf{x})$.

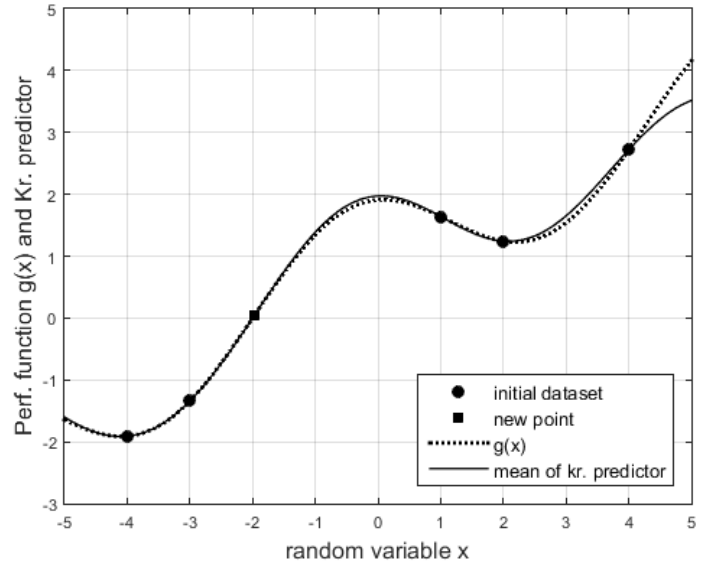
In the one-dimensional example of 2.2 the probabilistic classification function $\pi(\mathbf{x})$ is proven to be an unreliable approximation, since the estimated failure probability $\mathrm{p}_{\mathrm{f}\varepsilon}$ differs significantly from the actual failure probability p_{f}. Nevertheless, the $\pi(\mathbf{x})$ function can provide useful information to enrich the dataset by 1 additional observation and improve the estimation $\mathrm{p}_{\mathrm{f}\varepsilon}$.

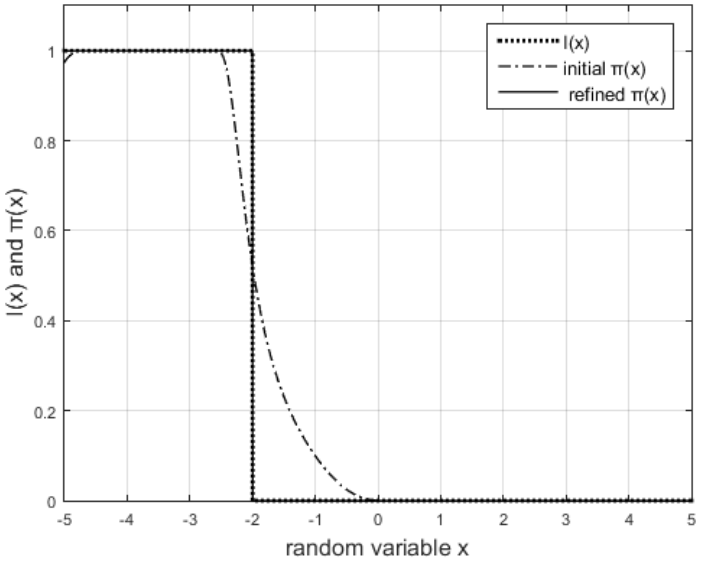
In this simple, one-dimensional problem the criterion of choosing the new point, is the following:

$$\mathrm{x} \in \mathbb{X}^* \equiv (-5,5): \mu_{\widehat{\mathrm{G}}}(\mathrm{x}) = 0 \Leftrightarrow \pi(\mathrm{x}) = 0.5$$

which is the most probable root of $\mathrm{g}(\mathrm{x}) = 0$ and distinguishes the failure domain (in this example $\mathrm{x} \leq -2$) from the safe domain ($\mathrm{x} > -2$). Therefore, by solving $\mu_{\widehat{\mathrm{G}}}(\mathrm{x}) = 0$ the new observation is acquired:

$$\mathrm{x}^{(6)} = -1.9769, \qquad \mathrm{g}\left(\mathrm{x}^{(6)}\right) = 0.035$$



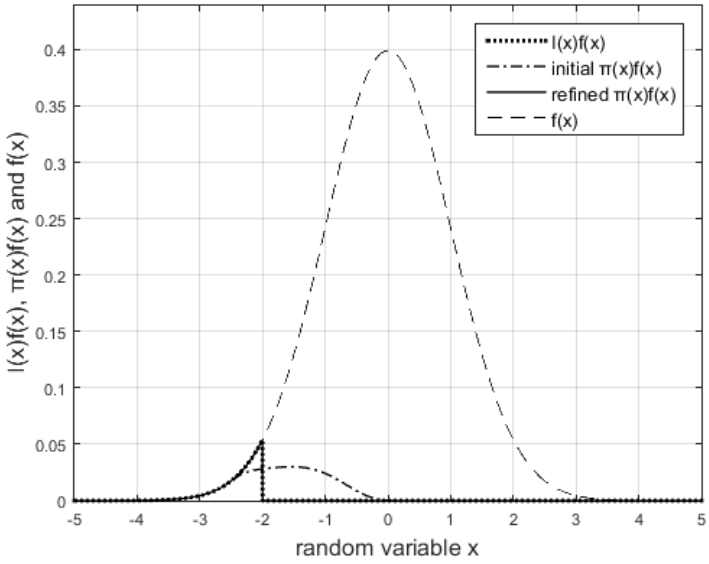


As shown in the figures above, the new point improves the $\pi(x)$ function significantly. The kriging model is refined and the estimation of the failure probability is accurate:

$$g(\mathbf{x}) \qquad \widehat{G}(\mathbf{x}) \sim \mathcal{N}\left(\hat{\mu}_G(\mathbf{x}), \hat{\sigma}_G{}^2(\mathbf{x})\right)$$

$$p_f = 2.275 \cdot 10^{-2} \qquad p_{f\varepsilon} = 2.276 \cdot 10^{-2}$$

The $\pi(x)$ function, as well as the $\pi(x)f_X(x)$, appear to be identical to $I_F(x)$ and $I_F(x)f_X(x)$ respectively.



It is worth noticing, that the failure probability is accurately estimated, with only $5+1$ evaluations of the performance function $\mathrm{g}(\mathrm{x})$. However, in problems with more than 1 dimensions, the additional observations to enrich the dataset are harder to find, and a simple criterion, such as the one used here, is not efficient. In the rest of the chapter, the most efficient strategies of adaptive refinement are presented. Those strategies, mainly focus on two objectives: locating the limit $\mathrm{g}(\mathbf{x})=0$ and prioritizing areas with high values of $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$.

3.2 Refinement Criteria

The following refinement criteria aim to identify the $\mathrm{g}(\mathbf{x})=0$ limit, by highlighting the areas with uncertainties in classifying a point $\mathbf{x} \in \mathbb{X}$. Consequently, the probabilistic classification function $\pi(\mathbf{x})$ has a prominent role.

3.2.1 The margin indicator function

The margin indicator function, relies on a chosen confidence level $1-\alpha$. It indicates whether a point $\mathbf{x} \in \mathbb{X}$ resides in the "grey area", as defined in 1.5.6.

$$\mathcal{C}(\mathbf{x})=\mathrm{I}_{1-\alpha}(\mathbf{x})=\begin{cases}1 & \alpha/2 \leq \pi(\mathbf{x}) \leq 1-\alpha/2 \\ 0 & \text{else}\end{cases} \tag{3. 2}$$

The confidence level could be set to $1-\alpha=95\%$. The second example of 1.5.6 is illustrated below. The $\pi(\mathbf{x})$ is also shown for reference.

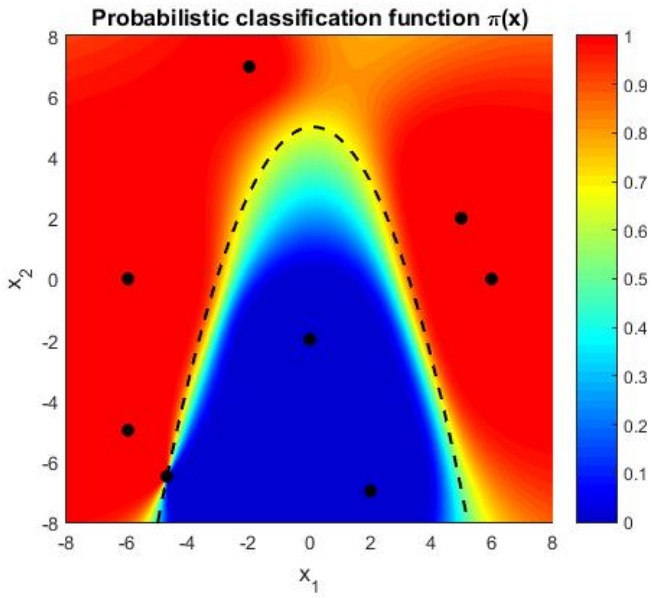
3.2.2 The margin probability function

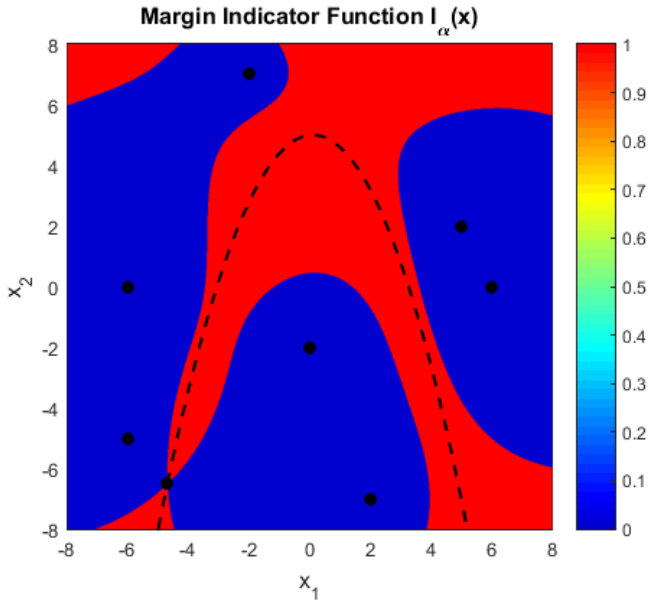
The margin probability function expresses the probability a point $\mathbf{x} \in \mathbb{X}$ to reside in the "grey area":

$$\mathcal{C}(\mathbf{x})=\mathrm{MP}(\mathbf{x})=\boldsymbol{\mathcal{P}}\left(\widehat{\mathrm{G}}(\mathbf{x}) \leq \mathrm{k}_{1-\alpha}\widehat{\sigma}_{\mathrm{G}}(\mathbf{x})\right)-\boldsymbol{\mathcal{P}}\left(\widehat{\mathrm{G}}(\mathbf{x}) \leq -\mathrm{k}_{1-\alpha}\widehat{\sigma}_{\mathrm{G}}(\mathbf{x})\right)=$$

$$=\Phi\left(\frac{\mathrm{k}_{1-\alpha}\widehat{\sigma}_{\mathrm{G}}(\mathbf{x})-\hat{\mu}_{\mathrm{G}}(\mathbf{x})}{\widehat{\sigma}_{\mathrm{G}}(\mathbf{x})}\right)-\Phi\left(\frac{-\mathrm{k}_{1-\alpha}\widehat{\sigma}_{\mathrm{G}}(\mathbf{x})-\hat{\mu}_{\mathrm{G}}(\mathbf{x})}{\widehat{\sigma}_{\mathrm{G}}(\mathbf{x})}\right) \tag{3. 3}$$

where $k_{1-\alpha} = \Phi^{-1}(1-\alpha/2)$. For a confidence level $1-\alpha = 95\% \implies k_{95\%} = 1.96$.





3.2.3 The function of most probable misclassification

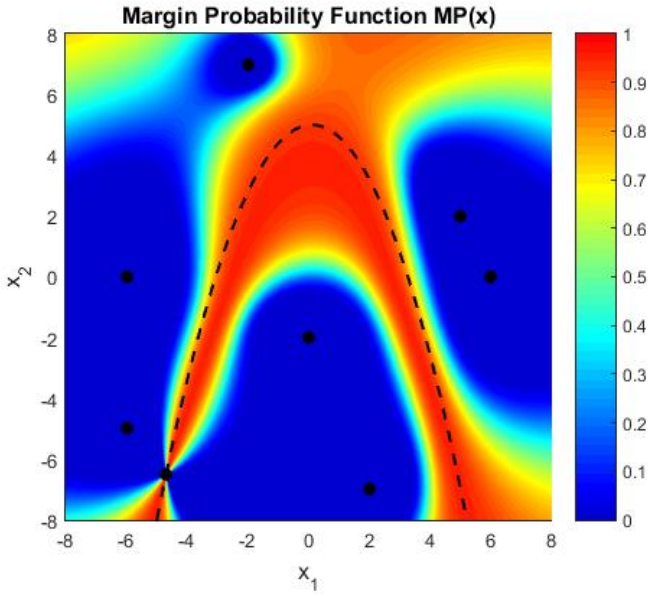
The most probable misclassification function expresses the probability that $\widehat{G}(\mathbf{x})$ is classified in the complementary domain to the one that $\hat{\mu}_G(\mathbf{x})$ is classified. This probability is:

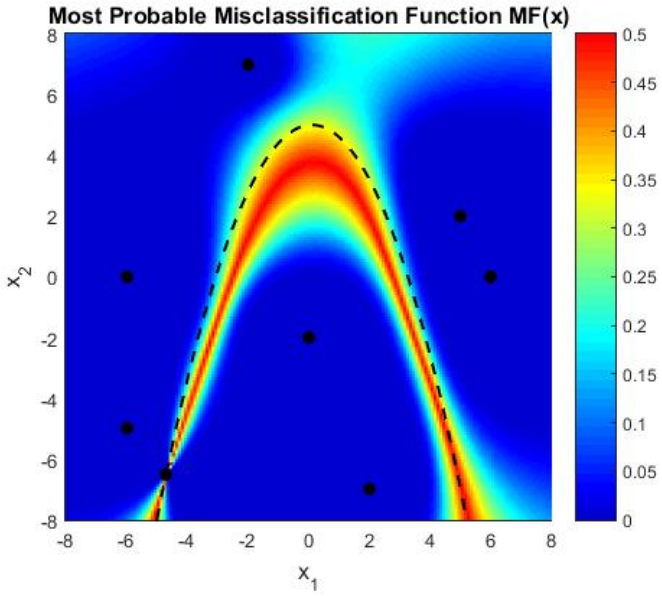
$$\min\left(\pi(\mathbf{x}); 1-\pi(\mathbf{x})\right) = \Phi\left(-\frac{|\hat{\mu}_G(\mathbf{x})|}{\hat{\sigma}_G(\mathbf{x})}\right)$$

The criterion is therefore defined as:

$$\mathcal{C}(\mathbf{x}) = \mathrm{MF}(\mathbf{x}) = \Phi\left(-\frac{|\hat{\mu}_{\mathrm{G}}(\mathbf{x})|}{\hat{\sigma}_{\mathrm{G}}(\mathbf{x})}\right) \tag{3. 4}$$

The function takes values from 0 to 0.5. The $\mathrm{MF}(\mathbf{x})$ is a more strict criterion, compared to the $\mathrm{MP}(\mathbf{x})$. To the author's experience, both are equally efficient, and most preferred in this thesis.





