Estimation of rare event probabilities and extreme quantiles. Applications in the aerospace domain

Rudy Pastel

N° d'ordre : 4497 **ANNÉE 2012**

UNIVERSITÉ DE RENNES 1

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Mathématiques et applications

Ecole doctorale MATISSE

présentée par

Rudy Pastel

préparée à l'unité de recherche INRIA Rennes — Bretagne Atlantique
Institut National de Recherche en Informatique et en Automatique

---

Estimation de probabilités d'évènements rares et de quantiles extrêmes. Applications dans le domaine aérospatial

Thèse soutenue à Rennes le 14 février 2012

devant le jury composé de :

Benoît CADRE
Professeur, ENS Cachan / *président*

Michel BRONIATOWSKI
Professeur, université Pierre et Marie Curie / *rapporteur*

Bertrand IOOSS
Ingénieur de recherche HDR, EDF R&D / *rapporteur*

Pascal LEZAUD
Enseignant-chercheur, ENAC / *examinateur*

François LE GLAND
Directeur de recherche INRIA / *directeur de thèse*

Jérôme MORIO
Ingénieur de recherche ONERA / *co-directeur de thèse*

Contents

Nomenclature

Acronyms

AST	Adaptive Splitting Technique, page 38.
ANAIS	Adaptive Non parametric Adaptive Importance Sampling , page 69.
CLT	Central Limit Theorem , page 9.
CEP	Circular Error Probable, page 112.
CE	Cross-Entropy, page 26.
CMC	Crude Monte Carlo, page 7.
cdf	Cumulative distribution function (cdf) *i.e.* $\mathbb{P}[X \leq x]$, page 14.
ecdf	Empirical cumulative distribution function (ecdf) via CMC *i.e.* $\overline{F}_X(x)$, page 14.
IS	Importance Sampling, page 21.
iscdf	IS cumulative distribution function (iscdf) via IS *i.e.* $\widehat{F}_X(x)$, page 33.
ST	Splitting Technique, page 37.
MH	Metropolis-Hastings algorithm, page 44.
MVS	Minimum Volume Sets, page 113.
NEF	Natural Exponential Family, page 28.
NAIS	Non-parametric Adaptive Importance Sampling , page 30.
NIS	Non-parametric Importance Sampling , page 29.
PEF	Polynomial Exponential Family, page 95.
pdf	Probability density function, page 22.
SLLN	Strong law of large numbers, page 8.
w.r.t.	With respect to, page 9.

Conventional notations

Γ AST estimator simulation cost, page 47.

θ Density parameter, page 26.

Θ Density parameter set, page 26.

m Number of estimation, page 10.

ω Number of transition kernel application , page 43.

n Number of particle in the cloud, page 46.

α, Quantile level, page 39.

N Total number of throws when deterministic , page 10.

Measures and densities

f_Z Probability density function of auxiliary random variable Z, page 21.

f_X Probability density function of input random variable X, page 21.

ν Probability measure of auxiliary random variable Z, page 21.

μ Probability measure of input random variable X, page 21.

η Probability measure of output random variable Y, page 25.

M,Q Transition kernel, page 40.

Estimators

$\widetilde{Y}$ AST probability of exceedance estimator, page 47.

$\widetilde{q}$ AST extreme quantile estimator, page 48.

$\overline{X}_N$ Crude Monte Carlo expectation estimator , page 8.

$\overline{q}$ Crude Monte Carlo quantile estimator , page 15.

$\widehat{Y}_N$ Importance Sampling expectation estimator , page 22.

$\widehat{q}$ Importance Sampling quantile estimator , page 33.

$\breve{Y}$ ST probability of exceedance estimator, page 46.

Functions

$F_X(x)$ Cumulative distribution function (cdf) *i.e.* $\mathbb{P}[X \leq x]$, page 14.

g,h General purpose test functions , page 41.

ζ Satisfaction criterion, page 37.

Laws

Mappings

Operators

Random variables

Résumé étendu de la thèse

Dans cette thèse, nous nous donnons pour objectif d'identifier et au besoin de concevoir les outils mathématiques permettant d'évaluer des probabilités d'événement rares et des quantiles extrêmes dans un cadre réaliste et applicable dans l'industrie aérospatiale. L'Office National d'Etudes et de Recherches Aérospatiales (ONERA) doit en effet se préparer à prendre en compte les risques liés à des évènements rares du fait de l'élévation des attentes réglementaires et contractuelles.

Nous commencerons par puiser dans la littérature les outils théoriques probabilistes dédiés: la technique de Monte-Carlo, l'échantillonnage d'importance ou Importance Sampling et les estimations particulaires de type Importance Splitting. Ensuite, nous confronterons les algorithmes associés à des cas de difficulté croissante afin de déterminer leurs domaines d'application propres et d'éprouver leurs capacités à répondre aux besoins spécifiques de l'ONERA. Cela nous amènera à proposer un algorithme d'échantillonnage d'importance à base de noyaux de densité. De façon plus inattendue, la question de l'extension aux dimensions supérieures du quantile sera aussi abordée et nous proposerons également un algorithme dédié.

De la théorie à l'application

Tout d'abord, qu'est-ce que le risque? De quoi tout un chacun a-t-il peur? Le risque est la combinaison d'une menace et d'une vulnérabilité *i.e.* de la probabilité d'un évènement inopportun et de ses conséquences éventuelles.

$$\begin{array}{rcccc} \text{Risque} & = & \text{Menace} & \times & \text{Vulnérabilité} \\ \text{Risque} & = & \text{Probabilité d'un problème} & \times & \text{Coût de son occurrence} \end{array} \tag{1}$$

Il n'y a par conséquent que deux façons de réduire le risque: diminuer la probabilité du problème et réduire le coût de son occurrence. Toutes les politiques de gestion du risque visent à réduire l'un des deux termes de cette équation. Dans le cadre de l'ingénierie des systèmes, cela signifie concevoir les systèmes de façon à rendre les problèmes improbables et, s'ils arrivent malgré tout, non nuisibles. La question fondamentale est celle du dimensionnement des filets de sécurité.

L'ONERA fait partie depuis plus de soixante ans de tous les principaux projets aéronautiques et aérospatiaux de la France. Le laboratoire remplit des missions d'expertise technologique et de recherche scientifique pour le compte de l'Etat comme d'entreprises privées et son champ d'action s'étend de la recherche fondamentale au prototype. En particulier, le Département Conception et évaluation des Performances des Systèmes (DCPS) mesure les probabilités de panne et dimensionne les systèmes de façon à ce qu'il soient capables de résister à des dégâts spécifiés avec une probabilité minimale donnée. Mathématiquement, il s'agit

d'estimations de probabilités et de quantiles.

$$\mathbb{P}\left[X \in \mathbf{A}\right] =? \qquad \inf_{v \in \mathbb{R}} \left\{\mathbb{P}\left[X \leq v\right] \geq \alpha\right\} =? \tag{2}$$

Les composantes du risque sont généralement mesurées par la simulation numérique ou l'inférence à partir de données *i.e.* par les probabilités ou les statistiques. Cependant, seuls les évènements fréquents sont ainsi observés à moins de créer ou de posséder une très large base de données. En pratique, les évènements très peu fréquents sont souvent négligés c'est-à-dire que leurs probabilités sont considérées comme nulles, quels que soient leurs coûts. L'expérience et l'histoire montrent que les évènements rares arrivent et peuvent causer des conséquences dramatiques. Les risques majeurs ne devraient pas être identifiés *a posteriori*.

Les évènements de faibles probabilités peuvent être si coûteux qu'ils ne peuvent plus être ignorés. Cependant, les standards et normes de sécurité ont augmenté avec la complexité des systèmes développés tant et si bien que la seule puissance de calcul, bien qu'en grande augmentation elle aussi, ne suffit plus. Démontrer le respect des normes devient de plus en plus difficile pour le laboratoire. Pour éviter d'être dépassés, ses outils d'estimation du risque doivent être mis à jour et augmentés de façon à pouvoir gérer les probabilités infimes et le dimensionnement extrêmement sûr *i.e.* gérer les évènements rares.

Qu'est-ce donc qu'un évènement rare? Il n'y pas de définition mathématique absolue. Toutes les approches sont des compromis entre le temps de simulation disponible, la puissance des moyens de calcul et la probabilité cible. Nous les résumerons comme suit. Un évènement est dit rare si sa probabilité est inférieure à l'inverse du nombre maximal de simulations qu'il est possible de générer à l'aide des moyens de calculs disponibles durant le temps imparti.

De façon à faire face aux évènements rares (ER), l'ONERA s'est mis en quête de techniques dédiées. En raison de sa position centrale dans le processus de gestion de projet, le DCPS a été choisi pour réaliser une recherche amont dans le but d'introduire des outils dédiés aux ER dans l'industrie aérospatiale. Cela s'est fait par le biais de la présente thèse.

L'objectif de cette thèse est de tester la capacité des techniques dédiées aux ER existantes à répondre aux besoins spécifiques de l'ONERA. Ceux-ci sont représentés par deux cas d'études: la collision entre les satellites Iridium et Cosmos et la zone de sécurité lors de la retombée en chute libre d'un booster de fusée.

Le cadre formel considéré peut être présenté comme suit.

$$\begin{array}{ccccc} \mathbb{R}^d & \rightarrow & \mathbb{R}^n & \rightarrow & \mathbb{R}^l \\ X & \overset{\phi}{\mapsto} & Y & \overset{\zeta}{\mapsto} & V \end{array} \tag{3}$$

X représente l'entrée aléatoire d'un système complexe d'intérêt ϕ. X est une variable aléatoire et non un processus stochastique. Il s'agit donc d'un cas *statique* et non *dynamique*. Le système ϕ est simulé par le biais d'un code numérique coûteux à exécuter. C'est pour cela que ϕ est tout à la fois une boîte noire et le facteur limitant du processus d'estimation. En effet, à cause du coût de calcul de ϕ, les évaluations de Y, la sortie du système, sont coûteuses. Y est la quantité d'intérêt mais n'est perçue qu'au travers d'un critère ζ, usuellement scalaire. En toute rigueur, toutes les grandeurs estimées sont donc des espérances de V. Cependant,par abus de notation et dans un soucis de simplicité vis-à-vis du contexte immédiat, la notation de la variable aléatoire considérée changera et ϕ sera définie de multiples fois au cours de cette thèse.

Tout d'abord, dans la partie I, nous étudierons la littérature académique dédiée aux ER. Au cours de nos premières implémentations, dans la partie II, nous proposerons une amélioration de l'une de ces techniques. Nous testerons alors la robustesse de chacune et finalement nous les confronterons au cas réaliste de l'estimation de la probabilité de collision des satellites Iridium et Cosmos. Enfin, dans la partie III, nous traiterons un problème d'estimation de quantile extrême *i.e.* la détermination de la distance de sécurité lors de la chute libre d'un booster de fusée. De fait, nous reformulerons ce problème et le plongerons dans le cadre des événements rares.

Trois techniques probabilistes

Lorsque l'on fait face à des évènements rares, il faut faire des choix. Le premier est entre les probabilités et les statistiques. Le second est de choisir une technique dans la discipline.

Les probabilités et les statistiques sont deux disciplines très liées et traitent toutes deux des ER. La principale différence est chronologique. Les probabilités utilisent l'aléa pour produire des données. Les statistiques utilisent les données pour inférer l'aléa. Ainsi, on peut motiver son choix entre probabilités et statistiques par sa position vis-à-vis de la génération des données *i.e.* selon qu'on produise des données ou que l'on se serve de données existantes. L'ONERA, de part son positionnement amont, utilise la simulation pour produire des données sur le système d'intérêt. Comme la théorie statistique propose des outils dédiés aux ER comme la théorie des valeurs extrêmes [27], on peut essayer de combiner ces derniers avec des outils probabilistes, comme dans [57], ou par les copules comme [74]. Cependant, nous avons décidé de nous concentrer sur les approches purement probabilistes, même si une revue plus complète peut être trouvée dans [69].

Quand il s'agit d'utiliser les probabilités pour produire des évènements rares, le facteur limitant est le budget de simulation. La technique de Monte Carlo (MC) est la plus répandue des techniques d'estimation probabiliste et est bien maîtrisée d'un point de vue théorique. Cependant, simuler le comportement du système en respectant son aléa naturel par MC est trop coûteux. Les techniques d'Importance Sampling (IS) et de Splitting (ST) visent à produire un aléa plus propice à l'occurrence des évènements rares d'intérêt.

Cross-Entropy et *Non-parametric Adaptive Importance Sampling*

L'IS semble proche de MC tant formellement que d'un point de vue opérationnel. Il s'agit en effet d'effectuer une estimation MC après avoir effectué un changement de mesure de probabilité, ce qui se traduit en pratique par le choix d'une nouvelle densité d'échantillonnage.

$$\begin{aligned}
\mathbb{E}\left[\phi\left(X\right)\right] &= \int_{\mathbb{X}} \phi\left(x\right) f_X\left(x\right) \lambda\left(\mathrm{d}x\right) \\
&= \int_{\mathbb{X}} \phi\left(z\right) \frac{f_X\left(z\right)}{f_Z\left(z\right)} f_Z\left(z\right) \lambda\left(\mathrm{d}z\right) \\
\mathbb{E}\left[\phi\left(X\right)\right] &= \mathbb{E}\left[\phi\left(Z\right) \frac{f_X\left(Z\right)}{f_Z\left(Z\right)}\right]
\end{aligned} \tag{4}$$

Toute la difficulté est alors concentrée dans le choix de cette nouvelle densité. En effet, même si une densité auxiliaire idéale existe, elle demande de déjà connaître l'espérance voulue! Même si elle n'est pas disponible en pratique, cette densité idéale fournit néanmoins une densité cible qu'il s'agit d'approcher ou d'apprendre au mieux.

La *Cross-Entropy* (CE) invite à choisir la densité auxiliaire au sein d'une famille paramétrique. Le choix optimal est alors la densité qui réalise, au sein de la famille, la plus petite divergence de Kullback-Leibler vis-à-vis de la densité idéale. Le *Non-parametric Adaptive Importance Sampling* (NAIS) propose quant à lui d'apprendre itérativement la densité idéale à l'aide d'une estimation par noyaux pondérés.

Dans les deux cas, l'apprentissage est coûteux. L'idée de [5, 15] est d'entraîner gratuitement à l'infini les densités auxiliaires sur une approximation du système appelée méta-modèle et construite à peu de frais, par krigeage [22] par exemple. Cela implique une nouvelle couche technique, l'estimation du méta-modèle, mais permet de mieux choisir les simulations à effectuer à l'aide du vrai système. Comme cette approche peut être utilisée avec toutes les techniques IS, nous nous sommes concentrés sur CE et NAIS.