3.3 Sampling-based Adaptive Refinement

3.3.1 Principle

In reliability analysis, the objective is to solve the integral:

$$\mathrm{p_f} = \int_{\mathbb{X}} \pi(\mathbf{x}) \mathrm{f}_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}$$

From a different perspective, $\pi(\mathbf{x})$ is integrated in $\mathbb{X}$ "weighted" by $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$. Consequently, the refinement criterion has to be weighted by $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$, in order to take into account, that certain areas of the $\mathbb{X}$ domain are more important and should be predicted accurately from the kriging model. The refinement criterion is updated:

$$\mathcal{C}^*(\mathbf{x}) = \mathcal{C}(\mathbf{x}) \cdot \mathrm{f}_{\mathbf{X}}(\mathbf{x}) \tag{3. 5}$$

The $\mathcal{C}^*(\mathbf{x})$ function takes high values, close to $\mathrm{g}(\mathbf{x}) = 0$, and is proportionate to $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$. Therefore, expresses the uncertainty of the kriging model in terms of classification, and prioritizes areas that affect the most the measurement of the failure probability. However, finding new points to improve the kriging prediction, cannot be achieved from the optimization problem of $\max \mathcal{C}^*(\mathbf{x})$. In order to find an efficient new point, $\mathcal{C}^*(\mathbf{x})$ should be high-valued in the vicinity of the candidate point, because every support point affects the prediction in the near area. Consequently, $\mathcal{C}^*(\mathbf{x})$ is seen as a probability density function of random vector $\mathbf{C}$:

$$\mathrm{f}_{\mathbf{C}}(\mathbf{c}) \propto \mathcal{C}(\mathbf{c}) \cdot \mathrm{f}_{\mathbf{X}}(\mathbf{c}), \qquad \mathbf{c} \in \mathbb{X}^* \tag{3. 6}$$

A sample from $\mathrm{f}_{\mathbf{C}}(\mathbf{c})$, can be a list of candidate points, to add to the kriging dataset and improve the predictor. However, this list of candidate points is very big, because $\mathrm{f}_{\mathbf{C}}(\mathbf{c})$ is, in general, multimodal and a Markov Chain Monte Carlo sampler requires a large seed. Therefore, only the most representative of those are selected, from solving a clustering problem.

The K-means clustering problem is utilized in this thesis, to choose new support points from the samples of $\mathrm{f}_{\mathbf{C}}(\mathbf{c})$. The K-means algorithm (Steinhaus 1956, MacQueen 1967, Lloyd 1982) and the K-means ++ algorithm (Vassilvitskii 2007), provide clusters of points with respective centers. From each cluster a new support point is derived.

3.3.2 Proposed Refinement Method

Sampling from $\mathrm{f}_{\mathbf{C}}(\mathbf{c})$ can be difficult, because it is a multimodal PDF, and the initial seed should be chosen carefully. In this thesis a different approach is presented, which does not require sampling from $\mathrm{f}_{\mathbf{C}}(\mathbf{c})$.

The refinement procedure is part of an iterative algorithm, which requires the estimation of the failure probability in each step, in order to observe the improvement of the kriging model. Therefore the refinement procedure is usually preceded by a simulation technique (MCS, SS). Using the sample generated from the simulation, weighted by $\mathcal{C}(\mathbf{x})$, the clustering problem is equivalent to using the sample from $\mathrm{f}_{\mathbf{C}}(\mathbf{c})$. More extensively:

- If preceded by Monte Carlo Simulation, then there is an available sample generated according to $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$:

  $$\mathbf{x}_{\mathrm{c}}{}^{(\mathrm{i})}, \qquad \mathrm{i} = 1, 2, \dots \mathrm{N}$$

  The K-means algorithm is performed with weights, given by:

  $$\mathrm{w}_{\mathrm{MC}}\left(\mathbf{x}_{\mathrm{c}}{}^{(\mathrm{i})}\right) = \mathcal{C}\left(\mathbf{x}_{\mathrm{c}}{}^{(\mathrm{i})}\right), \mathrm{i} = 1,2, \dots \mathrm{N} \tag{3. 7}$$

- If preceded by Subset Simulation, then there is an available sample generated according to $\mathrm{h}_{\mathrm{s}}(\mathbf{x})$:

  $$\mathbf{x}_{\mathrm{c}}{}^{[\mathrm{s}](\mathrm{i})}, \qquad \mathrm{i} = 1, 2, \dots \mathrm{N}$$

The K-means algorithm is performed with weights, given by:

$$w_{SS}\left(\mathbf{x}_c^{[s](i)}\right) = \frac{\mathcal{C}\left(\mathbf{x}_c^{[s](i)}\right)}{\pi_{s-1}\left(\mathbf{x}_c^{[s](i)}\right)}, i = 1,2,\dots N \tag{3. 8}$$

At this point, the clusters are available, and the new support points are selected. The new points are not the cluster centers, because they may correspond to low values of $\mathcal{C}^*(\mathbf{x})$. The closest point to the respective cluster center is chosen, with a restriction in weight:

$$\mathbf{x}^{(n+k)} = \arg_{\mathbf{x}_{cl}{}^{(k)(i)}} \min\left\|\mathbf{o}^{(k)} - \mathbf{x}_c{}^{(k)(i)}\right\|_2, i = 1,2,..N^{(k)}, w\left(\mathbf{x}_c{}^{(k)(i)}\right) \geq 0.5 \cdot w_{max}{}^{(k)}$$

$$\mathfrak{O}^* = \left\{\mathbf{x}^{(n+1)}, \mathbf{x}^{(n+2)}, \dots, \mathbf{x}^{(n+K)}\right\} \tag{3. 9}$$

where:

$\mathfrak{O}^*$ the collection of new support points
n the size of the existing dataset $\mathfrak{O}$
$k = 1,2,\dots K$ the clusters
K the number of new support points, and the number of clusters
$\mathbf{o}^{(k)}$ the center of each cluster
$\mathbf{x}_c{}^{(k)(i)}$ the samples of the clustering problem that belong to cluster k
$\|\cdot\|_2$ the distance in $\mathbb{R}^m$
$w_{max}{}^{(k)}$ the maximum weight of all points in cluster k
$N^{(k)}$ the size of cluster k

3.3.3 Stopping criterion of the refinement procedure

It is proposed that the following quantity is used as a criterion to indicate the confidence of the measured failure probability $p_{f\varepsilon}$ from the kriging predictor:

$$p_{MF} = \int_{\mathbb{X}} MF(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{3. 10}$$

which depends on the previously defined, most probable misclassification function.

The quantity p_{MF} can be estimated from the same sample, that the failure probability is estimated.

- In Monte Carlo Simulation:

$$\hat{p}_{MF} = \frac{1}{N}\sum_{i=1}^{N} MF\left(\mathbf{x}^{(i)}\right) \tag{3. 11}$$

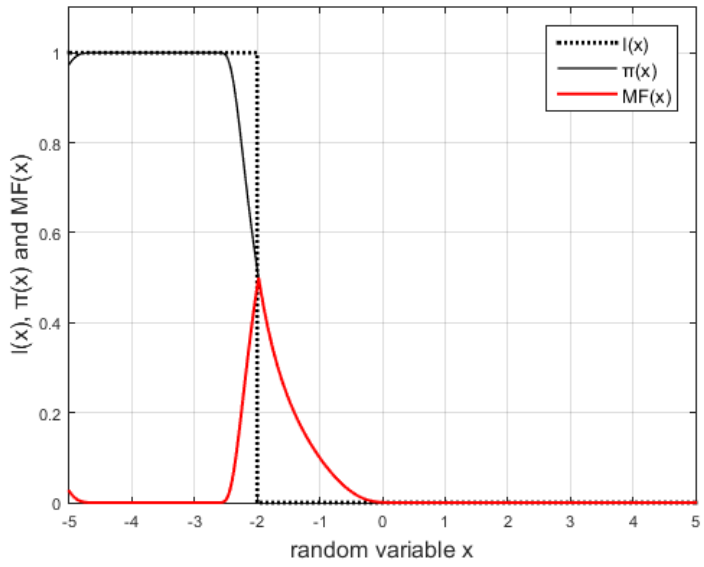
- In Subset Simulation:

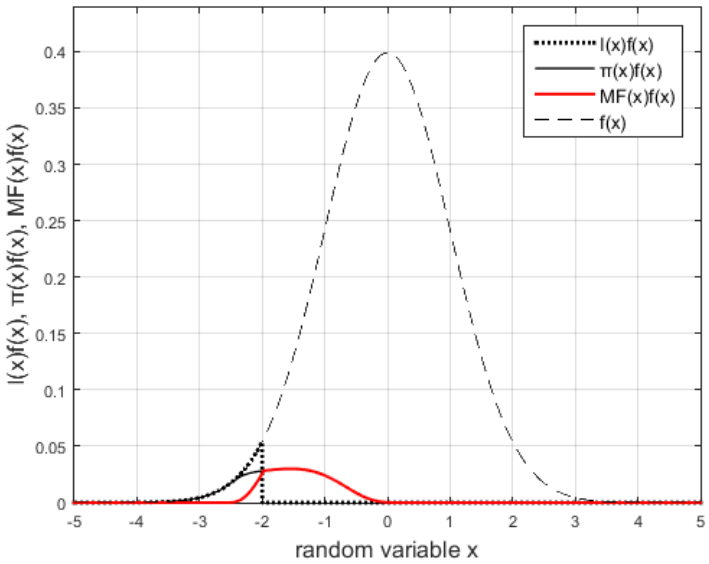
$$\hat{p}_{MF} = \frac{1}{N}\sum_{i=1}^{N} \frac{MF\left(\mathbf{x}^{[s](i)}\right)}{\pi_{s-1}\left(\mathbf{x}^{[s](i)}\right)} \cdot p_{s-1} \tag{3. 12}$$

It is always $p_{MF} < p_{f\varepsilon}$ and while $p_{MF}/p_{f\varepsilon}$ goes to zero, the kriging prediction is more confident. The refinement should stop when $p_{MF}/p_{f\varepsilon}$ is lower than a limit of $\ell_r = 1\%\sim5\%$. If $p_{MF}/p_{f\varepsilon}$ is not decreasing in several iterations, or the dataset is too big to add more support

points ($n > 1000$), the algorithm should resort to the Meta-model Based Importance Sampling.

In the example of 2.2 and 3.1, the $\hat{p}_{MF}$, is estimated before, and after the refinement. The following diagrams correspond to the initial kriging model:





Before the refinement, p_{MF} has a large value:

$$p_{MF} = 4.296 \cdot 10^{-2} \Longrightarrow p_{MF}/p_{f\varepsilon} = 75.3\%$$

but after the refinement:

$$p_{MF} = 0.005 \cdot 10^{-2} \Longrightarrow p_{MF}/p_{f\varepsilon} = 0\%$$

3.4 Complete Algorithm of Kriging Based Reliability Analysis

This subchapter presents the complete procedure that is partially described in chapters 1 through 3:

Step 0: Set target coefficients of variation δ_r, δ, initial support points n_0 and additional K per refinement iteration.

The coefficient δ is usually set to $\delta = 1\%\sim10\%$, and depends on the desirable accuracy of the estimated failure probability p_f. The δ_r coefficient defines the sample size of the simulation that measures the failure probability after each refinement iteration. For the reason that the same sample is used for the clustering, δ_r should be low enough, that there are many points with high weight value $w(\mathbf{x})$. A safe choice is to set $\delta_r = \delta$. However, at the first refinement iterations, δ_r can be lower than δ.

The number of initial support points n_0 is set to $n_0 = \max\big((2\sim4)\cdot m; 20\big)$, where m the dimensionality of the problem. This rule is not proven, but is based on the fact that more observations are needed to build a reliable kriging model, as dimensionality rises. Similarly, the additional observations per refinement iteration is set to $K = \max\big((1\sim2)\cdot m; 5\big)$.

Step 1: Build a kriging model on initial $n = n_0$ observations, $i = 0$.

The initial n_0 observations are sampled uniformly in the area of interest $\mathbb{X}^*$. The Latin Hypercube Sampling is proven to be very efficient, because the points are less likely to be close to each other, and more areas of $\mathbb{X}^*$ are covered.

It is worth mentioning, that in high-dimensional problems there is a probability that all of the n_0 observations reside exclusively on the safe or the failure domain. In this case, it might be helpful that the user provides manually a point in the domain with no observations, but that is not mandatory. For example, if in a structural reliability problem all initial observations reside in the failure domain, the vector of mean values could be an additional support point, since it represents a safe state.

At this point the kriging model is not expected to be accurate. However, it provides useful information to start distinguishing the safe and failure domains.

Step 2: Perform a simulation method on the kriging model. Estimate $p_{f\varepsilon}, p_{MF}$.

The Monte Carlo Simulation or the Subset Simulation estimate the model based failure probability $p_{f\varepsilon}$ and the p_{MF} which expresses the uncertainty of the model, with a targeted δ_r. The ratio $p_{MF}/p_{f\varepsilon}$ should be observed in every iteration i, as it provides the stopping criterion. It is proposed that even if $p_{MF}/p_{f\varepsilon}$ reaches values less than $\ell_r = 1\%$, the refinement should stop after $2\sim3$ iterations. In case $p_{MF}/p_{f\varepsilon}$ stabilizes in several iterations and $p_{MF}/p_{f\varepsilon} \leq \ell_r$ is not true, that means that any additional observations, do not improve the kriging model considerably. In this case the refinement also stops, and the algorithm continues to step 4. Otherwise, meaning $p_{MF}/p_{f\varepsilon}$ is decreasing, the algorithm continues to step 3.

Step 3: Perform weighted K-means clustering and add K new points to the dataset, $i = i + 1$

The sample of the simulation performed in step 2, is clustered with the K-means algorithm, weighted according to the $w(\mathbf{x})$ function. It is recommended, that the samples with low weights (e.g. $w(\mathbf{x}) < 0.02 \cdot w_{max}$) to be discarded, in order to facilitate the K-means

algorithm, and avoid local maxima. The new observations are acquired and added to the dataset: $\mathrm{n} = \mathrm{n} + \mathrm{K}$. The algorithm continues to step 2.