Adaptive Splitting Technique

La technique de splitting adaptatif (AST) consiste à reformuler l'infime probabilité cible en un produit de probabilités conditionnelles plus importantes à estimer itérativement. Cette technique est dite adaptative parce que les conditions sont définies à la volée.

$$\mathbb{P}[X \in \mathbf{A}] = \prod_{i=1}^{\kappa} \mathbb{P}\left[X \in \mathbf{A}_i \middle| X \in \mathbf{A}_{i-1}\right] \tag{5}$$

Bien que cette formulation semble à peine plus complexe que l'originale, l'AST demande un noyau Markovien réversible par rapport à la mesure de probabilité naturelle *i.e.* celle de X, et un tel noyau n'est pas toujours explicitement connu. C'est le cas si X est Gaussien, mais en pratique, on devra souvent avoir recours à des noyaux de type Metropolis-Hastings. De tels noyaux sont délicats d'utilisation et augmentent le nombre déjà important de paramètres de l'AST. Le calibrage est donc la principale difficulté de cet algorithme car il est en grande partie livré à l'arbitraire.

Après cette revue, nous sommes passés aux applications pratiques avant de commencer les cas réalistes.

NAIS adaptatif et le cas Iridium Cosmos

Nous avons commencé les applications par des cas d'entraînements. Ceux-ci nous ont amenés à modifier l'algorithme NAIS. Puis nous avons effectué le cas réaliste de l'estimation de la probabilité de collision entre les satellites Iridium et Cosmos. Ce dernier cas nous a permis d'éprouver la maturité expérimentale de chaque technique.

ANAIS

NAIS ne peut être utilisé sans la connaissance *a priori* d'une densité auxiliaire qui génère d'emblée des évènements rares, alors qu'une telle densité est exactement le but recherché. Cela

réduit fortement l'intérêt pratique de NAIS. Pour palier cette difficulté d'utilisation, nous avons proposé le NAIS adaptatif (ANAIS) en nous inspirant de CE et AST. La nouvelle stratégie est d'utiliser une suite croissante de seuils définis à la volée pour apprendre éventuellement la densité auxiliaire idéale. Cela permet de contourner l'écueil du choix de la densité auxiliaire initiale et autorise même à choisir la densité de X. C'est pour cela que malgré son besoin d'approfondissement théorique, ANAIS est utilisé à la place de NAIS.

Cas d'entraînement

Les cas d'entraînement impliquaient des lois de Rayleigh. Ces lois ont l'avantage d'être bien connues et de limiter les difficultés à la rareté des événements impliqués. Elles semblaient donc un bon choix pour une première expérience. Il fallait estimer, avec un budget donné, des probabilités d'évènements rare – le dépassement d'un seuil élevé – ou des quantiles extrêmes. De plus le seuil, la dimension de l'espace et le niveau du quantile variaient. De cette étude, effectuée avec un budget de simulation constant de 50000 réalisations, on retiendra surtout les classements expérimentaux suivants.

Nous avons testé la robustesse à l'augmentation de la dimension en faisant varier celle-ci de 1 à 20. CE et AST ont été quasiment insensibles alors que la performance de ANAIS s'est détériorée avec l'augmentation de la dimension. Dans l'ordre des variances croissantes, on peut schématiquement les classer comme suit.

$$1 \leq d \leq 10 \qquad 3.\text{AST} \qquad 2.\text{ANAIS} \qquad 1.\text{CE} \tag{6}$$

$$11 \leq d \leq 20 \qquad 3.\text{ANAIS} \qquad 2.\text{AST} \qquad 1.\text{CE} \tag{7}$$

ANAIS, en tant qu'estimateur par noyau d'une densité n'a pas supporté l'élévation de la dimension.

Nous avons fait varier la probabilité cible et le niveau du quantile cible de $1 - 10^{-5}$ à $1 - 10^{-13}$. ANAIS et CE ont gardé une variance basse et stable alors que celle d'AST était haute et croissante. Nous les avons classées comme suit, dans l'orde des variances croissantes.

$$3.\text{AST} \qquad 2.\text{ANAIS} \qquad 1.\text{CE} \tag{8}$$

Il est également possible que la calibrage d'AST n'ait simplement pas été adapté.

Pour ce qui est du calibrage, chacune des trois techniques testées a posé des problèmes spécifiques. On peut néanmoins les ranger comme suit dans l'ordre des difficultés croissantes.

$$3.\text{CE} \qquad 2.\text{AST} \qquad 1.\text{ANAIS} \tag{9}$$

Cette première place peu commode n'est cependant apparue véritablement que lors de l'expérience Iridium-Cosmos, plus réaliste.

Le cas Iridium-Cosmos

La collision inattendue en 2009 de ces deux satellites a poussé l'ONERA à réfléchir aux processus de gestion de la sécurité des satellites. Mesurer les probabilités de collision en est une étape importante. Il s'agissait donc tant d'un cas d'étude réaliste que d'un acte de prospection visant à supprimer les hypothèses permettant actuellement de faire cette estimation.

L'estimation par CE n'a pas pu être menée à terme parce que la densité paramétrique choisie était uni-modale et de ce fait non adaptée. Nous avons bien réfléchi à des changements de mesures exponentiels multi-modaux à base de polynômes, mais échantillonner selon de telles lois pose problème.

ANAIS n'a généré aucune collision et a estimé la probabilité nulle. Certains points clés de la méthode ont été reconsidérés mais un nouvel algorithme n'a pas encore été écrit.

AST a estimé la densité cible sans biais et avec un écart-type relatif de $\frac{2}{5}$. Cette valeur est similaire à celle des cas d'entraînement. Même si son calibrage demande à être approfondi d'un point de vue théorique, AST semble être le meilleur choix.

Conclusion opérationnelle: probabilités d'évènements rares

A l'issue de cette étude, nous avons pris deux décisions. La première était de ne pas utiliser CE sans une information sur le système suffisante. En effet, cette technique demande trop d'information *a priori* pour pouvoir faire un choix de famille paramétrique adapté à une boîte noire. Nous n'avons par conséquent plus utilisé CE par la suite. La seconde décision a été de désigner AST comme une technique d'estimation de probabilité d'évènements rares adaptée à notre cadre d'étude caractérisé par la présence d'une boîte noire, une fois qu'une règle théorique et opérationnelle de calibrage aura été établie. Quant à ANAIS, cet algorithme semblait digne d'une étude approfondie.

Nous sommes alors passés au second cas d'étude, afin d'effectuer une estimation de quantile extrême réaliste. Il s'agissait de déterminer la distance de sécurité dans le cadre de la chute libre d'un booster de fusée.

Estimation de zones de sécurité à l'aide de *Minimum Volume Sets* extrêmes.

Nous nous sommes initialement attelés à l'estimation de la distance à au point de chute prévu d'un booster de fusée assurant un haut niveau de sécurité prescrit. Poussés par le sens pratique, nous avons en effet été amenés à dépasser le cadre l'estimation de quantiles extrêmes.

Pourquoi le quantile extrême ?

Comme prévu, MC n'a pas été capable d'estimer distinctement les quantiles extrêmes de distance entre le point de chute effectif du booster et le point initialement prévu. Comme espéré, AST et ANAIS ont pu et d'une façon similaire. Cependant ces quantiles induisaient des zones de sécurité circulaires qui ne correspondent pas du tout à la géométrie de la distribution des points de chute sur le plan. Nous avons donc recherché dans la littérature en quête d'un outil plus approprié. Bien que la recherche d'une extension de la définition du quantile aux dimensions quelconques soit un sujet actif chez les statisticiens, nous avons choisi une approche probabiliste: le *Minimum Volume Set* (MVS). Dans notre cas, il s'agissait de la plus petite surface contenant une masse de probabilité donnée. Le MVS présente donc une interprétation ayant un sens concret et pratique: un MVS concentre les points de grande densité de probabilité de façon à réunir la probabilité voulue sur la surface la plus petite possible. De plus, il épouse la

géométrie de la densité de la distribution du point de chute sur le plan car c'est essentiellement un ensemble de niveau de cette dernière. Ce niveau est cependant défini comme un quantile de la variable $f_Y(Y)$ où Y est le point de chute aléatoire sur le plan et f_Y sa densité. L'estimateur MC de MVS couple donc d'une part l'estimation par noyaux de f_Y et d'autre par l'estimation des quantiles de $f_Y(Y)$. Par conséquent, cet estimateur est incapable d'estimer un MVS de niveau extrême parce qu'il est fondé sur un quantile de niveau extrême. Pour cela, il faut des outils dédiés.

Estimation de MVS extrêmes

Nous avons adapté AST et ANAIS pour produire des estimateur de MVS extrêmes que nous avons comparés tant du point de vue de la géométrie que de la variance. La version AST a été handicapée par une question théorique non résolue qui ne lui a permis d'exploiter pleinement que la moitié du budget de simulation. C'est pour cela qu'en l'état, la version ANAIS apparaît comme le meilleur choix. Cela dit, les deux versions demandent des approfondissements théoriques. De fait, AST et ANAIS nécessitant déjà individuellement des approfondissements, c'est *a fortiori* le cas lorsqu'ils sont couplés avec un estimateur de densité à base de noyaux pondérés.

Conclusion opérationnelle: quantiles et MVS extrêmes

Nous n'avons de pas réponse parfaitement tranchée quant à quelle technique utiliser pour estimer des quantiles extrêmes. En effet, AST et ANAIS posent des questions théoriques non encore résolues. Un argument pratique est de décider en fonction de la dimension de l'espace d'entrée. Si elle est inférieure à 10, ANAIS semble le meilleur choix alors qu'on devrait choisir AST pour des dimensions supérieures. Quant à l'estimation de MVS extrêmes, elle bénéficiera grandement des études théoriques portant sur AST et ANAIS et sera sûrement alors opérationnelle.

Perspectives de recherches

De fait, toute ces études mettent en évidence de nombreuses questions théoriques. Les principales perspectives de recherche sont les suivantes.

Dans le cas de la CE, il serait intéressant de pouvoir échantillonner selon un changement de mesure exponentiel dont l'argument est un polynôme de la variable d'espace, une fois que le nombre de composantes connexes d'intérêt est connu.

Quant à ANAIS, rendre monotone la suite des seuils rendrait peut-être l'algorithme plus efficace. Savoir comment comment le calibrer aussi.

Pour ce qui est de l'AST, le calibrage est également un problème. De plus, les algorithmes de type Metropolis-Hastings étant un problème en soi, il serait commode de disposer de plus de paires de densité et de noyaux Markoviens réversibles explicites.

En ce qui concerne spécifiquement l'estimation de MVS extrêmes, le première question est celle des estimations de densité par des noyaux pondérés.

Il serait pratique des savoir comment, dans chacun des cas sus-cités, répartir au mieux son budget de simulation.

Pour ce qui est des questions de paramétrage, il serait intéressant de voir ce que peuvent apporter des procédures d'automatisation comme celles décrites dans [58].

De la pratique à la théorie

Cette thèse a permis à l'ONERA de commencer sa préparation aux ER en ciblant parmi les outils théoriques existants ceux qui correspondent à ses besoins dans un contexte de boîte noire ayant en entrée une variable aléatoire. L'algorithme AST apparaît comme un bon candidat. Quant à CE, il ne devrait pas être utilisé avec une boîte noire mais seulement si un minimum d'information sur le système est disponible. Pour sa part, l'algorithme ANAIS que nous avons créé demande encore à murir. Par ailleurs, nous avons amené l'ONERA à introduire les MVS parmi ses outils de quantification de la dispersion spatial d'une variable aléatoire. L'estimateur MC de MVS n'étant pas adapté aux MVS de niveaux extrêmes, nous avons construit deux algorithmes dédiés: ANAIS-MVS et AST-MVS.

Ce travail a été principalement tourné vers l'application en raison du désir de l'ONERA de disposer d'outils opérationnels, les probabilités n'étant pas son cœur de métier. Tous ces outils demandent cependant un approfondissement théorique.

Introduction: from theory to practice

The contemporary western society wants to eliminates risks. Actually, at levels of the western society, risk is an issue. The public punishes it *via* protests and elections. Clientele avoids it *via* online goods experience pooling and online discontent publicity. Politics forbids it *via* safety norms and regulations. Therefore, companies and industries have to assess and parry risks. The huge amounts of money now at stake in any business or scientific endeavor invite entrepreneurs to be conservative and look for assurances. This thesis is about how the *Office National d'Etudes et de Recherches Aérospatiales* (ONERA), the French Aerospace lab, is preparing to deal with an emerging class of risk: rare events.

First things first: what is risk? What is it everyone is so afraid of? Risk is the combination of a threat and a damage *i.e.* the combination of the probability of a undesired event and its cost.

$$\begin{array}{rcccc} \text{Risk} & = & \text{Threat} & \times & \text{Damage} \\ \text{Risk} & = & \text{Probability of failure} & \times & \text{Cost of failure} \end{array} \tag{10}$$

There are therefore only two ways reduce risk: decreasing the probability of failure and lessening the cost of failure. All risk avoidance procedures aim at diminishing either one of, or both, the factors of the risk equation. With respect to system engineering this means designing the system so that failures are unlikely and, if they do occur, harmless. The fundamental questions are how to measure unlikeliness and how to dimension safety nets.

The ONERA has been participating to all major French aeronautics and aerospace projects over the last six decades, delivering research and technology services to the State and private companies alike, from theoretical studies to prototype testings. In particular, the System design and performance evaluation Department (DCPS) assesses the probabilities of failure and dimension systems so they are capable of enduring specified damages with a given minimal probability. Mathematically, this is probability and quantile estimations.

$$\mathbb{P}\left[X \in \mathbf{A}\right] =? \qquad\qquad \inf_{v \in \mathbb{R}} \left\{\mathbb{P}\left[X \leq v\right] \geq \alpha\right\} =? \tag{11}$$

These components of risk are usually assessed *via* system simulation or inference from data *i.e.* probability and statistics. However, only the frequent events can be observed that way unless one grows or possesses very large data. In practice, very unlikely events used to be disregarded *i.e.* their probabilities used to be deemed zero, whatever their costs. Experience showed that unlikely events do happen and can be very costly. Major risks should not be acknowledged *a posteriori*.

Very low probability events can be so costly they can not be ignored anymore. However, safety and requirement standards increased with the complexity of the developed systems so

that sheer calculation power, though soaring, can not keep up with them. Demonstrating the fulfillment of the requirements might become impossible for the lab. To avoid this, the lab's risk estimation toolboxes need to be updated and refurbished so as to cope with minute probabilities and dimensioning to extreme safety *i.e.* so as to be able to handle rare events.

What is a rare event, then? There is no clear cut mathematical definition of a rare event. One can give a threshold, and claim that any event with a lesser probability is rare. One can give a simulation timespan, say two hours, and claim any event not simulated during that time rare. One can give a simulation budget *i.e.* a number of calls to a software, and claim any event never realized when this budget is depleted rare. All these approaches balance simulation time, calculation power and targeted probability. They can be summed up as follows. A random event is said rare if its probability is less than the inverse of the maximum number of simulations that can be done within the available simulation time using the calculation power at hands.