Step 4: Estimate the model based failure probability $\mathrm{p}_{\mathrm{f}\varepsilon}$ with targeted c.o.v δ_ε.

The final, model based failure probability $\mathrm{p}_{\mathrm{f}\varepsilon}$ is estimated with MCS or SS.

If $\mathrm{p}_{\mathrm{MF}}/\mathrm{p}_{\mathrm{f}\varepsilon}$ stabilized in a higher value than the limit ℓ_r, the algorithm resorts to Meta-model based Importance Sampling. Therefore, a low targeted e.g. $\delta_\varepsilon = 1\%$ is set. Else, set $\delta_\varepsilon = \delta$ and end at step 4.

Step 5: Meta-model based Importance Sampling

The MIS is performed with targeted $\delta_{\mathrm{corr}} = \sqrt{\delta^2 - \delta_\varepsilon^2}$. The $\mathrm{a}_{\mathrm{corr}}$ is estimated, and the final corrected failure probability is $\mathrm{p}_{\mathrm{f}} = \mathrm{a}_{\mathrm{corr}} \cdot \mathrm{p}_{\mathrm{f}\varepsilon}$.

The last step can also be considered, as a verification of the model based estimation. This means that MIS is useful even if the ratio $\mathrm{p}_{\mathrm{MF}}/\mathrm{p}_{\mathrm{f}\varepsilon}$ is zero.

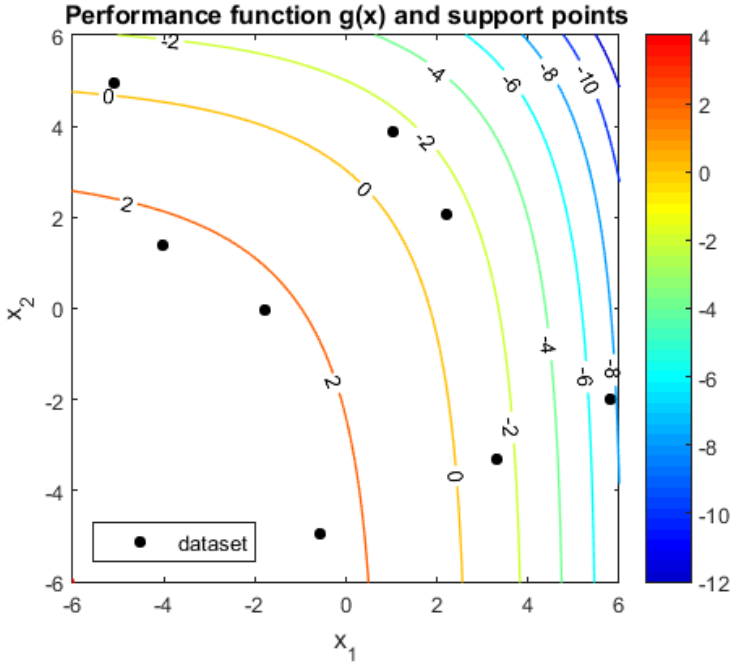
3.5 Refinement Example

A kriging model is built on the performance function of example 1.3:

$$\mathrm{g}(\mathbf{x}) = 4.5 - \mathrm{e}^{0.3\cdot \mathrm{x}_1 + 0.7} - \mathrm{e}^{0.3\cdot \mathrm{x}_2}, \qquad \mathrm{x}_1, \mathrm{x}_2 \in \mathbb{R}^2$$

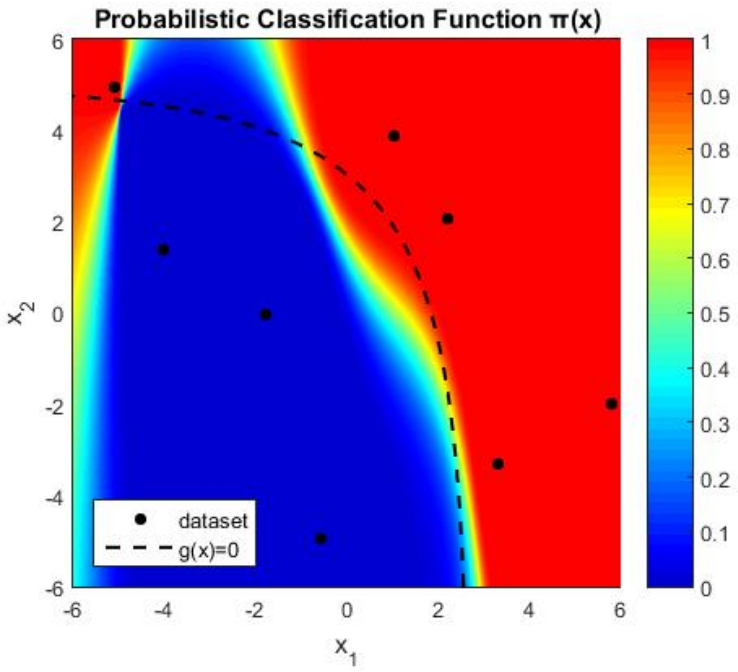
where, $\mathrm{x}_1, \mathrm{x}_2$ normally distributed random variables, with zero mean and unit variance

The regressor is constant and the autocorrelation function is the squared exponential. The initial support points for the kriging model is $\mathrm{n}_0 = 8$. Those are uniformly distributed (with LHS) in the area of interest $\mathbb{X}^* \equiv [-6, 6] \times [-6, 6]$.



The probabilistic classification function $\pi(\mathbf{x})$ of the initial kriging model is displayed below. The transition from the red area (probable failure) to the blue area (probable safety) does not

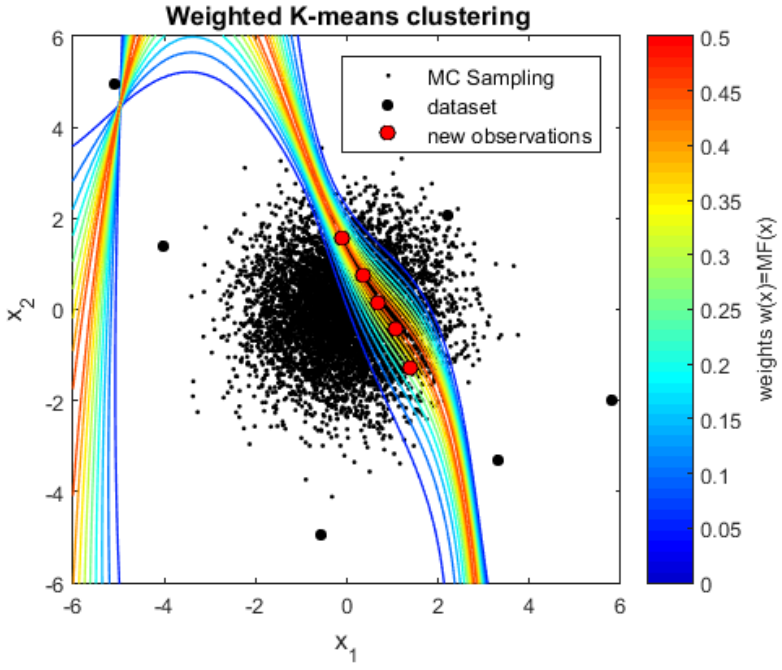
correspond with the dashed line, which is the real $\mathrm{g}(\mathbf{x}) = 0$ limit:



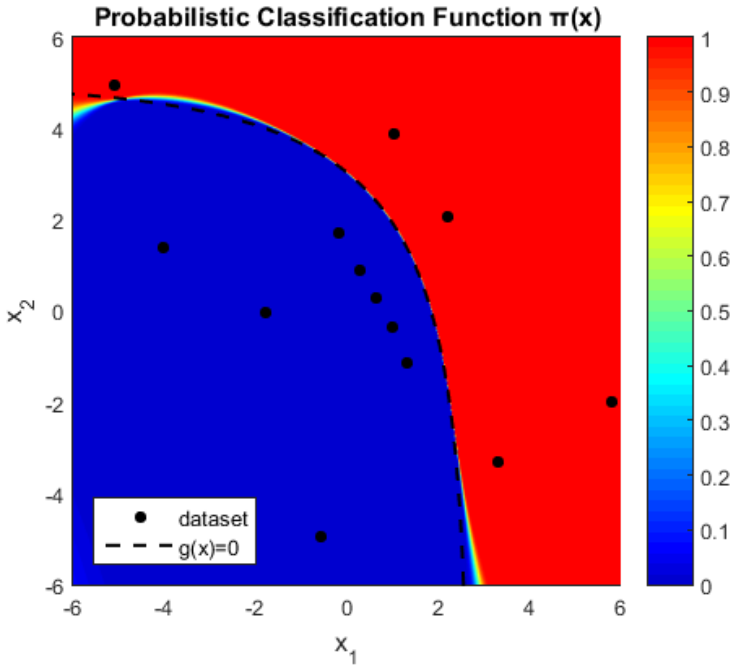
The failure probability is estimated on the $\pi(\mathbf{x})$ function. For the purpose of this example, the failure probability is numerically integrated, to avoid confusion between the error of the simulation method and that of the kriging model:

$$\mathrm{p}_{\mathrm{f}\varepsilon} = 0.2473, \qquad \mathrm{p}_{\mathrm{MF}}/\mathrm{p}_{\mathrm{f}\varepsilon} = 42.6\%$$

For reference, the actual failure probability is $\mathrm{p}_{\mathrm{f}} = 0.0462$. Clearly, the $\mathrm{p}_{\mathrm{MF}}/\mathrm{p}_{\mathrm{f}\varepsilon}$ is too high and, therefore, the kriging model is not considered accurate. For the refinement of the model, a sample of $\mathrm{N} = 100000$ according to $\mathrm{f}_{\mathbf{X}}(\mathbf{x})$ is generated, and the clustering problem is solved with $\mathrm{K} = 5$. In the following figure, a part of the MC sample, along with the weight function $\mathrm{w}(\mathbf{x}) = \mathrm{MF}(\mathbf{x})$ are illustrated. Also, the resulting new support points are shown as red dots:



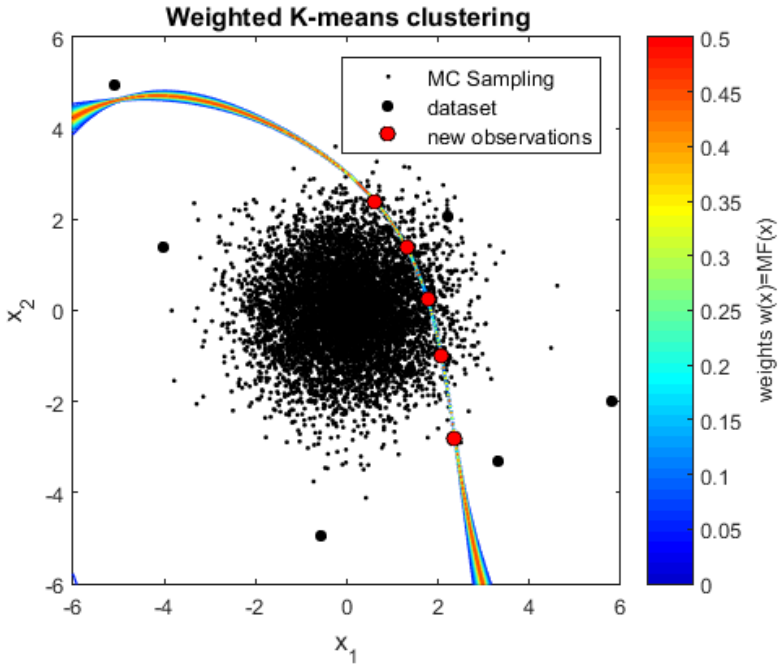
The new support points are added to the dataset and the kriging model is built again. The refined probabilistic classification function is shown below:



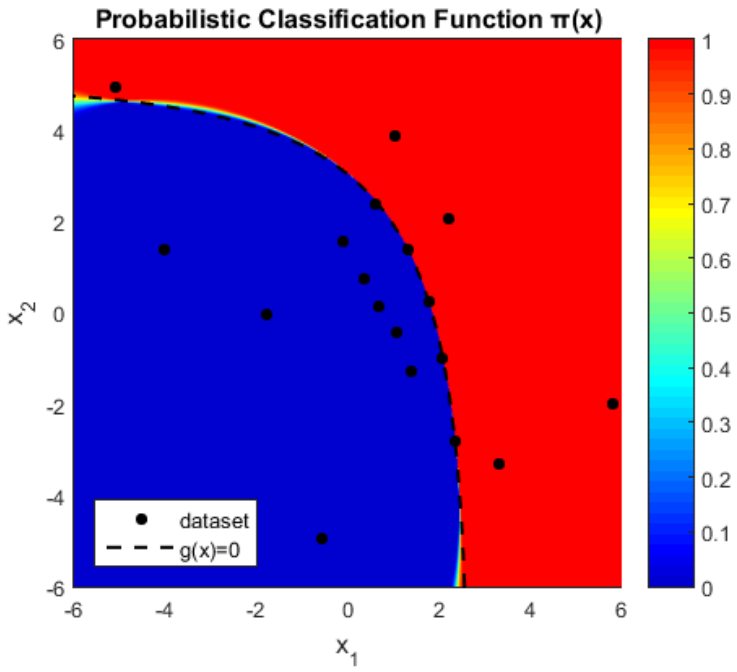
The new points are close to the actual $g(\mathbf{x}) = 0$ curve, and form a parallel line. As a consequence, the variance of the model in the vicinity of the new points has dropped greatly, and the kriging prediction is more accurate:

$$p_{f\varepsilon} = 0.0468, \qquad p_{MF}/p_{f\varepsilon} = 0.84\%$$

However, the refinement continues for one last iteration:



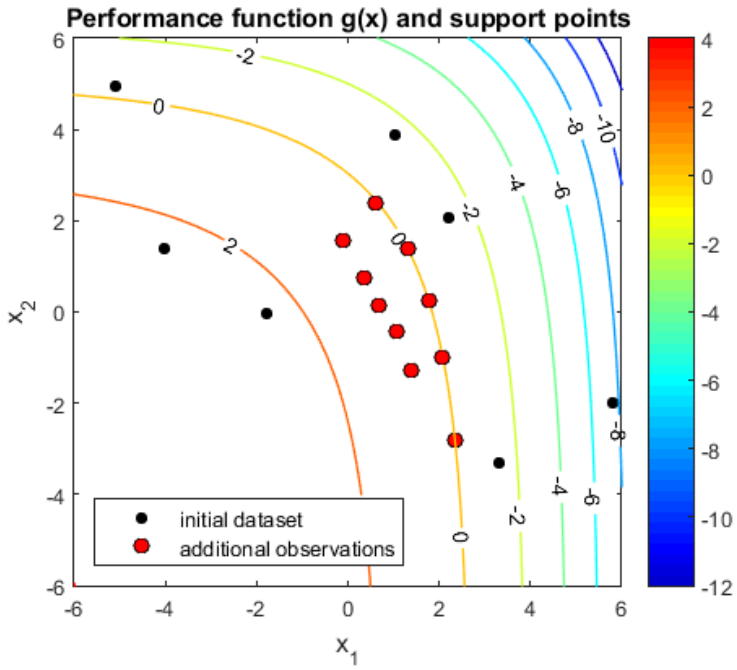
Visually, the $\pi(\mathbf{x})$ function has not changed considerably:



The model based failure probability is measured:

$$p_{f\varepsilon} = 0.0462, \qquad p_{MF}/p_{f\varepsilon} = 0.01\%$$

which by at least 3 significant figures, is equal to the actual one $p_f = 0.0462$. Lastly, the final dataset is illustrated on the real $g(\mathbf{x})$ performance function:



In conclusion, the kriging model based reliability analysis required only $8 + 10 = 18$ evaluations of the performance function $g(\mathbf{x})$, while a simulation technique based on $g(\mathbf{x})$ would require more than 10000 for the same result. In lower failure probabilities than the one of the example, the required support points to build a model does not rise, in contrast to the

evaluations of a simulation technique. Therefore, the benefit of substituting the performance function with a surrogate model is achieved. In the next chapter, applications of higher dimensional problems are presented. Finally, the above example is run multiple times, and the results of each independent run are gathered in the following table. The relevant error, of the measured failure probability to the actual one, is also shown: $e_p = |p_f - p_{f\varepsilon}|/p_f$, where $p_f = 0.0462$.

		refinement iteration		
		$i = 0, n = 8$	$i = 1, n = 13$	$i = 2, n = 18$
1	$p_{f\varepsilon}$	0.0755	0.0465	0.0462
	$p_{MF}/p_{f\varepsilon}$	22.38%	0.97%	<0.01%
	e_p	63.42%	0.65%	<0.01%
2	$p_{f\varepsilon}$	0.4034	0.0443	0.0462
	$p_{MF}/p_{f\varepsilon}$	50.94%	4.06%	<0.01%
	e_p	773.2%	4.11%	<0.01%
3	$p_{f\varepsilon}$	0.0834	0.0464	0.0462
	$p_{MF}/p_{f\varepsilon}$	54.44%	0.24%	<0.01%
	e_p	80.52%	0.43%	<0.01%
4	$p_{f\varepsilon}$	0.1091	0.0465	0.0462
	$p_{MF}/p_{f\varepsilon}$	39.32%	1.13%	<0.01%
	e_p	136.0%	0.65%	<0.01%
5	$p_{f\varepsilon}$	0.0323	0.0461	0.0462
	$p_{MF}/p_{f\varepsilon}$	25.08%	0.76%	<0.01%
	e_p	30.09%	0.22%	<0.01%
6	$p_{f\varepsilon}$	0.0459	0.0463	0.0462
	$p_{MF}/p_{f\varepsilon}$	30.28%	0.33%	<0.01%
	e_p	0.65%	0.22%	<0.01%
7	$p_{f\varepsilon}$	0.0353	0.0468	0.0462
	$p_{MF}/p_{f\varepsilon}$	28.61%	0.26%	<0.01%
	e_p	23.59%	1.30%	<0.01%

It is worth noticing that, at the first estimation $(i = 0)$, the estimated failure probability is most of the time higher than the actual one. This is quite possible to happen in structural reliability problems, because the area near the mean values resides in the safe domain. This is more obvious in the 2.2 example, where the error of $\pi(\mathbf{x})$ is enlarged by the $f_{\mathbf{X}}(\mathbf{x})$. Also, in the 6th run, $p_{f\varepsilon}$ is almost equal to the actual p_f, despite that $p_{MF}/p_{f\varepsilon}$ is large. This happened because the misclassification of the areas in the safe and failure domain negated completely each other by coincidence. In this example the iterative procedure converges rapidly, because 2-dimensional problems are generally easy to handle. However, in applications of the next chapter, it is shown that the high-dimensional problems require many more iterations and additional support points, in order an accurate solution to be achieved.

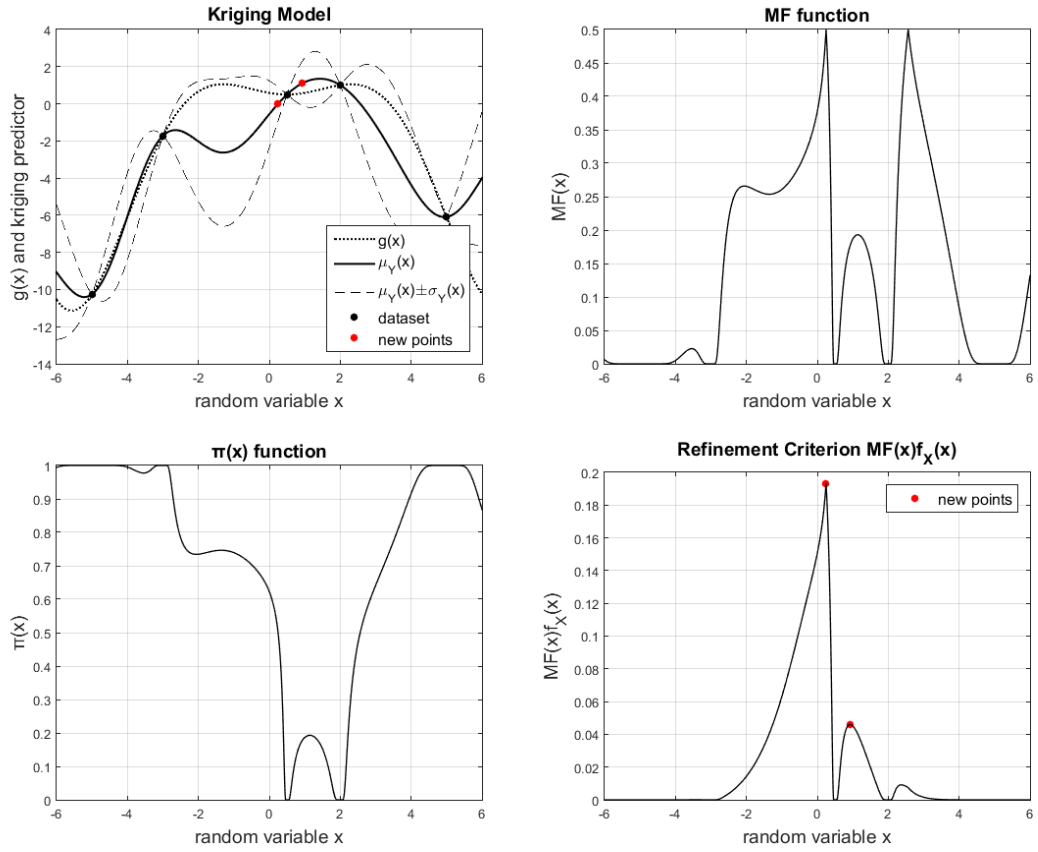
4 Applications

The applications of kriging based reliability analysis presented in this chapter, aim to educate the reader on the process and validate the method, in problems with many dimensions and multiple design points. The first example, is another one-dimensional example, with purpose to illustrate the role of the probabilistic classification function. The second example is a 2-dimensional problem, with a piecewise performance function, which causes difficulties in the refinement of the kriging model. In the third example, a benchmark problem is solved in 5, 10, 50 and 80 dimensions. The final two problems are structural, the one being a truss bridge with a restriction to the deformation. The last is a tall structure modeled with the assumption of a shear and flexural cantilever beam, which deforms according to a statistical earthquake spectrum by Boore and Atkinson.

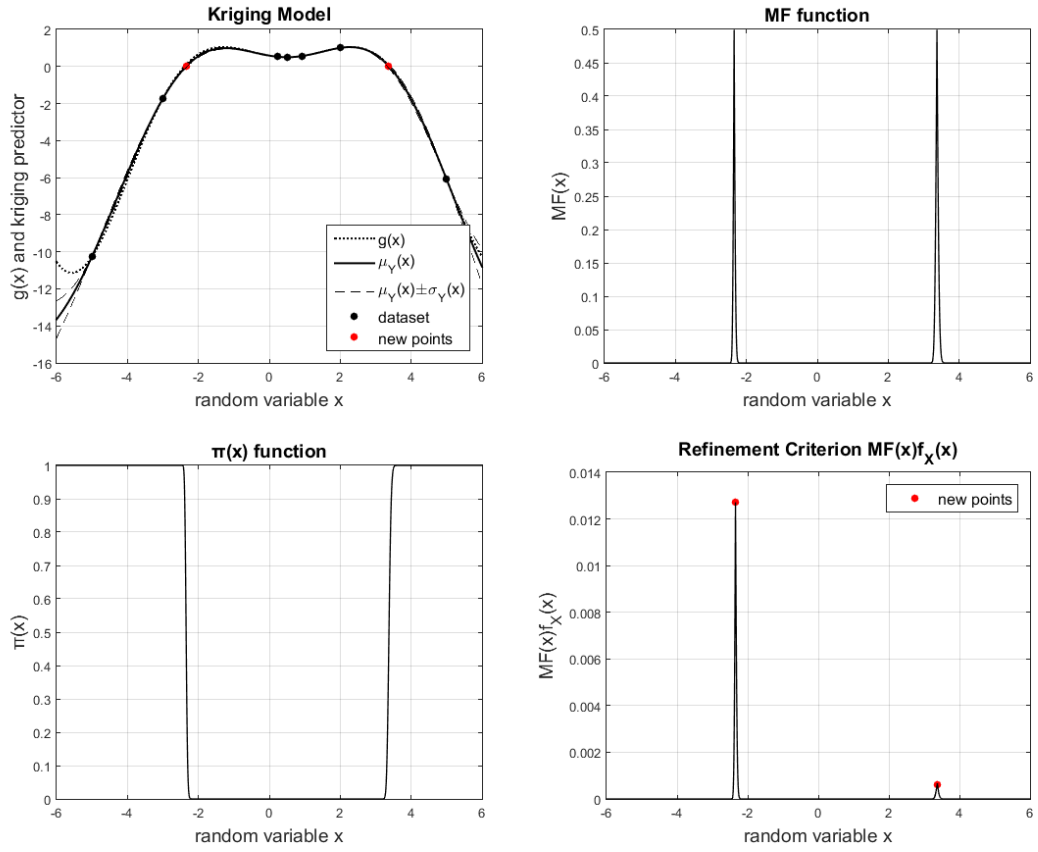
4.1 One-dimensional problem

The following one-dimensional problem is sampled at 5 points to build the kriging model. The random variable is $\mathbf{X} \in \mathbb{X} \equiv \mathbb{R}$ and is normally distributed with zero mean and unit variance:

$$g(x) = (0.6x - 0.3)^2 \cos(0.6x - 0.3) + 0.5$$



As shown above, the initial pre-refinement model fails to provide accurate classification. In the 4th graph the refinement criterion depicts the best points to enrich the dataset. Two local maxima are chosen, and the kriging model is reevaluated:



The second refinement is the last one, and the failure probability is correctly evaluated, with numerical integration:

	refinement iteration		
	$i = 0, n = 5$	$i = 1, n = 7$	$i = 2, n = 9$
$p_{f\varepsilon}$	0.4769	0.0099	0.0086
$p_{MF}/p_{f\varepsilon}$	52.57%	7.43%	0.04%
e_p	54.16%	14.92%	0.04%

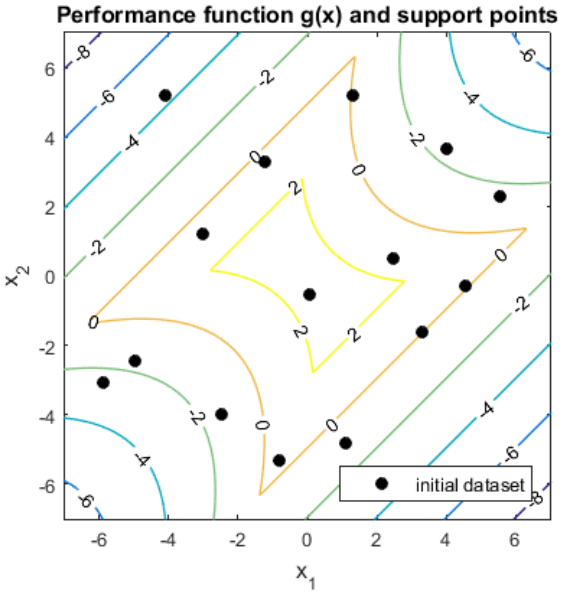
where, $e_p = |p_f - p_{f\varepsilon}|/p_f$, and the actual failure probability $p_f = 0.0086$.

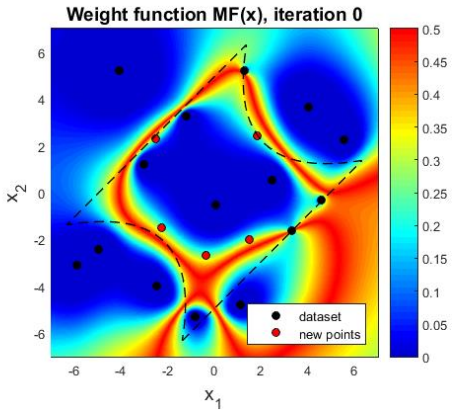
4.2 2-dimensional, piecewise performance function

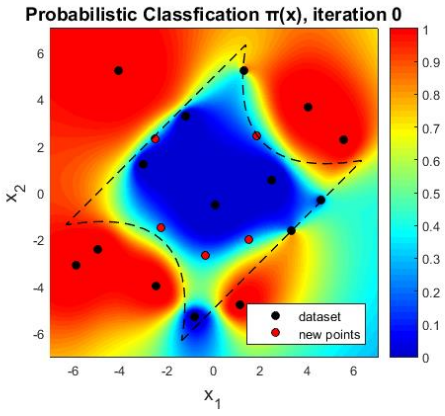
This function is difficult to emulate, because the limit $g(\mathbf{x}) = 0$ surrounds the mean values and many areas of the failure domain $\mathbb{F} \subset \mathbb{X} \equiv \mathbb{R}^2$ have to be accurately predicted. The random vector $\mathbf{X}$ consists of 2 standard normal variables:

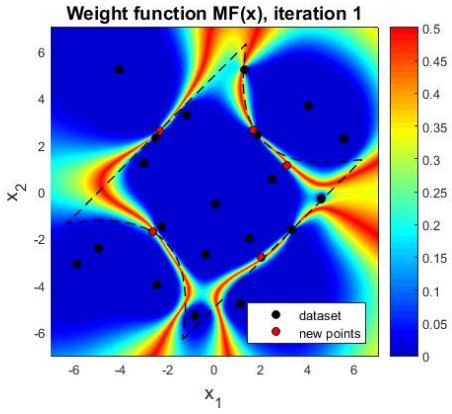
$$g(\mathbf{x}) = \min\left\{\begin{matrix} 3 + \frac{(x_1 - x_2)^2}{10} - \frac{x_1 + x_2}{\sqrt{2}} \\ 3 + \frac{(x_1 - x_2)^2}{10} + \frac{x_1 + x_2}{\sqrt{2}} \\ x_1 - x_2 + \frac{7}{\sqrt{2}} \\ -x_1 + x_2 + \frac{7}{\sqrt{2}} \end{matrix}\right\}, \mathbf{x} \in \mathbb{R}^2$$

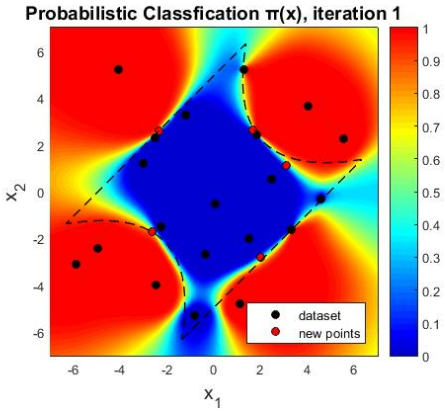
The initial dataset is size $n_0 = 15$ and the additional points per refinement iteration is $K = 5$. In iteration i the dataset consists of $n = n_0 + i \cdot K$ observations. In the following graphs the weight function $w(\mathbf{x}) = MF(\mathbf{x})$ and the probabilistic classification function are illustrated in iterations $i = \{0, 1, 2, 3, 4, 10\}$. The stopping criterion $p_{MF}/p_{f\varepsilon}$ takes low values before the 10th refinement, but the algorithm continues for the purpose of this example. In the last table the values of the estimated failure probability $p_{f\varepsilon}$ and $p_{MF}/p_{f\varepsilon}$ are shown at each iteration, with targeted $\delta_\varepsilon = 1.5\%$. For reference the failure probability, estimated on the actual performance function is $p_f = 2.208 \cdot 10^{-3}$ with $\delta = 1\%$ and $N = 4.5 \cdot 10^6$ evaluations.

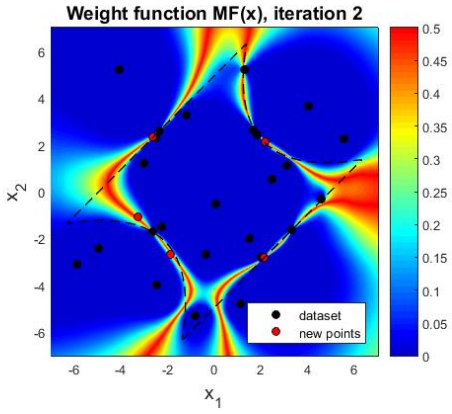


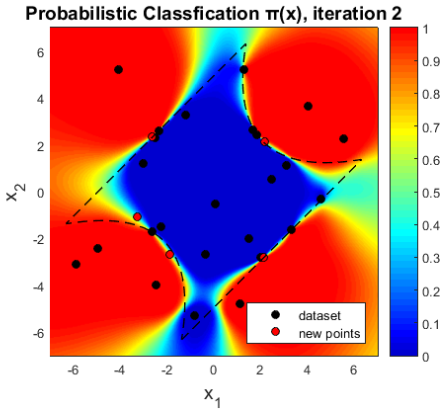


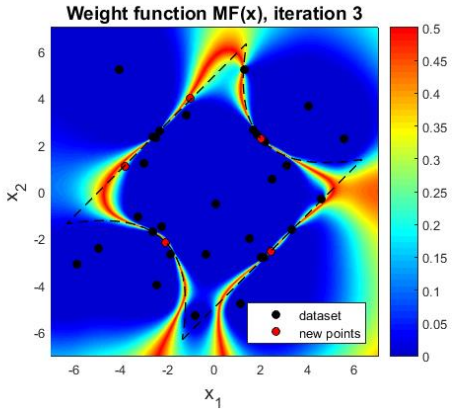


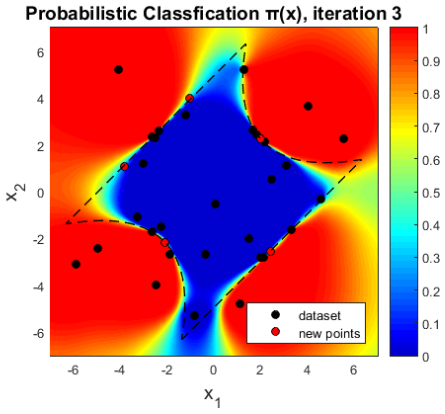

Weight function MF(x), iteration 1
dataset
new points
x1
x2

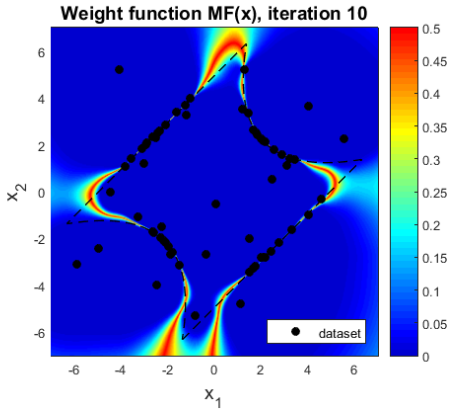

Probabilistic Classfication π(x), iteration 1
dataset
new points
x1
x2

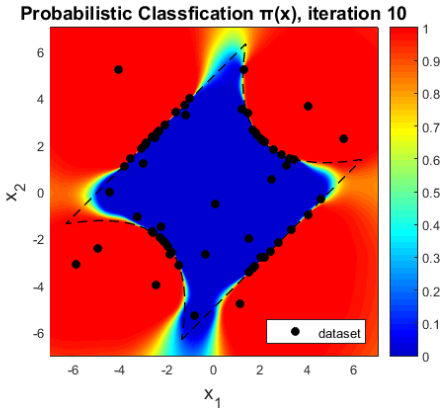

Weight function MF(x), iteration 2
dataset
new points
x1
x2

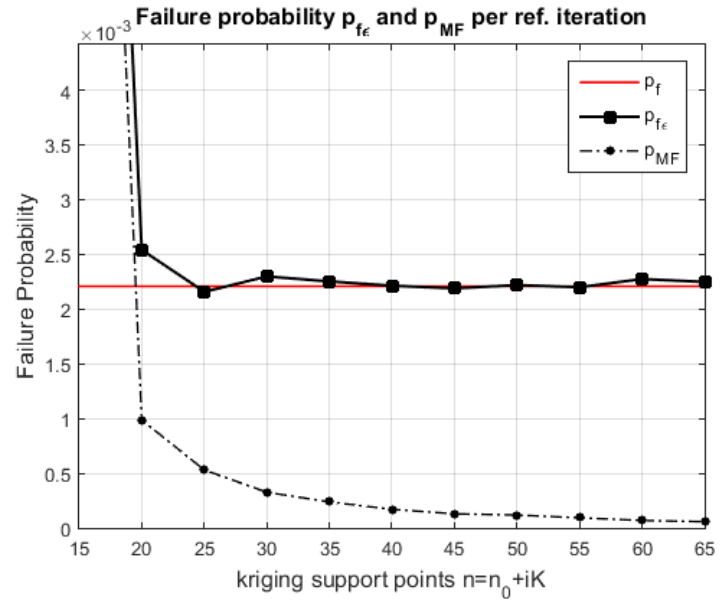

Probabilistic Classfication π(x), iteration 2
dataset
new points
x1
x2

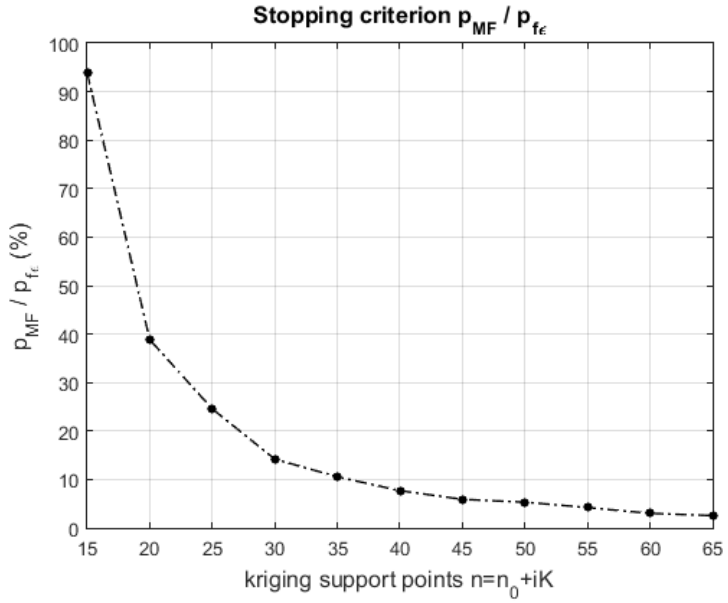

Weight function MF(x), iteration 3
dataset
new points
x1
x2


Probabilistic Classfication π(x), iteration 3
dataset
new points
x1
x2


Weight function MF(x), iteration 10
dataset
x1
x2


Probabilistic Classfication π(x), iteration 10
dataset
x1
x2





The horizontal axis of the last charts, shows the number of observations that were used to build the model and estimate the respective failure probability. Before the refinement, the estimated $p_{f\varepsilon}$ is off the chart, but after the first refinement, the estimation is very close to the actual one. At iteration $i = 5, n = 40$ the model is sufficient because the ratio $p_{MF}/p_{f\varepsilon}$ is low enough. Compared to the MCS on the real $g(\mathbf{x})$ which required $N = 4.5 \cdot 10^6$ evaluations, the kriging model approach required only $n = 40$.

4.3 High-dimensional problem

The following performance function by Rackwitz (2001), is adaptable, meaning the number of dimension is chosen freely. In this application, the kriging based reliability method is performed on $m = \{5, 10, 50\}$ dimensions.

$$g(\mathbf{x}) = m + 3\sigma_{\mathbf{X}}\sqrt{m} - \sum_{i=1}^{m} x_i$$

where vector $\mathbf{X}$ of m components with lognormal distribution, $\mu_{\mathbf{X}} = 1$, $\sigma_{\mathbf{X}} = 0.2$

The failure probability is estimated at each refinement iteration at a targeted $\delta_\varepsilon = 4.5\%$. When the refinement procedure is over, a Meta-model based Importance Sampling is performed, to verify or correct the kriging based $\mathrm{p_{f\varepsilon}}$.

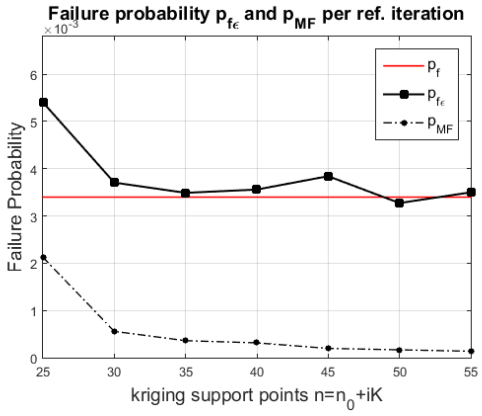
The actual failure probability with coefficient of variation $\delta = 1\%$, is estimated on the real $\mathrm{g}(\mathbf{x})$ with Monte Carlo Simulation. The respective model evaluations N are also displayed:

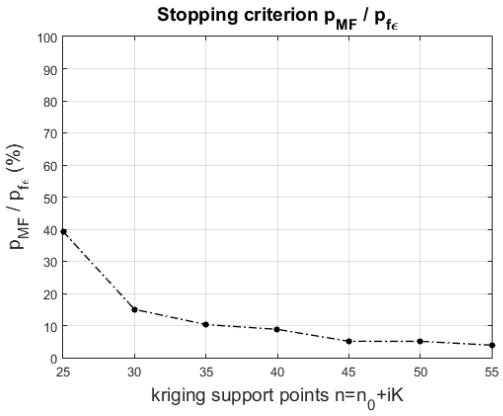
MSC on the real $\mathrm{g}(\mathbf{x})$

	$\mathrm{m} = 5$	$\mathrm{m} = 10$	$\mathrm{m} = 50$	$\mathrm{m} = 80$
$\mathrm{p_f}$	$3.399 \cdot 10^{-3}$	$2.736 \cdot 10^{-3}$	$1.899 \cdot 10^{-3}$	$1.771 \cdot 10^{-3}$
N	3,000,000	3,700,000	5,200,000	5,600,000

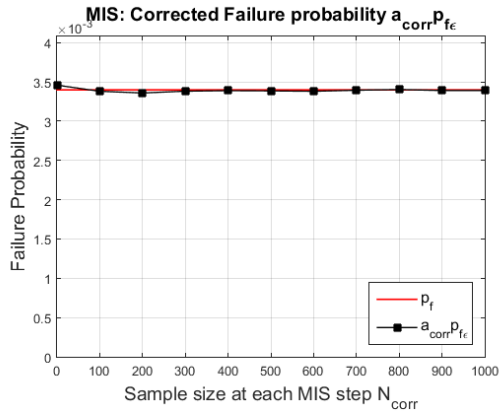
- $\mathrm{m} = 5$

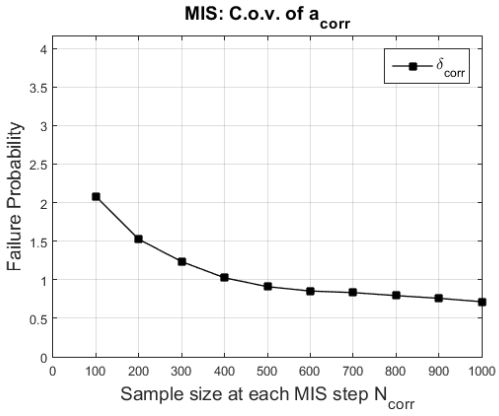
Starting with $\mathrm{n_0} = 25$ initial observations, the kriging model is refined with $\mathrm{K} = 5$ new observations at every step. It is reminded that each refinement step contains the statistical error of the simulation, which affects partly the change of $\mathrm{p_{f\varepsilon}}$.