In order to deal with rare events, the ONERA looked for rare event dedicated user-friendly tools to spread among its scientific staff. Because of it central position in the project management framework, the DCPS was chosen to perform upstream research about the introduction of rare event dedicated tools in the aerospace industry. This was done *via* this thesis. This thesis objective was testing the capability of existing rare event dedicated techniques to cope with the needs of the ONERA *via* two case studies, the Iridium-Cosmos satellite collision and a booster fallout safety zone.

The considered framework can be represented as follows.

$$\begin{array}{ccccc} \mathbb{R}^d & \to & \mathbb{R}^n & \to & \mathbb{R}^l \\ X & \overset{\phi}{\mapsto} & Y & \overset{\zeta}{\mapsto} & V \end{array} \tag{12}$$

X represents the random input of ϕ which stands in as the complex system of interest. X is a random variable and not a stochastic process. This case is therefore *static* and not *dynamic*. This system is usually simulated *via* a patchwork of costly to run softwares. This is why ϕ comes in as both a black box and the bottleneck of the estimation process. Events are rare because ϕ can not be ran many times. Actually because of the cost of ϕ runs, the evaluations of Y, the system output, are expensive. Y is the quantity of interest but is seen through a criterion ζ, which is usually scalar. Strictly speaking, the whole work is focused on estimating the expectation of V type quantities. However, by notation abuse and in order to keeps things as simple as possible with respect to their specific contexts, the notation of the considered random variable will change and ϕ will be multiply defined in the course of this presentation.

First, we scanned the rare event technique academic literature as summed up in Part I. During our first implementations, Part II, we came up with an improvement of one of these techniques. Then we tested their robustness and eventually used them in a realistic probability estimation case as we estimated the probability that satellites Iridium and Cosmos collided. Then, in Part III, we considered an extreme quantile estimation problem when dealing with the booster fallout safety zone. Actually, we reformulated the problem and adapted the dedicated tool to the rare event framework.

Part I

Crude Monte Carlo, Importance Sampling and Splitting

Introduction

When dealing with rare events, one has to make choices. The first one is between probability and statistics. The second is selecting the technique to use within the chosen field.

Both statistics and probability engage rare events and the disciplines are very linked. Their main difference is chronological. Probability uses *randomness to produce data*. Statistics uses *data to infer randomness*. Therefore, one's choice between statistics and probability completely depends on whether one generates the data or draws from the data. In our case at ONERA, we use simulation to produce data about the systems of interest. As statistics does come with rare event dedicated tools such as Extreme Value Theory [27], one can try to combine it with probability as in [57] or *via* Copulas [74]. We decided not to add technicalities up. As a consequence, we focused on probability, even though a more complete review can be read in [69].

When it comes to using probability to generate rare events, the limiting factor is the simulation cost. As detailed in Chapter 1, simulating the system's behaviour according to its natural randomness *via* Crude Monte Carlo is too costly. To circumvent this, one idea is simulating the system's behaviour according to randomness better suited to the purpose. This is the objective of Importance Sampling – Chapter 2 – and the Splitting Technique – Chapter 3 –.

Chapter 1

The Crude Monte Carlo method

The Crude Monte Carlo (CMC) is by far the most spread probability and quantile estimation method, for it is easy to understand theoretically, fast to implement and its results are convenient to communicate. However, what really accounts for its success, is the wide array of predicaments CMC's simplicity enables one to sort out.

1.1 A brief history

Though there had been previous isolated use of the technique before, such as an estimation of π throwing needles on a plane wooden surface with parallel equidistant lines [43], 1944 is a corner stone in the Monte Carlo method history.

During the Manhattan Project [77], the CMC was used in its modern form to simulate random neutron diffusion in fissile material [60]. As the whole project was secret, the technique was codenamed "Monte Carlo" after the Monte Carlo Casino, Monaco, where a scientist's relative would spend the money he had borrowed from his family [32, 61]. The technique benefited the birth of the Electronic Numerical Integrator Analyser and Computer a.k.a. ENIAC, the first electronic computer to run the otherwise impossible extensive calculation required.

This newly allowed practicability triggered a massive interest for the technique which bloomed just at par with the computer boom to become, according to [23], one the top ten algorithms in the XX$^{\text{th}}$ century, mostly for its very approachable technical features.

1.2 Probability estimation

1.2.1 A direct application of the strong law of large numbers

CMC comes as a direct use of the well-known strong law of large numbers which formalizes the link between the empirical mean of an integrable random variable and its expectation.

Theorem 1.2.1 (Strong law of large numbers)**.** *Let* $(X_1, \cdots, X_N)$ *be a set of* N *independent identically distributed (*iid*) as* X *and integrable random variables* $(\mathbb{E}[|X|] < \infty)$*. Then, the empirical mean* $\overline{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k$ *converges almost surely toward their expected value* $\mathbb{E}[X]$

i.e.

$$\overline{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k \xrightarrow[N\to\infty]{a.s.} \mathbb{E}[X] \tag{1.1}$$

Proof. Two different proofs can be found in [35, 44]. □

This strong law of large numbers (SLLN) justifies the use of the empirical mean to approximate an expectation. Furthermore, it is robust as it still holds when hypothesis are lightly changed.

1.2.2 CMC expectation estimator

The CMC estimator makes a direct use of the SLLN and estimates the expectation of a random variable X as the empirical mean of N *iid* throws $\overline{X}_N$.

Algorithm 1.2.1 (CMC expectation estimator)**.** *To estimate the expectation of integrable random variable X proceed as follows.*

1. *Generate the* iid *sample set $(X_1, \cdots, X_N)$ such that $\forall i, X_i \sim X$.*
2. *Calculate its empirical mean $\overline{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k$.*
3. *Approximate $\mathbb{E}[X] \approx \overline{X}_N$.*

This algorithm only requires being able to simulate iid *throws of X.*

Besides, the CMC estimator itself is a random variable while the expectation is a deterministic quantity. This fact triggers three questions:

1. What is $\overline{X}_N$ law for a fixed value of N?
2. In what sens does this law converge ? If so, what does it converge to?
3. How fast is this convergence?

The first question will be saved for Subsection 1.2.4 while the two last birds are hit with the same stone in Subsection 1.2.3.

1.2.3 The central limit theorem

$\overline{X}_N$ law's convergence comes directly from a very famous theorem as well, with a slight extra constraint: square-integrability.

Theorem 1.2.2 (Central Limit Theorem)**.** *Let $(X_1, \cdots, X_N)$ be a set of N* iid *as X and square-integrable random variables ($\mathbb{E}[|X|^2] < \infty$) and note X variance $\mathbb{V}[X]$ and the set empirical mean $\overline{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k$. The random variable sequence $\sqrt{N}\left[\overline{X}_N - \mathbb{E}[X]\right] = \frac{\sum_{k=1}^{N} X_k - N\mathbb{E}[X]}{\sqrt{N}}$ converges in law toward $\mathcal{N}(0, \mathbb{V}[X])$, the centered Gaussian with variance $\mathbb{V}[X]$, when the sample size N goes to infinity. Thus*

$$\sqrt{N}\left[\overline{X}_N - \mathbb{E}[X]\right] = \frac{\sum_{k=1}^{N} X_k - N\mathbb{E}[X]}{\sqrt{N}} \xrightarrow[N\to\infty]{in\ law} \mathcal{N}(0, \mathbb{V}[X]) \tag{1.2}$$

This Central Limit Theorem (CLT) provides very useful informations.

First, it states that one can approximate the law of $\overline{X}_N$ with a very convenient Gaussian to design confidence intervals for large enough N. That is to say, assuming $a > 0$ and the random variables on the line, equation (1.3) helps building a $\int_{-a}^{a} e^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}}$ confidence level interval.

$$\mathbb{P}\left[\mathbb{E}[X] \in \left[\overline{X}_N - a\sqrt{\frac{\mathbb{V}[X]}{N}}, \overline{X}_N + a\sqrt{\frac{\mathbb{V}[X]}{N}}\right]\right] \xrightarrow[n\to\infty]{} \int_{-a}^{a} e^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} \tag{1.3}$$

According to equation (1.4) , to build a 95% confidence level interval the length of which is 40% of $\mathbb{E}[X_1]$, a should be 2 and $\sqrt{\frac{\mathbb{V}[X]}{N}}$ a tenth of $\mathbb{E}[X]$.

$$\int_{-a}^{a} e^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = \begin{cases} 68\% & \text{if} \quad a = 1 \\ 95\% & \text{if} \quad a = 2 \\ 99.7\% & \text{if} \quad a = 3 \end{cases} \tag{1.4}$$

This standard deviation over expectation ratio is precisely what the empirical relative deviation estimates.

The estimator's expectation and variance can be combined

$$\mathbb{E}\left[\overline{X}_N\right] = \mathbb{E}[X] \tag{1.5}$$

$$\mathbb{V}\left[\overline{X}_N\right] = \frac{\mathbb{V}[X]}{N} \tag{1.6}$$

to notice that the CMC estimator has no bias and to define the empirical relative deviation accuracy measure

$$\rho\left(\overline{X}_N\right) = \frac{\sqrt{\mathbb{V}\left[\overline{X}_N\right]}}{\mathbb{E}\left[\overline{X}_N\right]} = \frac{1}{\sqrt{N}} \frac{\sqrt{\mathbb{V}[X]}}{\mathbb{E}[X]} \tag{1.7}$$

The empirical relative deviation establishes a link between N and $\mathbb{E}[X]$ and $\mathbb{V}[X]$ with respect to (w.r.t.) the estimation accuracy and the confidence interval as the smaller $\rho\left(\overline{X}_N\right)$, the more accurate the estimation. Eventually, the CLT provides both the SLLN and the empirical relative deviation with a convergence rate: $\frac{1}{\sqrt{N}}$.

In the making of confidence intervals, unknown variance $\mathbb{V}[X]$ is in practice substituted with its CMC proxy

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \approx \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \left(\frac{1}{m}\sum_{i=1}^{m} X_i\right)^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \overline{X}_N^2 \tag{1.8}$$

in the bound formula.

As seen, CMC exhibits convenient asymptotic features – biasless almost sure convergence – though its $\frac{1}{\sqrt{N}}$ convergence is quite slow, according to the CLT. A bit more about it can be said when N is fixed.

1.2.4 The CMC expectation estimator's law

The SLLN and CLT results hold when N, the total number of throws, grows infinitely large and they provide very useful information. In practice however, N is always finite. What can be said of $\overline{X}_N$ for a given fixed N?

Property 1.2.1 (CMC expectation estimator law)**.** *Let $(X_1, \cdots, X_N)$ be a set of N* iid *and integrable random variables ($\mathbb{E}[|X|] < \infty$) and $\overline{X}_N$ its empirical mean (Algorithm 1.2.1) designed to estimate $\mathbb{E}[X]$.*

$\overline{X}_N$ is a random variable itself. Assuming X_1 has density f_X w.r.t. the Lebesgue measure, $\overline{X}_N$ is distributed according to density

$$f_{\overline{X}_N} : \left| \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & N \underbrace{f_X * \cdots * f_X}_{N \text{ times}}(xN) \end{array} \right. \tag{1.9}$$

where $*$ stands for the convolution product: $g * h(x) = \int_{\mathbb{R}} g(z-x)\, h(z)\, \mathrm{d}z$.

In particular, if one estimates the probability that X lies in a given $\mathbb{X}$-subset $\mathbf{A}$ i.e. *$\mathbb{P}[X \in \mathbf{A}] = \mathbb{E}[\mathbf{1}_{\mathbf{A}}(X)]$,*

$$\forall k \in \mathbb{N}, \qquad \mathbb{P}\left[\overline{X}_N = \frac{k}{N}\right] = \mathrm{C}_N^k \mathbb{P}[X \in \mathbf{A}]^k (1 - \mathbb{P}[X \in \mathbf{A}])^{N-k} \tag{1.10}$$

$$\textit{whence} \qquad \mathbb{V}\left[\overline{X}_N\right] = \frac{\mathbb{P}[X \in \mathbf{A}](1 - \mathbb{P}[X \in \mathbf{A}])}{N} \tag{1.11}$$

as $N\overline{X}_N$ is distributed as $\mathcal{B}(N, \mathbb{P}[X \in \mathbf{A}])$ i.e. has binomial distribution with N throws and probability of success $\mathbb{P}[X \in \mathbf{A}]$.

Besides, the CMC estimator $\overline{X}_N$ is a random variable and is integrable or square-integrable as soon as X is. As a consequence, Crude Monte Carlo techniques can be used to assess its expectation and variance whenever affordable.

$$\mathbb{E}\left[\overline{X}_N\right] \approx \frac{1}{m}\sum_{i=1}^{m}\left(\overline{X}_N\right)_i, \qquad \mathbb{V}\left[\overline{X}_N\right] \approx \frac{1}{m}\sum_{i=1}^{m}\left(\overline{X}_N\right)_i^2 - \left(\frac{1}{m}\sum_{i=1}^{m}\left(\overline{X}_N\right)_i\right)^2 \tag{1.12}$$

One can therefore make CMC estimations of the expectation and variance of any given random estimator to proxy its empirical relative deviation:

$$\rho\left(\overline{X}_N\right) \approx \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\overline{X}_N\right)_i^2 - \left(\frac{1}{m}\sum_{i=1}^{m}\left(\overline{X}_N\right)_i\right)^2}}{\frac{1}{m}\sum_{i=1}^{m}\left(\overline{X}_N\right)_i} \tag{1.13}$$

The purpose at hands being estimating $\mathbb{P}[X \in \mathbf{A}]$, equation (1.11) will be focused on. Figure 1.1 shows a $\overline{X}_N$ density example in an easy case. Dealing with a rare event is trickier.