Before initiating the Meta-model based Importance Sampling, the $\mathrm{p_{f\varepsilon}}$ is more accurately estimated with targeted $\delta_\varepsilon = 1\%$, to reduce the final δ. The kriging based failure probability is $\mathrm{p_{f\varepsilon}} = 3.459 \cdot 10^{-3}$. The size of the initial seed, or the number of Markov Chains, for the sample generation, is set to $\mathrm{K_c} = 100$. The burn-in and thinning parameters are set to 20 and 10 respectively. For the purpose of the example, the MIS is performed for multiple steps after the targeted δ is reached.

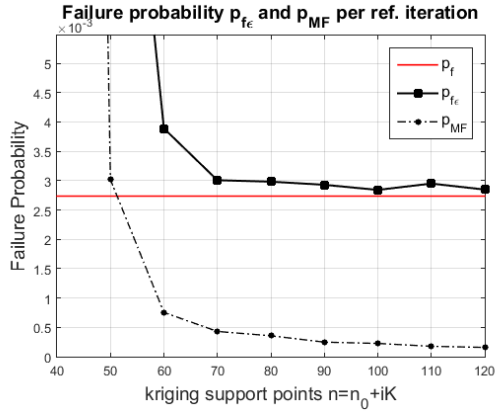


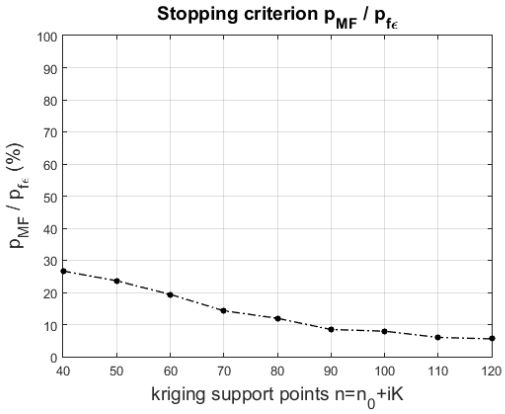


The above graphs show that at a targeted $\delta_{\text{corr}} = 1.5\%$, the final estimated failure probability is $\hat{p}_f = a_{\text{corr}} \cdot p_{f\varepsilon} = 0.9712 \cdot 3.459 \cdot 10^{-3} = 3.359 \cdot 10^{-3}$ with $55 + 200$ model evaluations and c.o.v. $\delta = 1.8\%$.

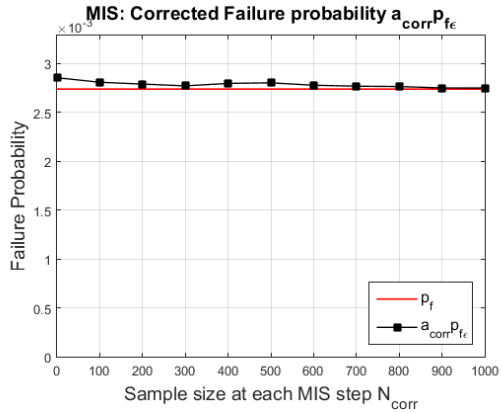
- $m = 10$

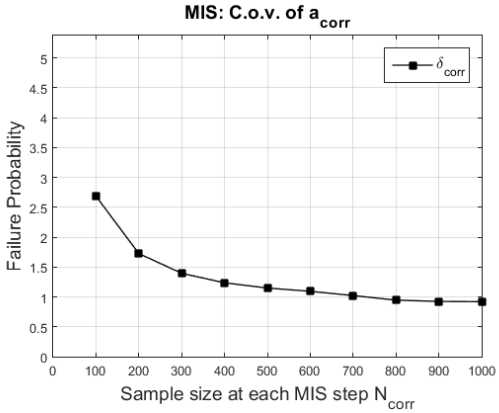
The procedure is the same, except the initial number of observations is set to $n_0 = 40$ and the additional per refinement iteration to $K = 10$.





The final model-based failure probability is estimated: $p_{f\varepsilon} = 2.852 \cdot 10^{-3}$ with $\delta_\varepsilon = 1.3\%$. Finally, the $p_{f\varepsilon}$ is corrected:

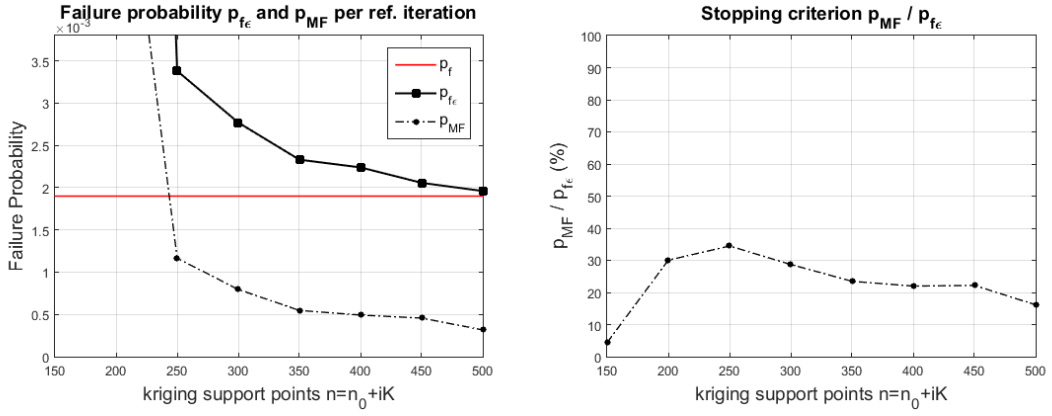




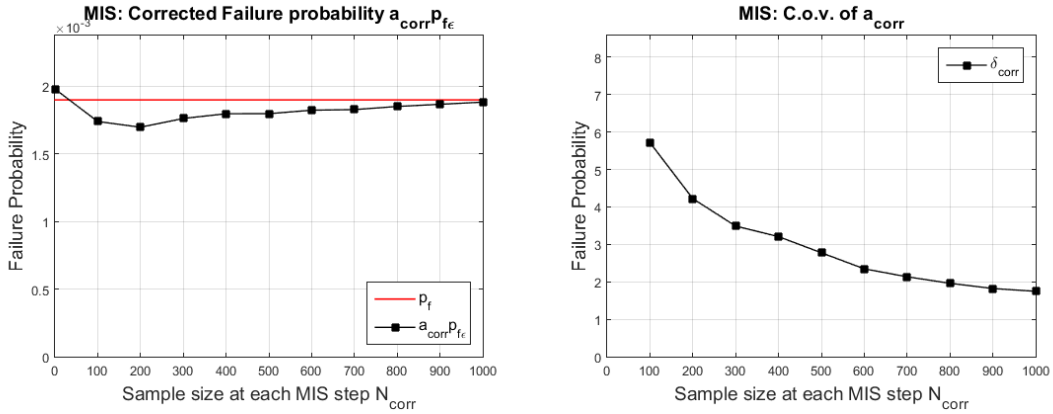
The targeted $\delta_{\text{corr}} = 1.5\%$ is reached at the 3rd MIS iteration. The corrected failure probability is $\hat{p}_f = a_{\text{corr}} \cdot p_{f\varepsilon} = 0.9708 \cdot 2.852 \cdot 10^{-3} = 2.769 \cdot 10^{-3}$ with c.o.v. $\delta = 2\%$, and the model evaluations were $120 + 300 = 420$.

- $m = 50$

The initial number of observations is set to $n_0 = 150$ and the additional per refinement iteration to $K = 50$.

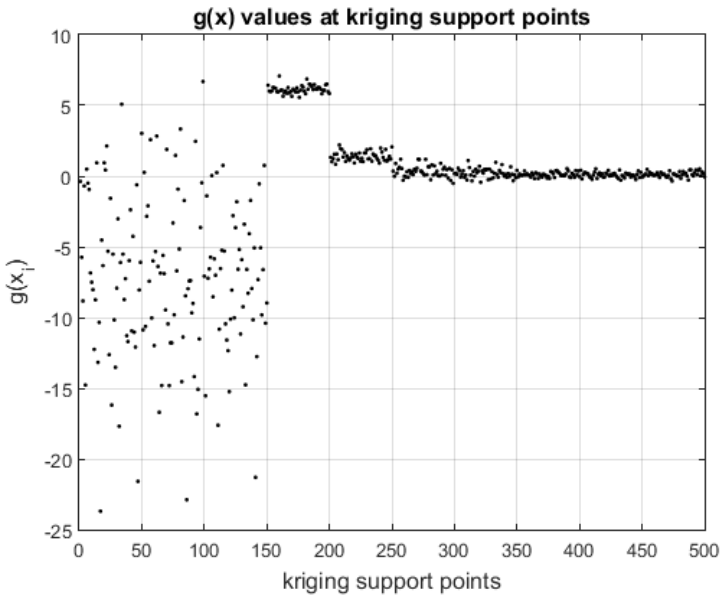


It is worth mentioning that at the first value of the ratio $p_{MF}/p_{f\varepsilon}$ is very low. This is attributed to the fact that the initial observations contained very few points in the safe domain (which is common in high-dimensional problems), and the prediction seemed confident. The final model-based failure probability is estimated: $p_{f\varepsilon} = 1.979 \cdot 10^{-3}$ with $\delta_\varepsilon = 1.5\%$. Finally, the $p_{f\varepsilon}$ is corrected:



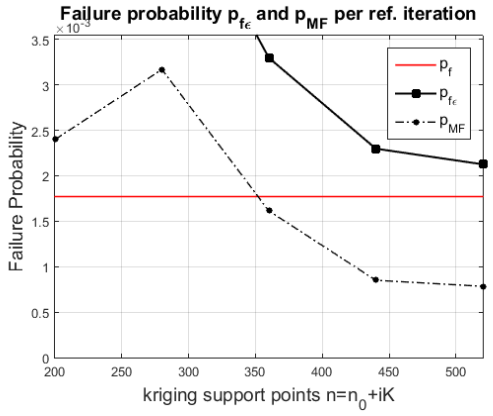
Stopping at the 10th iteration a c.o.v of $\delta_{\text{corr}} = 1.75\%$ is achieved. The final failure probability is $\hat{p}_f = a_{\text{corr}} \cdot p_{f\varepsilon} = 0.9512 \cdot 1.979 \cdot 10^{-3} = 1.882 \cdot 10^{-3}$ and $\delta = 2.30\%$ with $500 + 1000$ model evaluations.

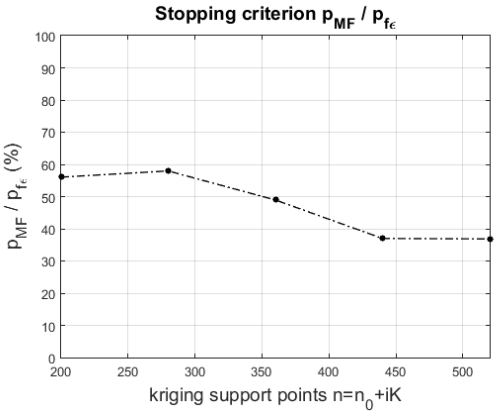
In the last graph below, the observations used to build the model, are shown by their $g(\mathbf{x})$ values, in the order that they were chosen. The first 150 are randomly distributed in the area of interest, while the rest are close to the estimated $g(\mathbf{x}) = 0$ limit.



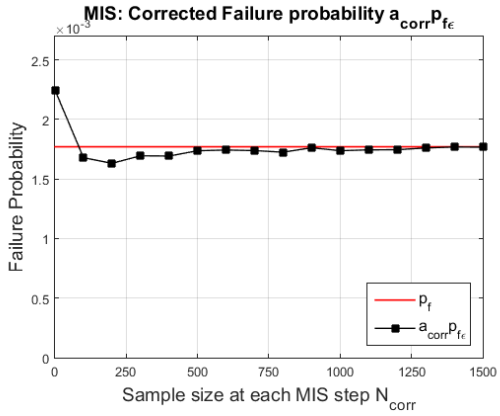
- $\mathrm{m} = 80$

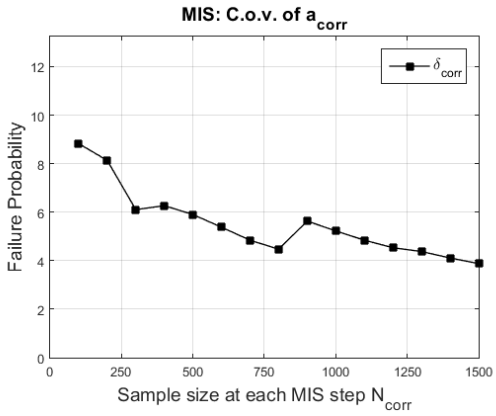
The initial number of observations is set to $\mathrm{n}_0 = 200$ and the additional per refinement iteration to $\mathrm{K} = 80$.





The final model-based failure probability is estimated: $\mathrm{p}_{\mathrm{f}\varepsilon} = 2.247 \cdot 10^{-3}$ with $\delta_\varepsilon = 1.5\%$. The $\mathrm{p}_{\mathrm{f}\varepsilon}$ is corrected:





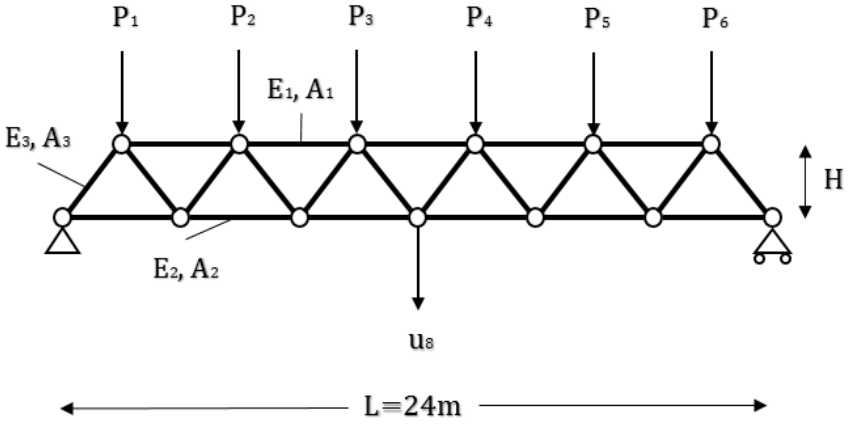
Stopping at the 10th iteration a c.o.v of $\delta_{corr} = 4\%$ is achieved. The final failure probability is $\hat{p}_f = a_{corr} \cdot p_{f\varepsilon} = 0.7866 \cdot 2.247 \cdot 10^{-3} = 1.767 \cdot 10^{-3}$ and $\delta = 4.30\%$ with $520 + 1500$ model evaluations.