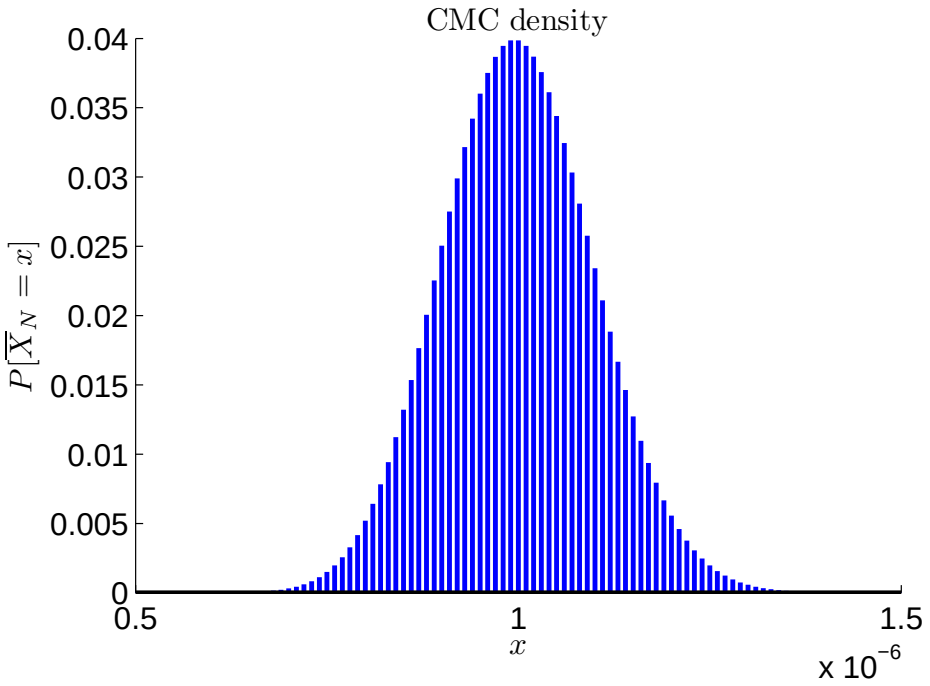


Figure 1.1: CMC probability estimator density with $N = 10^8$ and $\mathbb{P}[X \in \mathbf{A}] = 10^{-6}$.

1.2.5 The rare event case

Say we aim $\rho\left(\overline{X}_N\right) = 10\%$. Figure 1.2 shows what N should be for different values of $\mathbb{P}[X \in \mathbf{A}]$. With $\mathbb{P}[X \in \mathbf{A}] = 10^{-6}$, to ensure $\rho\left(\overline{X}_N\right) = 10\%$, N should be 10^8 (Figure 1.2). In such case, Figure 1.1 shows that CMC is an appropriate tool as small intervals around the targeted value have a high confidence level. Definition 1.3.1

$$\mathbb{P}\left[\overline{X}_N \in [0.9, 1.1] \cdot 10^{-6}\right] = 0.7065 \tag{1.14}$$
$$\mathbb{P}\left[\overline{X}_N \in [0.8, 1.2] \cdot 10^{-6}\right] = 0.9599 \tag{1.15}$$
$$\mathbb{P}\left[\overline{X}_N \in [0.7, 1.3] \cdot 10^{-6}\right] = 0.9976 \tag{1.16}$$

Besides, one can consider $\overline{X}_N$ law to be $\mathcal{N}\left(10^{-6}, \frac{10^{-6}\cdot\left(1-10^{-6}\right)}{10^8}\right)$ as stated in the CLT based on their cumulative distribution functions[1] (cdf) presented in Figure 1.3.

In practice however, such a simulation budget is very unlikely to be available. $N = 5 \cdot 10^5$

[1]Comparing cumulative distribution functions, as defined in Definition 1.3.1, to measure law convergence will be justified in Theorem 1.3.1.

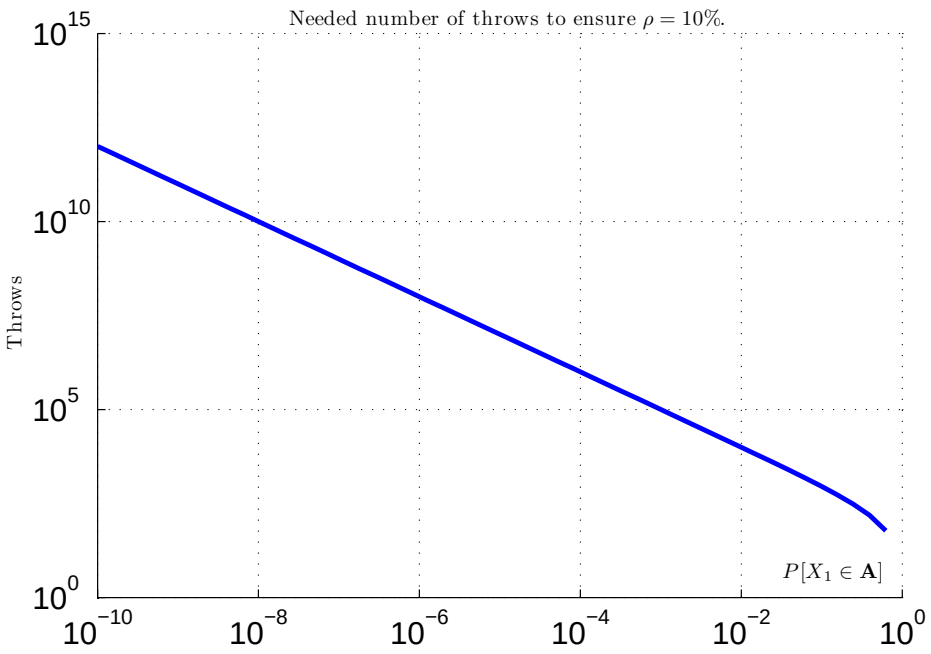


Figure 1.2: Needed number of throws to ensure 10% empirical relative deviation CMC expectation estimator for different values of $\mathbb{P}[X \in \mathbf{A}]$ with a loglog scale.

is a generous allocation. Then, the estimator law yields Figure 1.4 and

$$\mathbb{P}\left[\overline{X}_N = 0\right] \approx 60\% \tag{1.17}$$

$$\mathbb{P}\left[\overline{X}_N = \frac{1}{5 \cdot 10^5}\right] = 0\% \tag{1.18}$$

$$\mathbb{P}\left[\overline{X}_N = \frac{2}{5 \cdot 10^5}\right] \approx 30\% \tag{1.19}$$

$$\mathbb{P}\left[\overline{X}_N = \frac{4}{5 \cdot 10^5}\right] \approx 7\% \tag{1.20}$$

a more than 97% probability that the estimator misses its target value by at least 100% of it. Likewise, the convergence predicted by the CLT is still far as can be seen on Figure 1.5. CMC is not the tool to use in such conditions. Besides, the CMC estimator imposes an $\frac{1}{N}$ increment precision which can not be enough in some cases.

Generally speaking, as soon as $\mathbb{P}[X \in \mathbf{A}]$ or the needed precision is small w.r.t. $\frac{1}{N}$, CMC is not the tool to use to estimate a probability. Likely, given a simulation budget, there is a limit to how extreme a quantile can be estimated via CMC.

1.3 Quantile estimation

Assuming X is on the real line $\mathbb{R}$, q the α level quantile (α-quantile in short) is defined by

$$\forall \alpha \in [0,1], q = \inf_{v \in \mathbb{R}} \{\mathbb{P}[X \leq v] \geq \alpha\} \tag{1.21}$$

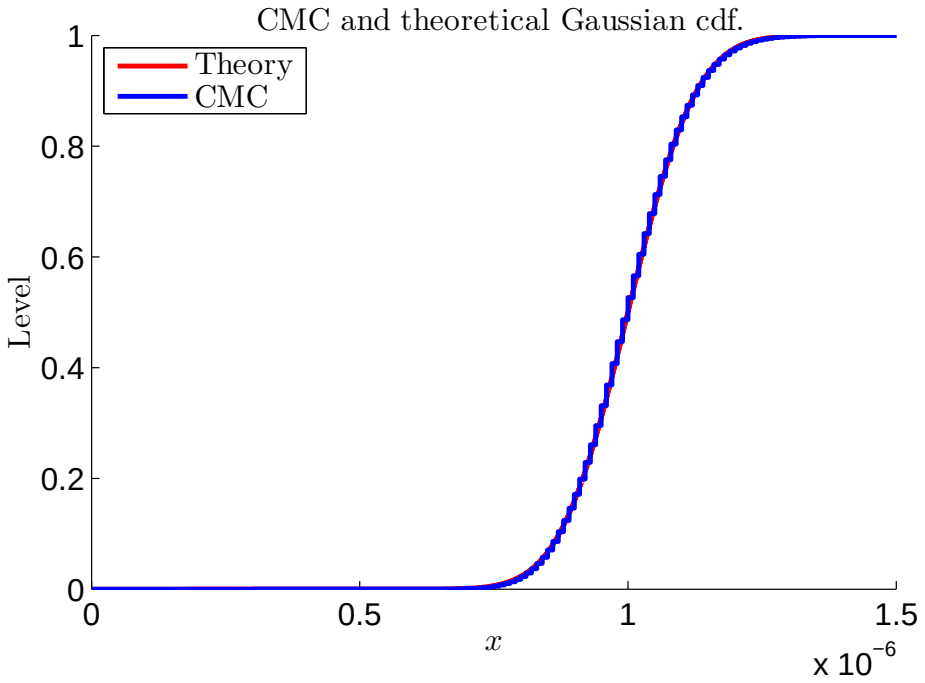


Figure 1.3: CMC probability estimator cdf with $N = 10^8$ and $\mathbb{P}[X \in \mathbf{A}] = 10^{-6}$ and the limiting $\mathcal{N}\left(10^{-6}, \frac{10^{-6}\cdot\left(1-10^{-6}\right)}{10^8}\right)$ Gaussian's cdf.

CMC can be used to estimate a quantile as well, taking advantage of the SLLN again.

1.3.1 The cumulative distribution function

Let us define the cumulative distribution function (cdf).

Definition 1.3.1 (Cumulative distribution function)**.** *Let X be a random variable on the $\mathbb{R}$ line. Its cumulative distribution function (cdf) is*

$$\forall x \in \mathbb{R}, F_X(x) = \mathbb{P}[X \leq x] \tag{1.22}$$

The relationship between quantile and cdf is then obvious:

$$\forall \alpha \in [0,1], q = \inf_{v \in \mathbb{R}} \{F_X(v) \geq \alpha\} \tag{1.23}$$

Given F_X, one can easily calculate any quantile. Unfortunately, in practice, cdfs are unknown and have to be estimated before estimating the quantiles.

Equation (1.22) suggests to use the SLLN to approximate F_X. This is precisely what the empirical cumulative distribution function is about.

Definition 1.3.2 (Empirical cumulative distribution function)**.** *Let $(X_1, \cdots, X_N)$ be a set of N* iid *and integrable random variables ($\mathbb{E}[|X|] < \infty$) on the $\mathbb{R}$ line. According to the SLLN and equation (1.22), the following approximation holds.*

$$\forall x \in \mathbb{R}, \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\{X_k \leq x\}} \xrightarrow[N\to\infty]{a.s.} F_X(x) \tag{1.24}$$

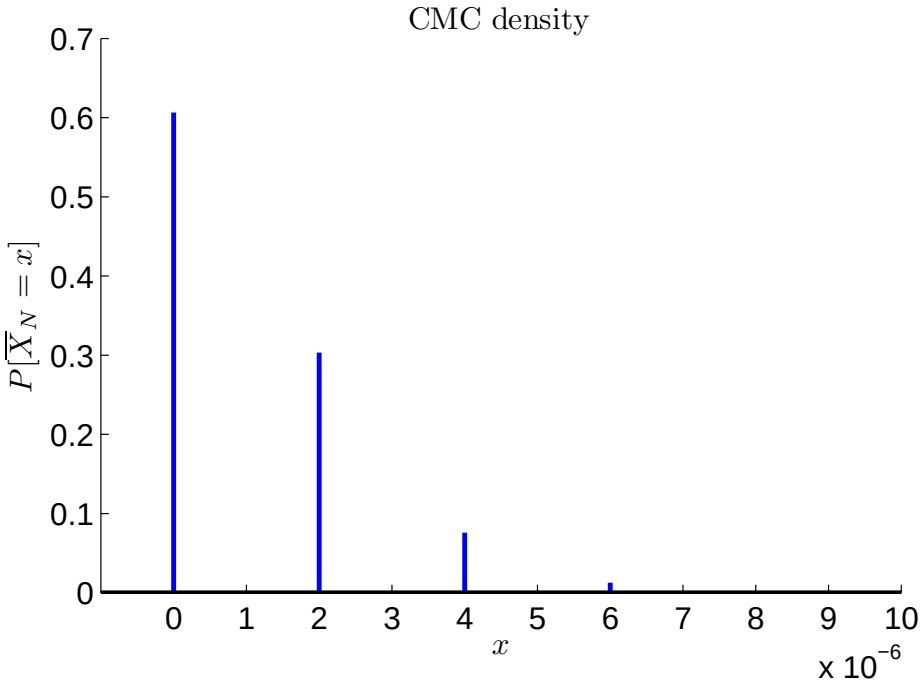


Figure 1.4: CMC probability estimator density with $N = 5 \cdot 10^5$ and $\mathbb{P}[X \in \mathbf{A}] = 10^{-6}$.

Then, the empirical cumulative distribution function (ecdf), defined as follows,

$$\forall x \in \mathbb{R}, \overline{F}_X(x) \equiv \overline{F}_X(x, N) = \frac{1}{N} \sum_{k=1}^{N} \mathbf{1}_{\{X_k \leq x\}} \approx \mathbb{P}[X \leq x] \tag{1.25}$$

yields almost sure (a.s.) pointwise convergence toward F_X.

$$\forall x \in \mathbb{R}, \overline{F}_X(x, N) \xrightarrow[N \to \infty]{a.s.} F_X(x) \tag{1.26}$$

A $\overline{F}_X$ variation exists and can be analysed just as pleasantly. For clarity sake, its description is postponed until Section 2.3.

The ecdf is a valuable choice of cdf estimator as stated in the next proposition.

Proposition 1.3.1 (Convergence of the ecdf toward the cdf)**.** *Let $(X_N, N \in \mathbb{N}^*)$ be a sequence of real random variables and $(F_N, N \in \mathbb{N}^*)$ the sequence of their cdfs.*

If $X_N \xrightarrow{d} X$, then $F_N \xrightarrow{d} F_X$ except where X cdf F_X is not continuous.

Reciprocally, assume $F_N \xrightarrow{d} F_X$ where F_X has right limits, except when not continuous and $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$. Then F_X is the cdf of a random variable X and $X_N \xrightarrow{d} X$.

One can thus have the idea to check law convergence through cdfs, as done on Figure 1.3, and even to test it such as via the Kolmogorov-Smirnov test [20]. In the situation at hand, quantiles are the target: what can be done about them for fixed N?

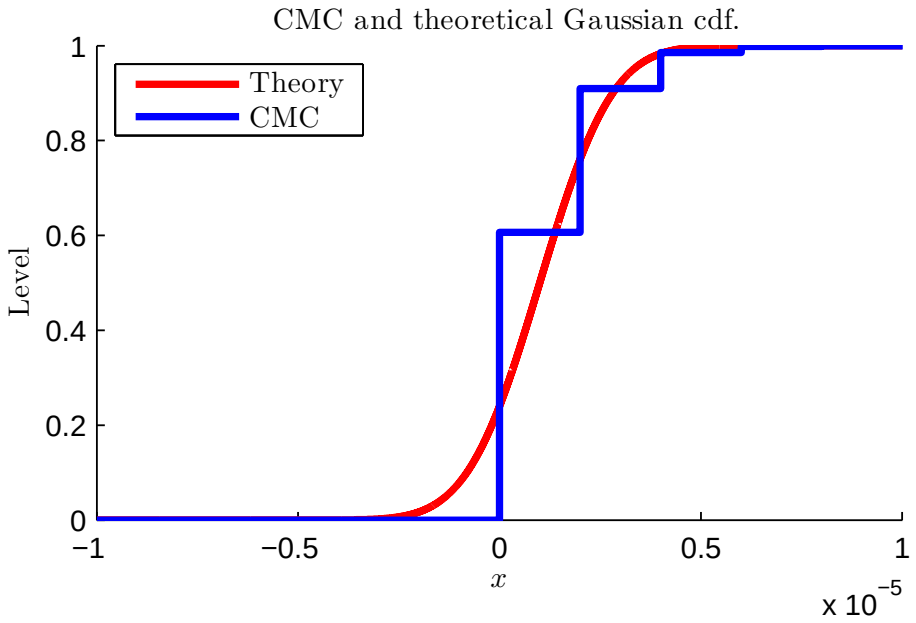


Figure 1.5: CMC probability estimator cdf with $N = 5 \cdot 10^5$ and $\mathbb{P}[X \in \mathbf{A}] = 10^{-6}$ cdf and the limiting $\mathcal{N}\left(10^{-6}, \frac{10^{-6}\cdot\left(1-10^{-6}\right)}{5\cdot 10^5}\right)$ Gaussian's.

1.3.2 CMC quantile estimator

The CMC method provides a direct way to estimate $q(\alpha)$, without even building $\overline{F}_X$ in practice.

Algorithm 1.3.1 (CMC quantile estimator)**.** *To estimate the cdf of random variable X distributed on the $\mathbb{R}$ line proceed as follows.*

1. *Generate the* iid *sample set $(X_1, \cdots, X_N)$ such that $\forall i, X_i \sim X$.*
2. *Sort them in ascending order to form $\left(X_{(1)}, \cdots, X_{(N)}\right)$.*
3. *Approximate*[2] *$q(\alpha)$ with $\overline{q} = X_{(\lfloor \alpha N \rfloor + 1)}$.*

This only requires being able to simulate iid *throws of X.*

However, three theoretical questions involving $\overline{F}_X$ are in order:

1. How close is $\overline{F}_X$ to F_X?
2. How close is $\overline{q}$ to q? What about choosing any other $\overline{q}$ in $\left[X_{(\lfloor \alpha N \rfloor)}, X_{(\lfloor \alpha N \rfloor + 1)}\right]$?
3. What is the asymptotic behaviour of $\overline{q}$?

They will be answered in this very same order.

[2] $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are respectively the floor and ceiling functions.

1.3.3 The empirical cumulative function error

$\overline{F}_X$ distance to F_X needs attention, both locally and globally.

1.3.3.1 Global results

The Glivenko-Cantelli theorem gives a first global precision result, proving that $\overline{F}_X$ yields almost sure uniform convergence toward F_X.

Theorem 1.3.2 (Glivenko-Cantelli). *Let $(X_1, \cdots, X_N)$ be a set of N iid random variables on the $\mathbb{R}$ line with cdf F_X. Then, almost surely*

$$\lim_{N\to\infty} \sup_{x\in\mathbb{R}} |\overline{F}_X(x, N) - F_X(x)| = 0 \tag{1.27}$$

where $\overline{F}_X$ is the ecdf of the set.

The Dvoretzky-Kiefer-Wolfowitz (DKW) inequality states how fast this uniform convergence takes place. Unfortunately, the given upper bound answers "Not so fast".

Theorem 1.3.3 (Dvoretzky-Kiefer-Wolfowitz inequality). *Let $(X_1, \cdots, X_N)$ be a set of N iid random variables on the $\mathbb{R}$ line with continuous[3] cdf F_X. Then, almost surely*

$$\forall \epsilon > 0, \mathbb{P}\left[\sup_{x\in\mathbb{R}} |\overline{F}_X(x) - F_X(x)| \geq \epsilon\right] \leq 2e^{-2N\epsilon^2} \tag{1.28}$$

where $\overline{F}_X$ is the ecdf of the set.

If the lowest level of interest is 10^{-4} or $1 - 10^{-4}$ and the admissible relative error is 10%, ϵ can be set to 10^{-5}. Then, to ensure $2e^{-2N\epsilon^2} = 0.05$ *i.e.* a 95% confidence level, one has to choose $N > 1.8 \cdot 10^{10}$! High global accuracy comes with a high price according to the loose upper bound.

1.3.3.2 Local results

When it comes to a local approach, the CLT holds.

$$\forall N \in \mathbb{N}, \forall x \in \mathbb{R}, \sqrt{N}\left(\overline{F}_X(x) - F_X(x)\right) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, F_X(x)(1 - F_X(x))) \tag{1.29}$$

The Berry-Esseen theorem ([7]-[34]) is an attempt at measuring the precision for a fixed N.

Theorem 1.3.4 (Berry-Esseen theorem). *Let $(X_1, \cdots, X_N)$ be a set of N iid random variables on the $\mathbb{R}$ line such that $\mathbb{E}[|X|^3] < \infty$. Then, denoting $m = \mathbb{E}[X]$, $\sigma^2 = \mathbb{V}[X]$ and $m_3 = \mathbb{E}[|X - m|^3]$*

$$\forall a \in \mathbb{R}, N \geq 1, \left|\mathbb{P}\left[\sqrt{N}\frac{\overline{X}_N - m}{\sigma} \leq a\right] - \int_{-\infty}^{a} e^{-x^2/2}\frac{\mathrm{d}x}{\sqrt{2\pi}}\right| \leq \frac{Cm_3}{\sigma^3\sqrt{N}} \tag{1.30}$$

where the constant C does not depend on neither a nor X.

[3] This is in practice a plausible hypothesis.

It is known that $0.40973 \approx \frac{\sqrt{10}+3}{6\sqrt{2\pi}} \leq C < 0.4784$ [50] and this theorem still holds when σ is replaced with its empirical estimator σ_N in the majored probability according to [91].

$$\sigma_N^2 = \frac{1}{N}\sum_{k=1}^{N} X_k^2 - \left(\frac{1}{N}\sum_{k=1}^{N} X_k\right)^2 \tag{1.31}$$

The same formula (1.30) can be applied in a pointwise fashion to the ecdf, with $\forall x \in \mathbb{R}$, $m = F_X(x)$, $\sigma = F_X(x)(1 - F_X(x))$ and $\overline{F}_X(x)$ in $\overline{X}_N$ stead. In any case, the upper bound depending not on the random variable law and σ and m_3 being unknown beforehand, this result is not of much help in practice.

1.3.4 Quantile estimator to real quantile mean spread

The Algorithm 1.3.1 estimates q via $\overline{q} = X_{(\lfloor \alpha N \rfloor + 1)}$. What can be said about this random variable? This is a matter of order statistics. [1] gives the expectation of $\overline{q}$ assuming F_X is regular.

Property 1.3.1 (CMC quantile estimator expectation and variance)**.** *Let $\overline{q}$ be CMC estimator of the α–quantile q of random variable X as defined in Algorithm 1.3.1. Assuming F_X has a non zero derivative at q and Taylor expansion in this point, $\overline{q}$ has the following expectation and variance*

$$\mathbb{E}[\overline{q}] = q + \frac{\alpha(1-\alpha)}{2(N+2)} \cdot F_X^{-1(2)}(\alpha) + \mathcal{O}\left(\frac{1}{(N+2)^2}\right) \tag{1.32}$$

$$\mathbb{V}[\overline{q}] = \frac{\alpha(1-\alpha)}{N+2} \cdot \left(F_X^{-1(1)}(\alpha)\right)^2 + \mathcal{O}\left(\frac{1}{(N+2)^2}\right) \tag{1.33}$$

where $F_X^{-1(j)}(\alpha)$ is the j^{th} order derivative of the inverse mapping of F_X at point α.

This estimator has therefore no bias *asymptotically*. One can ask for the $\frac{1}{(N+2)^2}$ coefficients: they depend on the derivatives in q of F_X which are usually unknown in practice. However, it might not be the real concern as this bias goes to zero with a $\frac{1}{N}$ rate while the standard deviation $\sqrt{\mathbb{V}[\overline{q}]}$ does it with a $\frac{1}{\sqrt{N}}$ rate. Hence, the standard deviation is often the issue.

Unaccustomedly however, $\mathbb{V}[\overline{q}]$ is not involved in the building of the confidence intervals. As a matter of facts, they are even independent of F_X and can be derived for a fixed N [16].

Property 1.3.2 (CMC quantile estimator confidence intervals)**.** *Let $\left(X_{(1)}, \cdots, X_{(N)}\right)$ be a set of N random variables* iid *according to X on the $\mathbb{R}$ line sorted in ascending order. Let q be the α-quantile of random variable X with $\alpha \in]0,1[$. Then, regardless to F_X, and for any rank i,j such that $1 \leq i < j \leq N$,*

$$\mathbb{P}\left[q \in \left[X_{(i)}, X_{(j)}\right]\right] = \sum_{r=i}^{j-1} \frac{N!}{r!(N-r)!} \alpha^r (1-\alpha)^{N-r} \tag{1.34}$$

An easily implementable practice to build up a c–level confidence interval is choosing i and j so that $\mathbb{P}\left[q \in \left[X_{(i)}, X_{(j)}\right]\right] \geq c$ and so that $\left[X_{(i)}, X_{(j)}\right]$ is as small as possible. The law of random variable $W_{j,i} = \left(X_{(j)} - X_{(i)}\right)$ is stated in [89] but requires a too cumbersome integration to be of practical interest. Besides, with respect to this confidence interval structure, all values in $\left]\, X_{(\lfloor \alpha N \rfloor)}, X_{(\lfloor \alpha N \rfloor + 1)}\, \right[$ are equivalent.

However useful these fixed N results are, an asymptotic approximation to $\overline{q}$ is of interest, at least in an attempt to enjoy a Gaussian comfort similar to the one enjoyed by the expectation estimator.

1.3.5 Asymptotic behaviour of the CMC quantile estimator

Conveniently enough, there is a Gaussian proxy to $\overline{q}$ when $N \to \infty$.

Theorem 1.3.5 (Asymptotic behaviour of the CMC quantile estimator)**.** *Let $(X_1, \cdots, X_N)$ be a set of N random variables that are independent and identically distributed (*iid*) as integrable random variable X $(\mathbb{E}[|X|] < \infty)$ with cdf F_X. Let $\alpha_1, \cdots, \alpha_k$ be k reals such that $0 < \alpha_1 < \cdots < \alpha_k < 1$. Then, for non zero integer N define the $r_1(N), \cdots, r_k(N)$ integers via*

$$\forall i \in \{1, \cdots, k\}, r_i(N) = \lceil \alpha_i N \rceil + 1 \tag{1.35}$$

Furthermore, denote $q_{\alpha_1}, \cdots, q_{\alpha_k}$ the F_X α_i–quantiles and assume them all to be unique and such that F_X has strictly positive finite derivatives for them

$$\forall i \in \{1, \cdots, k\}, 0 < F_X^{(1)}(q_{\alpha_i}) < \infty \tag{1.36}$$

Then, the random vector U_N defined by

$$U_N = \sqrt{N}\left(X_{(r_1(N))} - q_{\alpha_1}, \cdots, X_{(r_k(N))} - q_{\alpha_k}\right) \tag{1.37}$$

converges in distribution towards the $\mathcal{N}_k(0, \Delta)$ Gaussian law, where the Δ matrix is symmetrical and defined by

$$\forall i, j \in \{1, \cdots, k\} \text{ such that } i \leq j, \Delta_{ij} = \frac{\alpha_i (1 - \alpha_j)}{F_X^{(1)}(q_{\alpha_i}) F_X^{(1)}\left(q_{\alpha_j}\right)} \tag{1.38}$$

This result requires more hypotheseses than the CLT and unlike it directly relies on $F_X^{(1)}$, often making it impossible to use. However, it might come really handy in the happy few cases when it holds.

For instance, Theorem 1.3.5 holds if X has a dimension 2 Rayleigh distribution $\mathcal{R}_2$, that is to say is distributed as the Euclidean norm of a $\mathcal{N}(0_2, I_2)$. The theorem can in this case help forecast the budget needed to achieve desired accuracy.

$$\begin{aligned} \forall q \in \mathbb{R}, F_X(q) &= \mathbf{1}_{\mathbb{R}^+}(q)\left(1 - e^{-\frac{q^2}{2}}\right) \\ \forall \alpha \in \left]\, 0, 1\, \right[, F_X^{-1}(\alpha) &= \sqrt{-2\ln(1-\alpha)} \\ \forall r \in \mathbb{R}, F_X^{(1)}(r) &= \mathbf{1}_{\mathbb{R}^+}(r) r e^{-\frac{r^2}{2}} \end{aligned} \tag{1.39}$$

Then, with $\alpha = 10^{-1}$, and aiming $\frac{\Delta}{\sqrt{N}} = \frac{q_\alpha}{10}$

$$q_\alpha = 0.4590 \qquad F_X^{(1)}(q_\alpha) = 0.4131 \qquad \Delta = 0.6254 \qquad N \geq 186 \tag{1.40}$$

This N value is to be compared with the number of throws needed to estimate $\mathbb{P}[X \leq q_\alpha]$ with a 10% empirical relative deviation as stated in (1.11): 90. Estimating a quantile can therefore be costlier than estimating an expectation and stands as a problem of its own.

1.3.6 The extreme quantile case

The quantile is said extreme when its level α goes to 0 (or 1) and the associated event $\{X < q\}$ (or $\{q < X\}$) becomes rare. In practice, as soon as the desired level precision is smaller than $\frac{1}{N}$, the quantile is extreme with respect to the estimation budget if used with Algorithm 1.3.1. However, what really matters in CMC quantile estimation accuracy is $F_X^{(1)}$, which is unknown and usually unreachable. Therefore, and it is the usual practice in later mentioned techniques as well, most of the research attention is given to probability estimation. Quantile estimation is performed adapting probability estimation algorithms.

1.4 Conclusion

Crude Monte Carlo is the usual way when estimating expectations and quantiles, for it is conveniently usable and delivers in most frequent cases. However, it fails to deal with rare events or extreme quantiles with a reasonable simulation cost. CMC's calculation power and budget being depleted by always-increasing requirements triggered a quest for rare event dedicated tools.

The two main probabilistic ways are Importance Sampling –Chapter 2– and Importance Splitting –Chapter 3–. The former uses a probability measure change but keep the CMC practical aspects while the later rephrases the desired probability and proceeds iteratively.