The results are gathered in the following table:

	$m = 5$	$m = 10$	$m = 50$	$m = 80$
$p_{f\varepsilon}$	$3.459 \cdot 10^{-3}$	$2.852 \cdot 10^{-3}$	$1.979 \cdot 10^{-3}$	$2.247 \cdot 10^{-3}$
δ_ε	1%	1.3%	1.5%	1.5%
N_ε	55	120	500	520
a_{corr}	0.9712	0.9708	0.9512	0.7866
δ_{corr}	1.5%	1.5%	1.75%	4%
N_{corr}	200	300	1000	1500
$\hat{p}_f$	$3.359 \cdot 10^{-3}$	$2.769 \cdot 10^{-3}$	$1.882 \cdot 10^{-3}$	$1.767 \cdot 10^{-3}$
δ	1.8%	2%	2.3%	4.3%
N	255	420	1500	2020

4.4 Truss Bridge

The following truss bridge is modeled with 13 random variables.



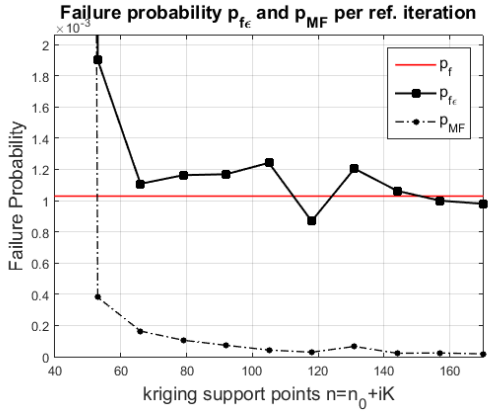
The vertical deformation of the center node is measured and restricted to $L/200$, where L the length of the bridge. The performance function is:

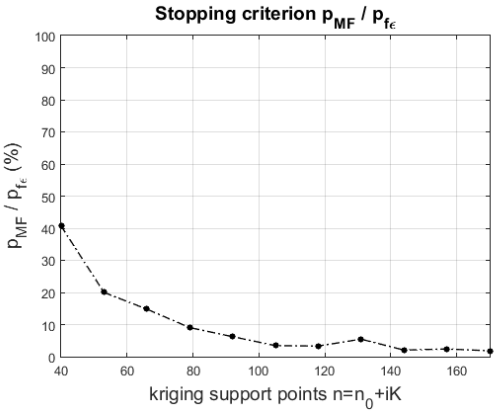
$$g(\mathbf{x}) = L/200 - u_8(\mathbf{x})$$

where $\mathbf{X}$ the random vector consisting of all the random normal variables shown below. All those variables are uncorrelated.

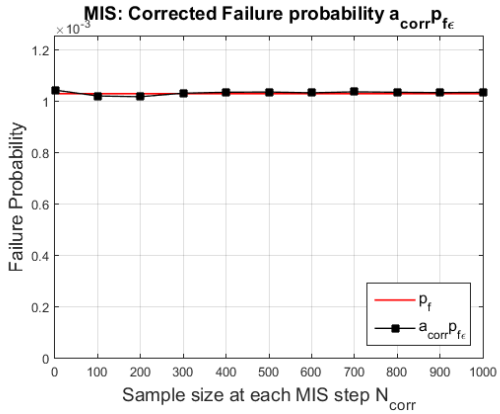
	μ	c.o.v.	σ
E_1, E_2, E_3	210 GPa	10%	21 GPa
A_1, A_2	20 cm^2	5%	1 cm^2
A_3	10 cm^2	5%	0.5 cm^2
$P_1 \sim P_6$	50 kN	15%	7.5 kN
H	2 m	5%	0.1 m

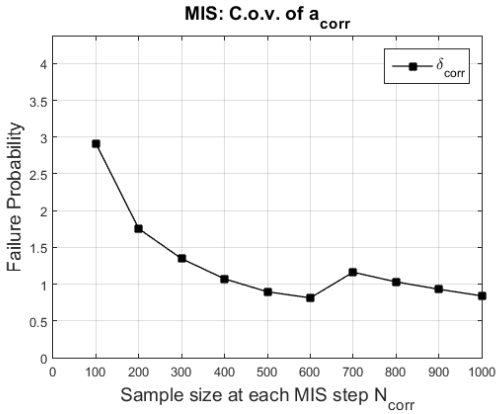
The kriging model is built with $n_0 = 40$ initial observations, and is refined with $K = 13$ new points at every step. The actual failure probability $p_f = 1.029 \cdot 10^{-3}, \delta = 1.5\%$ is estimated with MCS.





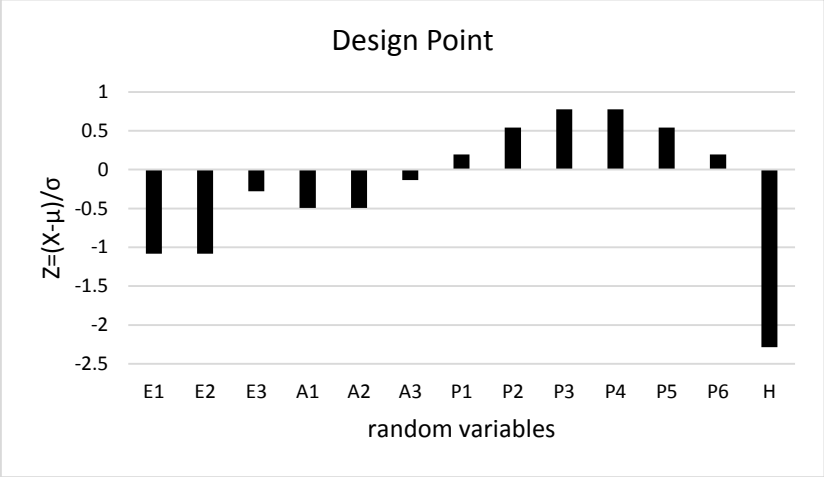
As shown above, 10 iterations are sufficient and the refinement stops at $n = 170$ total observations. The model based failure probability is $p_{f\varepsilon} = 1.043 \cdot 10^{-3}$, with $\delta_\varepsilon = 1.80\%$. The results of the MIS are the following:





At the 3rd iteration the $\delta_{corr} = 1.35\%$ is low enough and the $a_{corr} = 0.9885$ is close to unity. The final estimation is $\hat{p}_f = a_{corr} \cdot p_{f\varepsilon} = 0.9885 \cdot 1.043 \cdot 10^{-3} = 1.031 \cdot 10^{-3}$ with $\delta = 2\%$.

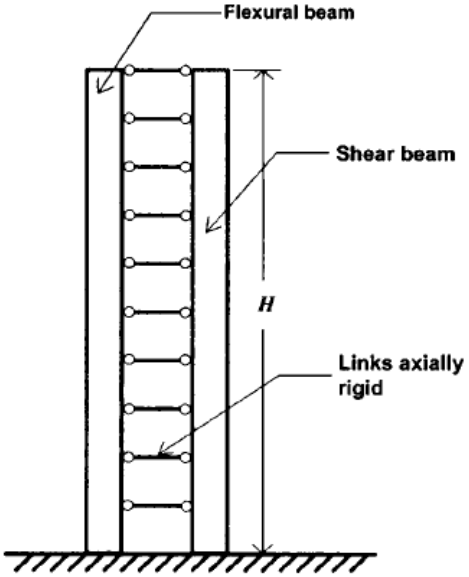
The design point, which is the point with the highest PDF value in the failure domain, is also illustrated below. Each random variable is normalized:



The variable that affects the most the deformation of the truss, is its height H. The E_3 and A_3 variables of the diagonals, affect it the less, as expected.

4.5 Tall building under seismic loads

This tall building of height $\mathrm{H} = 80\mathrm{m}$, is simulated as a simplified cantilever shear and flexural beam in parallel (Miranda, Taghavi 2005).



The quantities that determine the stiffness are EI and GA, which represent the flexural and the shear stiffness respectively. These two quantities are linked with the dimensionless factor α_0.

$$\alpha_0 = \mathrm{H}\left(\frac{\mathrm{GA}}{\mathrm{EI}}\right)^{\frac{1}{2}}$$

In this example, the α_0 is set to $\alpha_0 = 5$ which corresponds to a structural system that combines braced frames and moment-resisting frames. Assuming the geometry of the building and the geometric properties of its elements (columns, braces) are deterministic, and the elasticity module of the material is random, only one of the EI and GA quantities is set to be a random variable and the other one is evaluated from the expression of α_0. Therefore, the EI defined as the following stochastic process:

$$\mathrm{EI}(\mathrm{h}) \sim \boldsymbol{\mathcal{N}}\left(\mu_{\mathrm{EI}}, \mathrm{C}(\mathrm{h}, \mathrm{h}')\right)$$

where,

$\mu_{\mathrm{EI}} = 3.5 \cdot 10^8\ \mathrm{kNm}^2$ the mean values, constant along the height
$\mathrm{C}(\mathrm{h}, \mathrm{h}') = (0.1 \cdot \mu_{\mathrm{EI}})^2 \max(1 - |\mathrm{h} - \mathrm{h}'|/\mathrm{L}_{\mathrm{EI}}\,; 0)$, the correlation of $\mathrm{EI}(\mathrm{h})$ and $\mathrm{EI}(\mathrm{h}')$
$\mathrm{L}_{\mathrm{EI}} = 20\mathrm{m}$ the correlation length

The mass along the height of the build is also a stochastic process:

$$\mathrm{m}(\mathrm{h}) \sim \boldsymbol{\mathcal{N}}\left(\mu_{\mathrm{m}}, \mathrm{C}_{\mathrm{m}}(\mathrm{h}, \mathrm{h}')\right)$$

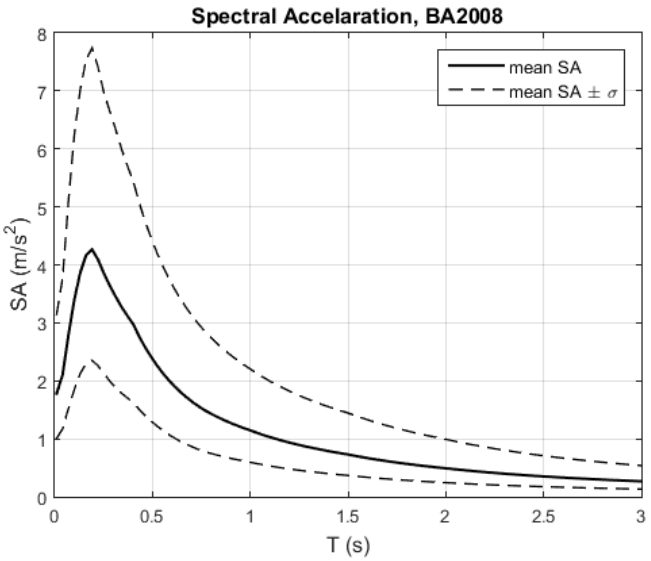
where,

$\mu_m = 120\ \mathrm{Mgr/m}$ the mean values, constant along the height
$C_m(h, h') = (0.1 \cdot \mu_m)^2 \max(1 - |h - h'|/L_m\ ; 0)$, the correlation of $EI(h)$ and $EI(h')$
$L_m = 12\mathrm{m}$ the correlation length

The beam is divided in 25 segments of constant EI and mass, therefore 50 correlated variables are defined so far. The seismic loads are arising from an earthquake spectrum, provided by the ground-motion prediction equations by Boore and Atkinson (2008). Assuming the following characteristics:

M	6.5	Moment magnitude
R_{jg}	$10\ \mathrm{km}$	Joyner-Boore distance
Fault type		Unspecified fault
$v_{s,30}$	$800\ \mathrm{m/s}$	Shear wave velocity

the following acceleration spectrum occurs:

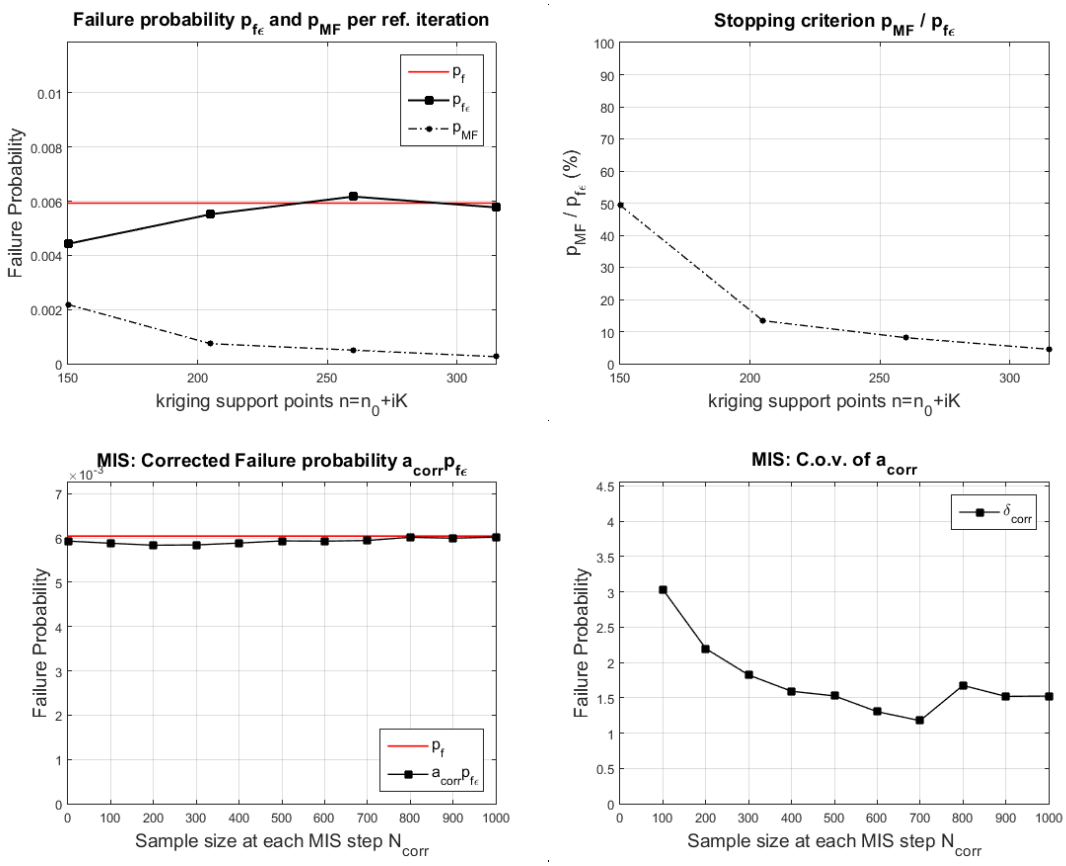


Lastly, the modal analysis provides with the 5 first eigenmodes, and the corresponding spectral accelerations $SA(T_{i=1,2,\ldots 5})$ are defined from the mean values and standard deviation by Boore and Atkinson, and the correlation $R_{SA}(T, T')$ by Baker and Jayaram (2008).