Chapter 2

The Importance Sampling Technique

Importance Sampling (IS) is an attempt at modifying CMC so as to deal properly with a given rare event case [6, 10, 30] and more generally stands as a variance reduction technique. Consider any transfer function ϕ with a random input $X \sim \mu$. IS takes advantage on the integral representation of the expectation of the random output to perform a change of measure. Actually, IS harvests all the available information to propose a hopefully better suited input space sampling random variable $Z \sim \nu$ called auxiliary random variable.

$$\begin{aligned} \mathbb{E}\left[\phi\left(X\right)\right] &= \int_{\mathbb{X}} \phi\left(x\right) \mu\left(\mathrm{d}x\right) \\ &= \int_{\mathbb{X}} \phi\left(z\right) \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\left(z\right) \nu\left(\mathrm{d}z\right) \\ \mathbb{E}\left[\phi\left(X\right)\right] &= \mathbb{E}\left[\phi\left(Z\right) \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\left(Z\right)\right] \end{aligned} \tag{2.1}$$

Formally, the only requirement so that $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ exists via the Radon-Nikodym theorem is that probability measure ν dominates μ.

$$\left[\mu \ll \nu\right] \Leftrightarrow \left[\forall \mathbf{A} \in \mathcal{X}, \left[\nu\left(\mathbf{A}\right) = 0\right] \Rightarrow \left[\mu\left(\mathbf{A}\right) = 0\right]\right] \tag{2.2}$$

This ensures that ν generates all the events μ does but with another probability, purposely making rare events more frequent. The underlying hope is that with a good choice of ν, the IS estimator, which is unbiased, *can* yield reduced variance with respect to CMC, keeping its convenient implementation.

1. How does the IS estimator variance compare with the CMC's?
2. What is a good choice of the auxiliary probability measure ν?
3. How to sample the chosen auxiliary random variable Z?

Elements of answers follow alike.

A few words about notations first. The random variable $\phi\left(X\right)$ will be henceforth denoted Y as well. Besides, if the probability measures μ and ν are assumed absolutely continuous with respect to a common nonnegative measure, which usually is the Lebesgue measure, denoted λ

-or often just $\mathrm{d}x$-, then μ and ν have respective probability density functions (pdf) f_X and f_Z with respect to λ.

$$\mu(\mathrm{d}x) = f_X(x)\,\lambda(\mathrm{d}x) \qquad \nu(\mathrm{d}x) = f_Z(x)\,\lambda(\mathrm{d}x) \tag{2.3}$$

Then, the Radon-Nikodym derivative $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ is $\frac{f_X}{f_Z}$ and (2.1) becomes

$$\begin{aligned}\mathbb{E}[\phi(X)] &= \int_{\mathbb{X}} \phi(x)\, f_X(x)\,\lambda(\mathrm{d}x) \\ &= \int_{\mathbb{X}} \phi(z)\,\frac{f_X(z)}{f_Z(z)} f_Z(z)\,\lambda(\mathrm{d}z) \\ \mathbb{E}[\phi(X)] &= \mathbb{E}\left[\phi(Z)\,\frac{f_X(Z)}{f_Z(Z)}\right]\end{aligned} \tag{2.4}$$

For both clarity and simplicity sakes and without loss of generality, $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ will be denoted w.

2.1 Auxiliary density design for the IS expectation estimator

Being of Monte-Carlo type, the Importance Sampling expectation estimator $\widehat{Y}_N$ is a random variable. Its performance depends on how appropriately chosen the auxiliary density is.

2.1.1 Importance Sampling expectation estimator

Equation (2.1) suggests a direct IS algorithm to estimate an expectation via the SLLN.

Algorithm 2.1.1 (IS expectation estimator)**.** *To estimate the expectation of the integrable random variable $Y = \phi(X)$ proceed as follows.*

1. *Choose or design ν w.r.t. which μ is absolutely continuous.*
2. *Generate the* iid *sample set $(Z_1, \cdots, Z_N)$ such that $\forall i, Z_i \sim \nu$.*
3. *Use the transfer and density functions to build $(\phi(Z_1), \cdots, \phi(Z_N))$ and $(\mathrm{w}(Z_1), \cdots, \mathrm{w}(Z_N))$.*
4. *Calculate the empirical mean $\widehat{Y}_N = \sum_{k=1}^{N} \mathrm{w}(Z_k)\,\phi(Z_k)$.*
5. *Approximate $\mathbb{E}[Y] \approx \widehat{Y}_N$.*

This requires the two following points

1. *Ability to simulate* iid *throws of Z,*
2. *Ability to evaluate* w *and ϕ at any point.*

to be performed.

The seducing part of this IS Algorithm 2.1.1 is its *trompe l'œil* closeness to the comfort of its CMC counterpart Algorithm 1.2.1. A closer looks quickly unveils two issues.

First, when ϕ is an indicator function, the IS probability estimator can return values greater than one as its formulation allows so. With $1 - 10^{-5} = \mathbb{E}\left[\mathbf{1}_{\{\phi(X)\leq q\}}\right] = \mathbb{E}\left[\mathbf{1}_{\{\phi(Z)\leq q\}}\,\mathrm{w}\,(Z)\right]$, it only takes little variance. Such values are difficult to interpret and could phase an unaware user. A practical solution is replacing the number of sample points in the normalising denominator with the weight sum. This can be useful as well when either f_X or f_Z is only known up to its normalising constant. It costs a little bias that is asymptotically negligible, but makes the analysis more complex [45, 81]. As using the sample size is advised in [41] when it comes to small probability estimation, the weight normalised estimators are not mentioned in this chapter until Section 2.3 which deals with quantile estimation.

Second is the major difficulty of Importance Sampling: the sampling density. Actually, designing and sampling according to ν can be a challenge as it may or may not, yield a reduced variance with respect to CMC estimator $\overline{Y}_N$.

2.1.2 Variance and optimal auxiliary density

$\widehat{Y}_N$ is designed so as to have no bias w.r.t. $\mathbb{E}[Y]$.

$$\mathbb{E}\left[\widehat{Y}_N\right] = \mathbb{E}\left[(\mathrm{w}\phi)\,(Z)\right] = \int \phi\,(z)\,\mathrm{w}\,(z)\,\nu\,(\mathrm{d}z) = \mathbb{E}\left[Y\right] \tag{2.5}$$

Its variance is to be compared to that of the CMC estimator $\overline{Y}_N$.

$$\begin{aligned}
\mathbb{V}\left[\widehat{Y}_N\right] &= \frac{\mathbb{E}\left[(\mathrm{w}\phi)^2\,(Z)\right] - \mathbb{E}\left[\phi\mathrm{w}\,(Z)\right]^2}{N} && (2.6)\\
&= \frac{\int (\mathrm{w}\phi^2\,(z))\,\mathrm{w}\,(z)\,\nu\,(\mathrm{d}z) - \mathbb{E}\left[\phi\,(X)\right]^2}{N} && (2.7)\\
&= \frac{\mathbb{E}\left[\mathrm{w}\phi^2\,(X)\right] - \mathbb{E}\left[\phi\,(X)\right]^2}{N} && (2.8)\\
&= \mathbb{V}\left[\overline{Y}_N\right] + \frac{\mathbb{E}\left[\phi^2\,(X)\,(\mathrm{w}\,(X)-1)\right]}{N} && (2.9)
\end{aligned}$$

According to equation (2.9),

$$\left[\mathbb{V}\left[\widehat{Y}_N\right] \leq \mathbb{V}\left[\overline{Y}_N\right]\right] \Leftrightarrow \left[\mathbb{E}\left[\phi^2\,(X)\,(\mathrm{w}\,(X)-1)\right] \leq 0\right] \tag{2.10}$$

However the ultimate goal is minimising $\mathbb{V}\left[\widehat{Y}_N\right]$ as stated in equation (2.8). Thus, using Jensen inequality in equation (2.11),

$$\mathbb{E}\left[\left(\mathrm{w}\phi^2\right)(X)\right] = \mathbb{E}\left[(\mathrm{w}\phi)^2\,(Z)\right] \geq \left(\mathbb{E}\left[(\mathrm{w}|\phi|)\,(Z)\right]\right)^2 = \left(\mathbb{E}\left[|\phi|\,(X)\right]\right)^2 \tag{2.11}$$

and as there is equality if and only if random variable $\mathrm{w}|\phi|\,(Z)$ is almost surely constant, one can conclude

$$\left[\mathbb{V}\left[\widehat{Y}_N\right] \text{ is minimal.}\right] \Leftrightarrow \left[\forall \mathbf{A} \in \mathcal{X}, \nu\,(\mathbf{A}) = \frac{(|\phi|\mu)\,(\mathbf{A})}{\mu\,(|\phi|)} \equiv \nu^*\,(\mathbf{A})\right] \tag{2.12}$$

Unfortunately, sampling according to ν^* directly, as in the ideal case represented in Figure 2.1, comes as both impossible and pointless as it requires a normalizing constant ($\mu(|\phi|)$) as difficult to calculate as the very sought value. This, nonetheless gives a hint about how to design the auxiliary measure ν: as similar to ν^* as possible.

Considering $Z \sim \nu$ and $Z^* \sim \nu^*$, comparing induced variances and using the optimal change of measure definition to calculate $\frac{\mathrm{d}\mu}{\mathrm{d}\nu^*} = \mathbf{1}_{\{|\phi|\neq 0\}} \frac{\mu(|\phi|)}{|\phi|}$,

$$\mathbb{V}[\mathrm{w}\phi(Z)] - \mathbb{V}[\mathrm{w}^*\phi(Z^*)] = \int \left(\phi \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)^2 \nu - (\mathbb{E}[|\phi|(X)])^2 \tag{2.13}$$

$$= \int \left(\phi \frac{\mathrm{d}\mu}{\mathrm{d}\nu^*} \frac{\mathrm{d}\nu^*}{\mathrm{d}\nu}\right)^2 \nu - (\mathbb{E}[|\phi|(X)])^2 \tag{2.14}$$

$$= (\mathbb{E}[|\phi|(X)])^2 \int \left(\frac{\mathrm{d}\nu^*}{\mathrm{d}\nu}\right)^2 \nu - (\mathbb{E}[|\phi|(X)])^2 \tag{2.15}$$

$$= (\mathbb{E}[|\phi|(X)])^2 \int \left(\left(\frac{\mathrm{d}\nu^*}{\mathrm{d}\nu}\right)^2 - 1\right) \nu \tag{2.16}$$

$$= (\mathbb{E}[|\phi|(X)])^2 \chi^2(\nu^*, \nu) \tag{2.17}$$

one can say more: the χ^2 divergence between ν^* and ν should be as small as possible. This is however of limited operational interest for ν^* is unknown. Because of this, as exposed in Section 2.2, practical algorithms make no obvious usage of this result.

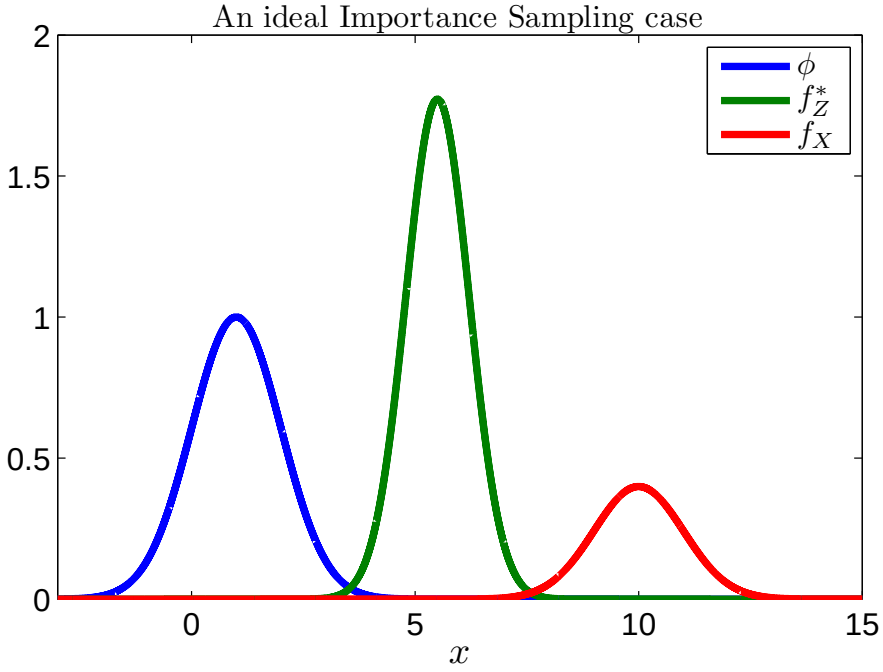


Figure 2.1: An ideal Importance Sampling expectation estimation configuration: given transfer function ϕ and input density f_X, the optimal change of measure f_Z^* is calculated.

2.1.3 When to use an auxiliary random variable?

ν^* is the optimal auxiliary probability measure the input measure can be substituted with. But why not directly substituting the output probability measure ($Y \sim \eta$) with an auxiliary measure? A very practical answer is that often too limited information is known about Y as the transfer function ϕ is too complicated to derive anything about η from it and μ.

However, if ϕ is a function composition, say

$$\phi = \phi_2 \circ \phi_1 \qquad Y_m = \phi_1(X) \qquad Y = \phi_2(Y_m) \tag{2.18}$$

IS can be applied to either input X or to intermediate random variable Y_m or output Y. Given the choice, when to apply IS?

[13] comes with an answer[1]: the closer to the output, the better *i.e.* the lesser variance. Applying IS to Y results in smaller variance than applying to Y_m and applying IS to Y_m results in smaller variance than applying it to X. From now on, it will be assumed that IS is applied as close to the output as possible.

2.2 Approximating the optimal auxiliary density

Though impossible to use directly, ν^* advocates to sample where both $|\phi|$ and μ are important so that simulated points really contribute to the estimation. Taking advantage of all the available information, ν is designed in compliance with this objective.

Furthermore, for it is usually the case in practice and assumed in hereinafter mentioned techniques, X and Z are assumed absolutely continuous with respect to the Lebesgue measure λ, hence, as already stated

$$\mu(\mathrm{d}x) = f_X(x)\,\lambda(\mathrm{d}x) \qquad \nu(\mathrm{d}x) = f_Z(x)\,\lambda(\mathrm{d}x) \tag{2.3}$$

Then, the associated optimal density can be expressed as

$$\forall x \in \mathbb{X},\ f_Z^*(x) = \frac{|\phi|(x)\, f_X(x)}{\mathbb{E}\left[|\phi|(X)\right]} \tag{2.19}$$

2.2.1 Translation and scaling

Two easily implementable thus well-spread ways are scaling and translating the natural density f_X:

- Density scaling: sampling according to $f_Z : z \mapsto \frac{1}{a} f_X\left(\frac{z}{a}\right)$, where a is the variance scaling parameter allows to simulate either close to or further from the input's mean.
- Density translation: sample according to $f_Z : z \mapsto f_X(z - m)$ to shift the input mean.

Scaling and translating benefit from the knowledge of the interesting subsets of the input space $\mathbb{X}$. Such knowledge is usually unavailable in practice and one can not ignoring the most likely unknown $|\phi|$. Therefore, translating and scaling are of very limited practical use.

[1] Section 4.3 page 63.

2.2.2 Cross-Entropy

The Cross-Entropy method (CE) aims at choosing f_Z within a parametric family of densities so as to minimise its Kullback-Leibler divergence to f_Z^* *i.e.* $\mathrm{KL}\left(f_Z^*, f_Z\right)$ [11, 80].

$$\mathrm{KL}\left(f_Z^*, f_Z\right) = \mathbb{E}\left[\ln\left(\frac{f_Z^*\left(Z^*\right)}{f_Z\left(Z^*\right)}\right)\right] = \int f_Z^*\left(x\right)\ln f_Z^*\left(x\right)\lambda\left(\mathrm{d}x\right) - \int f_Z^*\left(x\right)\ln f_Z\left(x\right)\lambda\left(\mathrm{d}x\right) \quad (2.20)$$

This operator comes from Shannon's information theory [24, 85] and can be here understood as the expected extra simulation needed when using f_Z instead of f_Z^* to perform an estimation.

The optimisation problem can be stated as the search for the optimal parameter θ^* within the parameter set[2] Θ. Equivalently, the best probability density change, in the cross-entropy sens, is any $f_Z^{\theta^*}$ that minimises the Kullback-Leibler divergence between the parametric density family $\left\{f_Z^\theta\right\}_{\theta\in\Theta}$ and f_Z^*, as represented in Figure 2.2.

$$f_Z = f_Z^\theta \qquad f_Z^* = \frac{|\phi| f_X}{\mathbb{E}\left[|\phi|\left(X\right)\right]} \qquad \theta^* = \arg\max_{\theta\in\Theta}\mathbb{E}\left[|\phi|\left(X\right)\ln f_Z^\theta\left(X\right)\right] \quad (2.21)$$

The optimisation strategy choice depends on ϕ and so does the whole algorithm. Assume $\phi\left(\cdot\right) = \mathbf{1}_{[s\,,\infty[}\left(\cdot\right)$ *i.e.* the desired quantity is a threshold exceedance probability.

To find θ^* so as to estimate $\mathbb{P}\left[X \in [s\,,\infty\,[\right]$, the Cross-Entropy strategy goes as follows. Instead of hardly maximising $\mathbb{E}\left[\mathbf{1}_{[s\,,\infty[}\left(X\right)\ln f_Z^\theta\left(X\right)\right]$ directly, an increasing sequence of thresholds is defined in an adaptive way as further detailed empirical X–quantiles.

$$s_0 < s_1 < \cdots < s_m \quad (2.22)$$

Then, successively, appropriate parameters are found solving intermediate easier optimisation problems

$$\theta^{k+1} = \arg\max_{\theta\in\Theta}\mathbb{E}\left[\mathbf{1}_{[s_{k+1}\,,\infty[}\left(Z^\theta\right)\ln f_Z^\theta\left(Z^\theta\right)\right] \quad (2.23)$$

benefiting from previous results *via* importance sampling.

$$\theta^{k+1} = \arg\max_{\theta\in\Theta}\mathbb{E}\left[\mathbf{1}_{[s_{k+1}\,,\infty[}\left(Z^{\theta^k}\right)\frac{f_X\left(Z^{\theta^k}\right)}{f_Z^{\theta^k}\left(Z^{\theta^k}\right)}\ln f_Z^\theta\left(Z^{\theta^k}\right)\right] \quad (2.24)$$

In practice however, these optimisation problems are substituted with their CMC proxies

$$\theta^{k+1} = \arg\max_{\theta\in\Theta}\sum_{j=1}^{N}\mathbf{1}_{[s_{k+1}\,,\infty[}\left(Z_j^{\theta^k}\right)\frac{f_X\left(Z_j^{\theta^k}\right)}{f_Z^{\theta^k}\left(Z_j^{\theta^k}\right)}\ln f_Z^\theta\left(Z_j^{\theta^k}\right) \quad (2.25)$$

until threshold s is reached for they can more easily be solved, some times even analytically. With this respect, exponential changes of measure are particularly convenient, especially if f_Z^θ belongs to the Natural Exponential Family.

[2] $Z^\theta \sim f_Z^\theta$ and $Z^* \sim f_Z^*$.

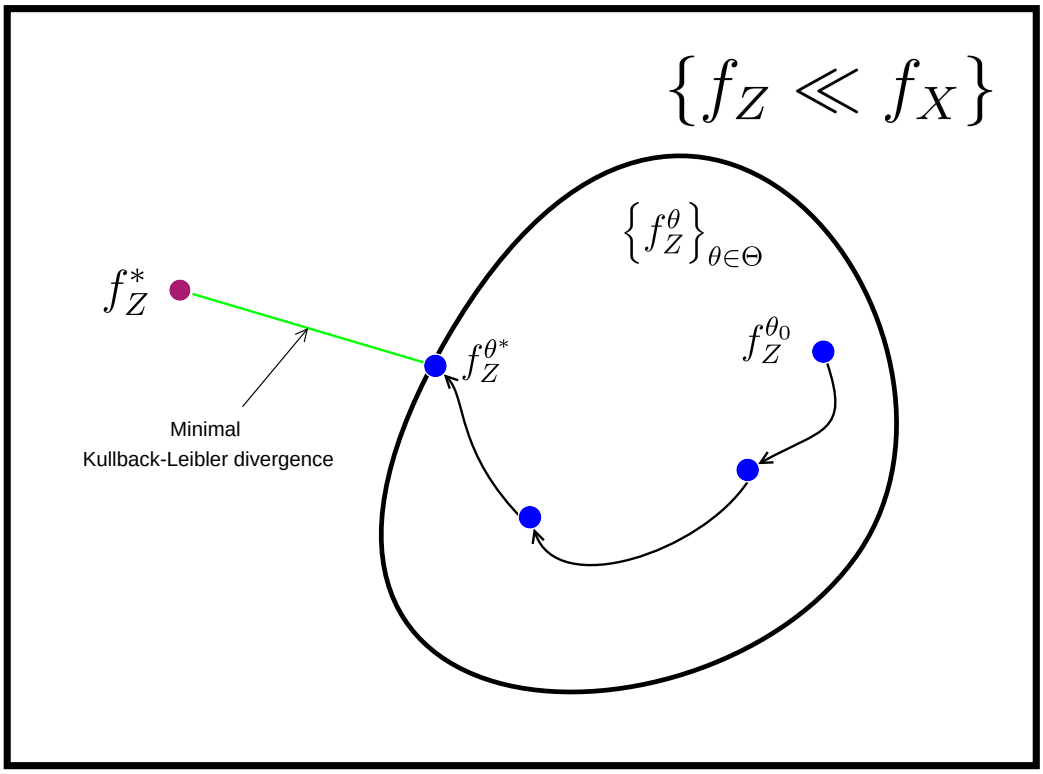


Figure 2.2: Importance Sampling Cross-Entropy based probability density change: θ moves within Θ so as to find a $f_Z^{\theta^*}$ that minimises the Kullback-Leibler divergence between the parametric density family $\left\{f_Z^\theta\right\}_{\theta\in\Theta}$ and f_Z^* the step by step.

Definition 2.2.1 (Exponential changes of measure)**.** *Let X be a random variable on $\mathbb{R}^d$. If the following holds,*

1. $\theta \in \mathbb{R}^l$ *and* $\forall i \in \{1, \cdots, l\}, g_i : \mathbb{R}^d \to \mathbb{R}$.
2. $x \mapsto \exp\left(\sum_{i=1}^l \theta_i g_i(x)\right)$ *is integrable and sums to* $\exp(-I(\theta))$,

then the following density function is said to be an exponential change of measure

$$f_Z^\theta : x \mapsto \exp\left(\sum_{i=1}^{l} \theta_i g_i(x) + I(\theta)\right).$$

Two famous classes of exponential changes of measure are:

1. *the Exponential Twisting or Tilting [12, 80]*

$$f_Z^\theta : x \mapsto f_X(x) \exp(\theta\phi(x) + I(\theta)),$$

2. *and the Natural Exponential Family (NEF) [71]*

$$f_Z^\theta : x \mapsto g_0(x) \exp\left(\theta^T x + I(\theta)\right).$$

The exponential changes of measure stem from large deviation theory [12]. Minimizing the KL-divergence between the exponential change of measure f_Z^θ and a given target density defines θ implicitly as the solution of a moment problem.

Equation 2.25 CMC formulation permits to define threshold s_{k+1} as the empirical β-quantile of the $f_Z^{\theta^k}$–generated sample. These two points translate into the following algorithm, where $\frac{f_X}{f_Z^\theta}$ is denoted w^θ.

Algorithm 2.2.1 (IS Cross entropy estimator)**.** *Let Y be $\mathbf{1}_{\{\phi(X)\in[s,\infty[\}}$. To estimate threshold exceedance probability $\mathbb{P}[\phi(X)\in[s,\infty[]=\mathbb{E}[Y]$ proceed as follows.*

1. *Choose or design a parametric density family $\left\{f_Z^\theta\right\}_{\theta\in\Theta}$ with respect to which f_X is absolutely continuous.*

2. *Set $k=0$, starting parameter θ^0 and threshold s_0 and quantile level α in accordance to the simulation budget.*

3. *Until $s_k\geq s$:*

   (a) *Generate the* iid *sample set $\left(Z_1^{\theta^k},\cdots,Z_N^{\theta^k}\right)$ such that $\forall i, Z_i^{\theta^k}\sim f_Z^{\theta^k}$.*

   (b) *Use the transfer and density functions to build*

$$\left(\mathbf{1}_{[s_k,\infty[}\left(\phi\left(Z_1^{\theta^k}\right)\right)\mathrm{w}^{\theta^k}\left(Z_1^{\theta^k}\right),\cdots,\mathbf{1}_{[s_k,\infty[}\left(\phi\left(Z_N^{\theta^k}\right)\right)\mathrm{w}^{\theta^k}\left(Z_N^{\theta^k}\right)\right)$$

   (c) *Define i^* as the index of the β-quantile of the previous sample set.*

   (d) *Set $s_{k+1}=Z_{i^*}^{\theta^k}$.*

   (e) *Calculate the next parameter*

$$\theta^{k+1}=\arg\max_{\theta\in\Theta}\sum_{j=1}^{N}\mathbf{1}_{[s_{k+1},\infty[}\left(Z_j^{\theta^k}\right)\mathrm{w}^{\theta^k}\left(Z_j^{\theta^k}\right)\ln f_Z^\theta\left(Z_j^{\theta^k}\right)$$

   (f) *Increment k by one: k=k+1.*

4. *Set $k^*=k-1$ and generate the* iid *sample set $\left(Z_1^{\theta^{k^*}},\cdots,Z_N^{\theta^{k^*}}\right)$ according to $f_Z^{\theta^{k^*}}$.*

5. *Approximate $\mathbb{P}[X\geq s]\approx\frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_{[s_{k+1},\infty[}\left(Z_j^{\theta^k}\right)\mathrm{w}^{\theta^{k^*}}\left(Z_j^{\theta^k}\right)=\widehat{Y_\flat}$.*

Requirements are the same as Algorithm 2.1.1 for all the $f_Z^{\theta^k}$.

Auxiliary quantile level β is chosen as a tradeoff between the forecasted number of thresholds and how many points should be generated at each step so as to make meaningful optimisations.

2.2.3 Non-parametric Adaptive Importance Sampling

The Non-parametric Importance Sampling (NIS) strategy is two folded [73, 94]: first build $f_Z^\bullet$, a proxy to f_Z^*, via a K kernel mixture, then use it to estimate $\mathbb{E}[Y]$.

Thus, from any chosen initial density f_Z°, one has to generate *iid* sample set $(Z_1^\circ, \cdots, Z_N^\circ)$, and then $\forall x = (x_1, \cdots, x_d) \in \mathbb{R}^d$, build following quantities, valid for positive ϕ.

$$h \in \left(\mathbb{R}^+\right)^d \text{ is the kernel bandwidth} \qquad |h| = \prod_{i=1}^{d} h_i \tag{2.26}$$

$$K_d(x,h) = \prod_{i=1}^{d} K\left(\frac{x_i}{h_i}\right) \qquad \mathrm{w}^\circ = \frac{f_X}{f_Z^\circ} \tag{2.27}$$

$$f_\phi(x) = \frac{\sum_{k=1}^{N} (\phi \mathrm{w}^\circ)(Z_k^\circ) K_d(Z_k^\circ - x, h)}{N|h|} \qquad \overline{\omega} = \frac{\sum_{k=1}^{N} (\phi \mathrm{w}^\circ)(Z_k^\circ)}{N} \tag{2.28}$$

If $\phi(x) = \mathbf{1}_{\mathbf{A}}(x)$, f_ϕ can be zero. We experienced this issue as explained in Section 4.1 and propose a variation of the NIS algorithm in Chapter 5. An other way out of this predicament is mixing f_ϕ with an other density f such that $f_X \ll f$.

$$f_Z^\bullet(x) = \frac{\overline{\omega} \times \frac{f_\phi(x)}{\overline{\omega}} + \delta_N f(x)}{\overline{\omega} + \delta_N} \tag{2.29}$$

Astutely designed, f and δ_N could provide an efficient guardrail in case f_ϕ can be zero somewhere. Let us keep things simple and assume f_ϕ is nowhere zero and then choose

$$f_Z^\bullet(x) = \frac{f_\phi(x)}{\overline{\omega}}. \tag{2.30}$$

$f_Z^\bullet$ can now be conveniently used to generate the *iid* sample $(Z_1^\bullet, \cdots, Z_m^\bullet)$ so as to perform the estimation

$$\mathrm{w}^\bullet = \frac{f_X}{f_Z^\bullet} \qquad \widehat{Y}_{m+N} = \frac{\sum_{i=1}^{m} (\phi \mathrm{w}^\bullet)(Z_i^\bullet)}{m} \tag{2.31}$$

This procedure obviously makes the choice of initial density f_Z° and kernel K of major importance. Actually, f_Z° samples should be close to where f_Z^* takes important values so that this sub-domain is well sampled. In [94], $\widehat{Y}_{m+N}$ is shown to converge toward $\mathbb{E}[Y]$ under the following assumptions.

1. K is bounded with finite second order moment.
2. ϕ has compact support on which both ϕ and f_X are twice continuously differentiable and bounded away from zero.
3. $\forall i \in \{1, \cdots, d\}, h_i = h$ and, $h \to 0$ and $mh^d \to \infty$ as $m \to \infty$.

Under this hypothesis, a central limit theorem type is derived as well but the variance explicitly requires ϕ first and second order derivatives, which are in practice often unknown. Therefore, the optimal bandwidth, which is based on the estimator variance can be used.