The performance function is the limit restriction in drift, meaning the maximum absolute value of the derivative of the beam's deflection.

$$g(\mathbf{x}) = 0.75\% - \max_h |\dot{u}(\mathbf{x}, h)|, \qquad h \in [0, H]$$

The kriging model is built with $n_0 = 150$ initial observations and is refined with $K = 55$ at each step. The actual failure probability is estimated on the real $g(\mathbf{x})$: $p_f = 6.043 \cdot 10^{-3}$ and $\delta = 1.2\%$.



The MIS reaches a very good approximation at the 10th iteration. The final failure probability is $\hat{p}_f = a_{corr} \cdot p_{f\varepsilon} = 1.0155 \cdot 5.930 \cdot 10^{-3} = 6.022 \cdot 10^{-3}$ with $\delta = 2.2\%$ and total number of $g(\mathbf{x})$ evaluations $N = 315 + 1000 = 1315$.

5 Conclusion

The kriging model based reliability analysis method is proven to fulfill its purpose, which is to estimate the failure probability with very few evaluations of the real performance function $g(\mathbf{x})$. To the author's experience, it is very efficient in problems with up to 20 dimensions, and practical up to 100 dimensions. In problems with more than 100 dimensions, the kriging correlation parameters are very difficult to be found and the method becomes computationally inefficient. This is additionally worsened, if the dataset becomes dense in certain areas, which is expected due to the refinement procedure. Such inefficiency can be treated by utilizing multiple kriging models, which are supported by a part of the total dataset. This approach is called big data regression and relies on clustering the original dataset.

It is worth mentioning though, that the kriging based analysis, does not lose efficiency as the failure probability reaches low values such as $< 10^{-7}$. This means that the number of $g(\mathbf{x})$ evaluations does not depend of the failure probability, but rather on the refinement procedure. This holds also for the Meta-model based Importance Sampling. The N_{corr} is high only when the refinement is incomplete. Consequently, the Subset Simulation performed on the kriging model, is the best approach to problems with low failure probability and hard-to-evaluate $g(\mathbf{x})$.

The meta-model based reliability analysis can be further combined with structural design optimization. In this case, there are multiple performance functions and each corresponds a failure probability which is bounded by an upper limit, with purpose to minimize the cost function.

Appendix A – Weighted K-means Clustering

The K-means (Steinhaus 1956, MacQueen 1967, Lloyd 1982) or the K-means++ (Arthur, Vassilvitskii 2007) algorithms solve the clustering problem. These algorithms aim to classify a dataset into clusters, according to the distance in the input space.

Step 0: Initial dataset $\mathfrak{O} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}\}$ with n points weighted by function $w(\mathbf{x}^{(i)})$, and number of clusters K

Step 1: Select K cluster centers randomly from the dataset $\mathfrak{O}$

$$\mathcal{C} = \{\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \dots, \mathbf{c}^{(K)}\}$$

The K-means++ algorithm utilizes careful initialization, achieving greater efficiency.

Step 2: Classify the data points into nearest cluster centers

$$m(\mathbf{c}^{(j)}|\mathbf{x}^{(i)}) = \begin{cases} 1 & \text{if } j = \arg\min_{k\in\{1,2,\dots K\}} \|\mathbf{x}^{(i)} - \mathbf{c}^{(k)}\|_2^2 \\ 0 & \text{else} \end{cases}$$

The above membership function is equal to 1, if $\mathbf{c}^{(j)}$ is the closest of all centers to the point $\mathbf{x}^{(i)}$, or else, it is equal to zero.

Step 3: Find new cluster centers

$$\mathbf{c}^{*(k)} = \frac{\sum_{i=1}^{n} w(\mathbf{x}^{(i)}) \cdot m(\mathbf{c}^{(k)}|\mathbf{x}^{(i)}) \cdot \mathbf{x}^{(i)}}{\sum_{i=1}^{n} w(\mathbf{x}^{(i)}) \cdot m(\mathbf{c}^{(k)}|\mathbf{x}^{(i)})}, k = 1, 2, \dots K$$

$$\mathcal{C} = \{\mathbf{c}^{*(1)}, \mathbf{c}^{*(2)}, \dots, \mathbf{c}^{*(K)}\}$$

Step 4: Calculate Inertia

$$I = \sum_{i=1}^{n} \sum_{k=1}^{K} w(\mathbf{x}^{(i)}) \cdot m(\mathbf{c}^{(k)}|\mathbf{x}^{(i)}) \cdot \|\mathbf{c}^{(k)} - \mathbf{x}^{(i)}\|_2^2$$

The algorithm continues to step 2. If the inertia does not change considerably in a few iterations, the algorithm stops.

Appendix B – Latin Hypercube Sampling

The Latin Hypercube Sampling is an alternative sampling method of the Monte Carlo Sampling. It ensures uniformity, but it may produce inaccurate results in reliability analysis. However, it is very useful in sampling the initial observations in constructing a kriging model.

The dataset of size n in sampled in the m-dimensional hypercube $\mathbf{x} \in [0,1]^{\mathrm{m}}$.

$$\mathfrak{D} = \left\{ \mathrm{x}_{\mathrm{j}}^{(\mathrm{i})} = \frac{\pi_{\mathrm{i}}^{(\mathrm{j})} - 1 + \mathrm{u}^{(\mathrm{i,j})}}{\mathrm{n}}, \qquad \mathrm{i} = 1,2, \dots \mathrm{n}, \mathrm{j} = 1,2, \dots \mathrm{m} \right\}$$

where, $\boldsymbol{\pi}^{(\mathrm{j})}$ a random permutation of the vector $\{1,2,3, \dots \mathrm{n}\}$, $\mathrm{u}^{(\mathrm{i,j})}$ a realization of a uniform distribution in the interval $[0,1]$.

Appendix C – Markov Chain Monte Carlo

In this thesis the Modified Metropolis-Hasting sampler (Au and Beck 2001) is used to generate sample from a targeted PDF that is not possible with Monte Carlo Sampling. The modified algorithm, is preferred to the usual in structural reliability. Such algorithms generate correlated samples, which is undesirable. The original algorithm, has a high reject rate in high dimensional problems, whereas the modified is unaffected by dimensionality. In this appendix the modified algorithm, the burn-in and thinning procedure and the numerical estimation of the γ coefficient are presented.

Modified Metropolis-Hasting Algorithm

The Modified Metropolis-Hastings Algorithm generates samples according to a target probability density function:

$$\mathrm{p}(\mathbf{x}) \propto \pi(\mathbf{x}) \cdot \mathrm{f}_{\mathrm{X}}(\mathbf{x})$$

where, $\pi(\mathbf{x})$ could be a positive function such as the probabilistic classification function or the indicator function, and $\mathrm{f}_{\mathrm{X}}(\mathbf{x})$ the PDF of random vector $\mathbf{X}$. It should be possible to define an isoprobabilistic transformation T and its inverse T^{-1}, from the random vector $\mathbf{X}$ to the random vector $\mathbf{U}$ with standard normal and independent components. Therefore the target PDF is rewritten:

$$\mathrm{p}(\mathbf{u}) \propto \pi(\mathbf{u}) \cdot \prod_{\mathrm{i}=1}^{\mathrm{m}} \varphi(\mathrm{u}_{\mathrm{i}})$$

where, $\varphi(\mathrm{u})$ the PDF of a standard normal variable.

Step 0: Target PDF $\pi(\mathbf{u}) \cdot \varphi_{\mathrm{m}}(\mathbf{u})$, initial seed $\mathbf{u}^{(0)}$ and length of Markov Chain N/K

Each seed is the start of an independent parallel chain. The number of seeds is K and the final sample is size N. $\mathrm{i} = 1, \mathrm{j} = 1$

Step 1: Propose a candidate for each component

A new candidate $\mathrm{u}_{\mathrm{i}}^{*}$ is randomly selected from a uniform distribution in:

$$\mathrm{u}_{\mathrm{i}}^{*} \sim \mathcal{U}\left(\mathrm{u}_{\mathrm{i}}^{(\mathrm{j}-1)} - 1, \mathrm{u}_{\mathrm{i}}^{(\mathrm{j}-1)} + 1\right)$$

It is accepted with probability:

$$\mathrm{r}_1 = \min\left\{\exp\left(\frac{1}{2}\left(\mathrm{u}_{\mathrm{i}}^{(\mathrm{j}-1)^2} - \mathrm{u}_{\mathrm{i}}^{*2}\right)\right); 1\right\}$$

Otherwise it is rejected, meaning $\mathrm{u}_{\mathrm{i}}^{*} = \mathrm{u}_{\mathrm{i}}^{(\mathrm{j}-1)}$. Set $\mathrm{i} = \mathrm{i} + 1$. If $\mathrm{i} > \mathrm{m}$, continue to step 2, else repeat step 1.

Step 2: Accept/Reject new sample

The vector $\mathbf{u}^{*}$ is accepted as $\mathbf{u}^{(\mathrm{j})} = \mathbf{u}^{*}$ with probability:

$$\mathrm{r}_2 = \min\left\{\frac{\pi(\mathbf{u}^{*})}{\pi(\mathbf{u}^{(\mathrm{j}-1)})}; 1\right\}$$

or rejected, $\mathbf{u}^{(j)} = \mathbf{u}^{(j-1)}$, otherwise. Set $j = j + 1$. If $j > N/K$ end, else set $i = 1$ and go to step 1.

If the $\pi(\mathbf{u})$ function is binary valued, such as the indicator function, the denominator of r_2 is always unit, and r_2 is also binary valued.

Burn-in and thinning

In Markov Chain samplers it is often difficult to obtain reliable initial seeds. Ideally, the initial seeds are distributed according to the target PDF, and sometimes this is true. However, the seeds are often distributed differently, or there are duplicate points. The burn-in procedure is to discard the initial $b = 10 \sim 20$ states of the Markov Chain and set the last state as seed, in order to achieve a better distribution.

The thinning procedure is to accept to the final sample, only one state every $t = 5 \sim 20$. Thinning achieves a less correlated sample as a result, which reduces greatly the γ coefficient.

Burn-in and thinning are useful in Meta-model based Importance Sampling, because the sample is chosen, based only on the kriging model, and not on the real $g(\mathbf{x})$. This means that the γ_{corr} and consequently the δ_{corr} can be reduced significantly without any computational cost. However, in Subset Simulation that is based on the real $g(\mathbf{x})$, thinning does not reduce the model evaluations, and for that reason it is not necessary.

Numerical estimation of coefficient $\boldsymbol{\gamma}$

A Markov Chain sample is used to evaluate the following integral:

$$\bar{p} = \int_{\mathbb{X}} \pi(\mathbf{x}) f(\mathbf{x}) dx$$

If the sample is distributed according to $f(\mathbf{x})$, then the estimation of $\bar{p}$ is:

$$\hat{\bar{p}} = \sum_{k=1}^{K} \sum_{i=1}^{N/K} \pi\left(\mathbf{x}^{[k](i)}\right)$$

where N the total size sample, and K the number of chains. The variance of that quantity is:

$$\sigma_{\bar{p}}^2 = \frac{\sigma_p^2}{N}(1 + \gamma)$$

where σ_p^2 is the variance of the samples:

$$\sigma_p^2 = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N/K} \pi\left(\mathbf{x}^{[k](i)}\right)^2 - \bar{p}^2$$

$$\gamma = 2 \sum_{l=1}^{N/K-1} \left(1 - \frac{lK}{N}\right) R(l)$$

and $R(l)$ the correlation between states that differ by lag distance l

$$R(l) = \frac{1}{\sigma_p^2 (N - lK)} \sum_{k=1}^{K} \sum_{i=1}^{N/K-l} \left(\pi\left(\mathbf{x}^{[k](i)}\right) - \bar{p}\right)\left(\pi\left(\mathbf{x}^{[k](i+l)}\right) - \bar{p}\right)$$

References

Au, S.-K. and J. Beck (2001). Estimation of small failure probabilities in high dimensions by subset simulation. Prob. Eng. Mech., 16(4), 263–277.

Au, S.-K., J. Ching, and J. Beck (2007). Application of subset simulation methods to reliability benchmark problems. Structural Safety, 29(3), 183–193.

Baker, J. W., Jayaram, N. (2008). Correlation of Spectral Acceleration Values from NGA Ground Motion Models. Earthquake Spectra, Volume 24, No. 1, pages 299–317

Bect, J., D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez (2011). Sequential design of computer experiments for the estimation of a probability of failure. Stat. Comput., In press.

Boore D. M. and Atkinson G. M. (2008). Ground-Motion Prediction Equations for the Average Horizontal Component of PGA, PGV, and 5%-Damped. Earthquake Spectra, Volume 24, No.1, pages 99-138

Der Kiureghian, A. and T. Dakessian (1998). Multiple design points in first and secondorder reliability. Structural Safety, 20(1), 37–49.

Dubourg, V., F. Deheeger, and B. Sudret (2011b). Metamodel-based importance sampling for structural reliability analysis. Submitted to Prob. Eng. Mech.

Marrel, A. (2005). Modélisation des codes de calcul dans le cadre des processus gaussiens stationnaires. Technical report, CEA (Stage INSA Toulouse).

Marrel, A. (2008). Mise en oeuvre et utilisation du métamodèle processus gaussien pour l'analyse de sensibilité de modèles numériques. Ph. D. thesis, Institut National des Sciences Appliquées de Toulouse.

Miranda, E., Taghavi, S. (2005). Approximate Floor Acceleration Demands in Multistory Buildings. I: Formulation

Picheny, V. (2009). Improving accuracy and compensating for uncertainty in surrogate modeling. Ph. D. thesis, University of Florida.

Picheny, V., D. Ginsbourger, O. Roustant, and R. Haftka (2010a). Adaptive designs of experiments for accurate approximation of a target region. J. Mech. Des., 132(7), 071008.

Rackwitz, R. (2001). Reliability analysis – A review and some perspectives. Structural Safety, 23, 365–395.

Rubinstein, R. and D. Kroese (2008). Simulation and the Monte Carlo method. Wiley Series in Probability and Statistics. Wiley.

Vazquez, E. (2005). Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications. Ph. D. thesis, Université Paris XI - UFR scientifique d'Orsay.