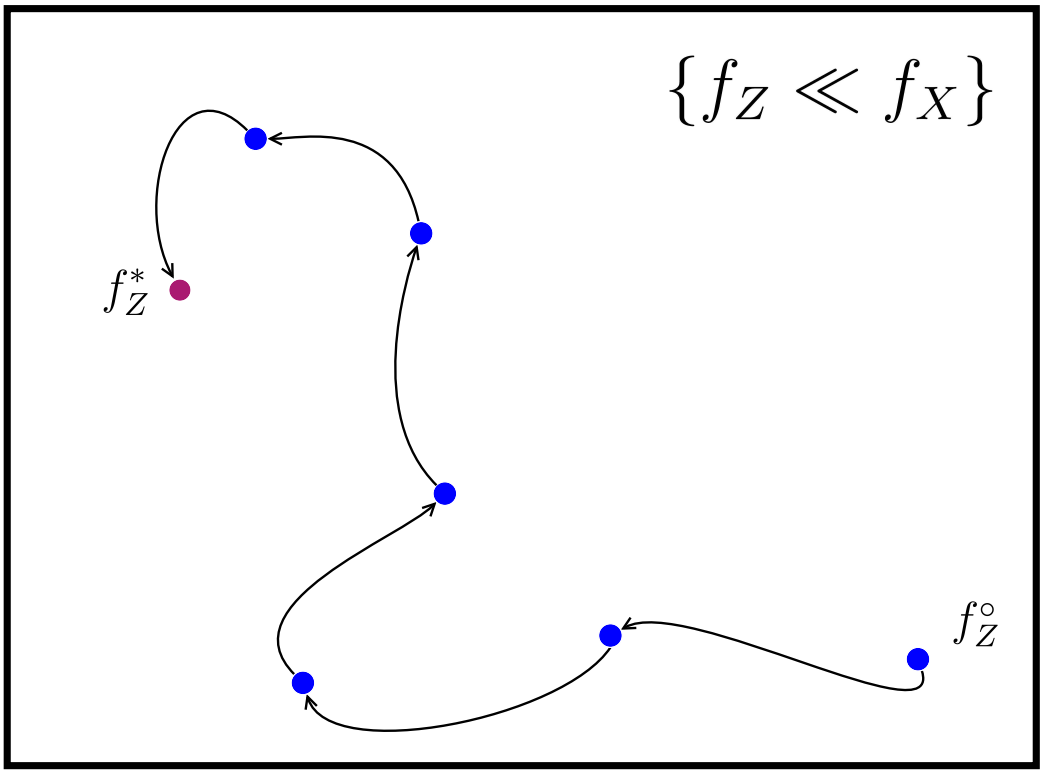


Figure 2.3: Importance sampling NAIS based probability density change. Under some hypothesis, the algorithm draws closer and closer to the optimal density f_Z^* as the number of steps increases and converges eventually.

This later paper suggests a NIS improvement called the Non-parametric Adaptive Importance Sampling (NAIS). The basic idea is iterating the above mentioned procedure *ad libitum*, say κ times, in an attempt to learn f_Z^* on the fly, ideally as in the Figure 2.3. All generated points are taken into account in the learning of f_Z^* because they are all designed to approximate it.

Algorithm 2.2.2 (Non-parametric Adaptive Importance Sampling estimator)**.** *To estimate the expectation of integrable random variable $Y = \phi(X)$ proceed as follows.*

1. *Choose an initial density f_Z° and generate* iid *sample set $(Z_1^\circ, \cdots, Z_N^\circ)$.*
2. *Build $f_Z^{\bullet,1}$ as per Equation 2.30.*
3. *Generate $\left(Z_1^{\bullet,1}, \cdots, Z_{m_1}^{\bullet,1}\right)$ from $f_Z^{\bullet,1}$ and compute*[3]

$$\mathrm{w}^{\bullet,1} = \frac{f_X}{f_Z^{\bullet,1}} \qquad \widehat{Y}_\bullet^{\,1} = \frac{\sum_{j=1}^{m_1} \left(\phi \mathrm{w}^{\bullet,1}\right)\left(Z_j^{\bullet,1}\right)}{m_1} \tag{2.32}$$

[3] $\forall z, \left(\phi \mathrm{w}^{\bullet,1}\right)(z) = \mathrm{w}^{\bullet,1}(z)\,\phi(z)$

4. *Update* $f_Z^\bullet$ *into*

$$f_Z^{\bullet,1}(x) = \frac{\sum_{j=1}^{m_1} (\phi \mathrm{w}^{\bullet,1}) \left(Z_j^{\bullet,1}\right) K_d \left(Z_k^{\bullet,1} - x, h^1\right)}{m_1 |h^1| \widehat{Y}_\bullet^1} \tag{2.33}$$

5. *Starting from* $k = 2$ *and until* $k = \kappa + 1$*, do as follows.*

   (a) *Generate* $\left(Z_1^{\bullet,k}, \cdots, Z_{m_k}^{\bullet,k}\right)$ *from* $f_Z^{\bullet,k-1}$ *and update the estimate of* $\mathbb{E}[Y]$ *into*

$$\widehat{Y}_\bullet^k = \frac{\sum_{i=1}^k \sum_{j=1}^{m_k} (\phi \mathrm{w}^{\bullet,i}) \left(Z_j^{\bullet,i}\right)}{\sum_{i=1}^k m_i} \tag{2.34}$$

   (b) *Update the estimate of* f_Z^*

$$\mathrm{w}^{\bullet,k} = \frac{f_X}{f_Z^{\bullet,k}} \qquad f_Z^{\bullet,k}(x) = \frac{\sum_{i=1}^k \sum_{j=1}^{m_i} (\phi \mathrm{w}^{\bullet,i}) \left(Z_j^{\bullet,i}\right) K_d \left(Z_j^{\bullet,i} - x, h^k\right)}{\sum_{i=1}^k m_i |h^k| \widehat{Y}_\bullet^k} \tag{2.35}$$

   (c) *Increment* k *by one.*

Requirements are the same as Algorithm 2.1.1.

The convergence hypothesis, detailed in [94], are very similar as well. The gist is that the NIS convergence hypothesis hold for each intermediate NAIS density. Though proven theoretically less efficient than NIS as the input dimension d grows above eight , which will be our case, NAIS gradually builds the proxy to f_Z^* and therefore might make it easier to use efficiently.

2.3 Auxiliary density design for quantile estimation

As already noticed in Subsection 1.3.5 when dealing with the CMC quantile estimator asymptotic behaviour, though linked, expectation estimation and quantile estimation are problems of their owns. This fact remains when using Importance Sampling.

2.3.1 Four Importance Sampling quantile estimators

Quantile estimation is no trickier than Algorithm 2.1.1 and likewise the CMC case – Definition 1.3.2 – an IS cumulative distribution function can be derived plugging $\phi(\cdot) = \mathbf{1}_{]-\infty,q]}(\cdot)$ in Equation 2.1.

Definition 2.3.1 (IS Cumulative distribution function and quantile estimator)**.** *Let* $(Z_1, \cdots, Z_N)$ *be a set of* N iid *according to* ν *and integrable random variables* $(\mathbb{E}[|Z|] < \infty)$ *on the* $\mathbb{R}$ *line such that* μ *is absolutely continuous w.r.t.* ν*. According to the SLLN and equation (2.1), the following approximation holds.*

$$\forall x \in \mathbb{R}, \frac{1}{N} \sum_{k=1}^N \underset{\{Z_k \leq x\}}{\mathbf{1}} (\mathrm{w}(Z_k)) \xrightarrow[N \to \infty]{a.s.} F_X(x) = \mathbb{P}[X \leq x] \tag{2.36}$$

Then, the importance sampling cumulative distribution function (iscdf), defined as follows,

$$\forall x \in \mathbb{R}, \widehat{F}_X(x) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\{Z_k \leq x\}} \mathrm{w}(Z_k) \equiv \widehat{F}_{1X}(x) \tag{2.37}$$

yields almost sure pointwise convergence toward F_X*.*

$$\forall x \in \mathbb{R}, \widehat{F}_X(x) \xrightarrow[N\to\infty]{a.s.} F_X(x) \tag{2.38}$$

Acually, according to [38], with $\overline{\mathrm{w}}_N = \frac{1}{N}\sum_{k=1}^{N} \mathrm{w}(Z_k)$*, one can use any of the following mappings.*

$$\widehat{F}_{1X}(x) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\{Z_k \leq x\}} \mathrm{w}(Z_k) \qquad \widehat{F}_{2X}(x) = 1 - \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\{Z_k > x\}} \mathrm{w}(Z_k) \tag{2.39}$$

$$\widehat{F}_{3X}(x) = \frac{\widehat{F}_{1X}(x)}{\overline{\mathrm{w}}_N} \qquad \widehat{F}_{4X}(x) = \frac{\widehat{F}_{2X}(x)}{\overline{\mathrm{w}}_N} \tag{2.40}$$

for they serve the same purpose: estimating quantiles.

$$\forall i \in \{1, \cdots, 4\}, \forall \alpha \in [0,1], \widehat{q}_i = \inf\left\{x \,\middle|\, \widehat{F}_{iX}(x) \geq \alpha\right\} \tag{2.41}$$

The interest of the three other estimators is that they are as easy to simulate as $\widehat{F}_X$ and can provide some asymptotic variance reduction with respect to it as further detailed in Subsection 2.3.2. Moreover, they can be used in a Crude Monte Carlo context, Section 1.3, and offer the same benefits

$$[\nu = \mu] \Rightarrow \left[\mathrm{w}(x) = 1, \widehat{F}_X(x) = \overline{F}_X(x), \widehat{F}_{2X}(x) = \overline{F}_{2X}(x) = 1 - \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{]-\infty,x]}(X_k)\right] \tag{2.42}$$

at no extra computational nor theoretical cost with respect to $\overline{F}_X$. The choice criterion between the four estimators is their asymptotic variance.

2.3.2 Asymptotic behaviour of the IS quantile estimator

Though their laws are (way) more complicated than that of F_X, $\widehat{F}_{1X}$, $\widehat{F}_{2X}$, $\widehat{F}_{3X}$ and $\widehat{F}_{4X}$ have similar asymptotic features with it [38]: normality and most likely intractable squared derivative $(F'_X(q))^2$ as variance denominator.

Theorem 2.3.1 (Asymptotic behaviour of the plain IS quantile estimator)**.** *Suppose that* $\mathbb{E}[\mathrm{w}^3(Z)] < \infty$ *and assume* F_X *is differentiable at* $q = F_X^{-1}(\alpha)$ *with* $F'_X(q) > 0$*. Then*

$$\forall i \in \{1,2,3,4\}, N^{\frac{1}{2}}(\widehat{q}_i - q) \xrightarrow[N\to\infty]{d} \sigma_i \mathcal{N}(0,1) \tag{2.43}$$

where

$$\sigma_1^2 = \frac{\mathbb{E}\left[\mathrm{w}^2(Z)\,\mathbf{1}_{]-\infty,q]}(\phi(Z))\right] - \alpha^2}{\left(F'_X(q)\right)^2} \qquad \sigma_3^2 = \frac{\mathbb{E}\left[\mathrm{w}^2(Z)\left(\mathbf{1}_{]-\infty,q]}(\phi(Z)) - \alpha\right)^2\right]}{\left(F'_X(q)\right)^2} \tag{2.44}$$

$$\sigma_2^2 = \frac{\mathbb{E}\left[\mathrm{w}^2(Z)\,\mathbf{1}_{[q,\infty[}(\phi(Z))\right] - (1-\alpha)^2}{\left(F'_X(q)\right)^2} \qquad \sigma_4^2 = \frac{\mathbb{E}\left[\mathrm{w}^2(Z)\left(\mathbf{1}_{[q,\infty[}(\phi(Z)) - (1-\alpha)\right)^2\right]}{\left(F'_X(q)\right)^2} \tag{2.45}$$

The numerator differences are of interest as they can be used as an objective crude choice criterion among same cost estimators and can be estimated *via* CMC approximation.

These estimator variations can helps further decrease variance, but are not the decrease main driver. The key to variance reduction is an appropriate change of measure.

2.3.3 The unknown optimal change of measure

As previously in the expectation estimator case, the algorithm's likeness to the CMC quantile estimator Algorithm 1.3.1 is deceivingly promising: designing and sampling according to an appropriate auxiliary probability measure *is* a challenge.

One can explain this by the intrinsic difficulty of analysing the inverse of a mapping random estimator. Another major reason is that there is not explicit zero variance measure to target. Therefore, as [33] puts it, "there is limited literature on the subject".

- [38] uses large deviation theory efficiently to deal the special case where X is a sum of random variables, which is not here the concern.
- [33] provides too theoretical a framework for its hypothesis to be checked in an industrial context.

Apart from that, in practice, the heuristic approach is designing f_Z so as to estimate with a low variance the probability $\mathbb{P}[\phi(X) \leq q_\alpha]$.

A priori knowledge about ϕ and q_α can help choosing where and how to sample the input space. Besides asymptotic results such as Theorem 2.3.1 can support the estimator choice. In a nutshell, one can go back to Section 2.2 about expectation estimation and use it to build up the auxiliary density *mutatis mutandis, servatis servandis* as exemplified in Algorithm 2.3.1.

2.3.4 The IS quantile estimation algorithm

The IS quantile estimator algorithm stems from the SLLN and the change of measure.

$$\mathbb{P}[\phi(X) \leq q] = \mathbb{E}\left[\mathbf{1}_{]-\infty,q]}(\phi(X))\right] = \mathbb{E}\left[\mathbf{1}_{]-\infty,q]}(\phi(Z))\mathrm{w}(Z)\right] \tag{2.46}$$

In practice however, computing thoroughly the cumulative distribution function proxies is not needed as only the weights are used *as per* Definition 2.3.1. Besides, the strategy is to reformulate the expectation estimation algorithms as *we* exemplify in the following Algorithm 2.3.1.

Algorithm 2.3.1 (IS quantile plain estimator)**.** *To estimate the α-quantile q of the random variable $Y = \phi(X)$ proceed as follows. The algorithm is stated for $\alpha \geq \frac{1}{2}$. If $\alpha \leq \frac{1}{2}$, one can easily adapt the algorithm.*

1. *Choose or design ν w.r.t. which μ is absolutely continuous.*
   - *CE: use Algorithm 2.2.1 and change the stopping criterion item 3 into*
   $$\textit{Until } \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{[s_{k+1},\infty[}\left(Z_j^{\theta^k}\right) \mathrm{w}^{\theta^{k^*}}\left(Z_j^{\theta^k}\right) \geq \alpha$$
   - *NAIS: no change from Algorithm 2.2.2.*
2. *Generate the* iid *sample set $(Z_1, \cdots, Z_N)$ such that $\forall i, Z_i \sim \nu$.*
3. *Build $(Y_1, \cdots, Y_N)$ so that $\forall i, Y_i = \phi(Z_i)$.*
4. *Choose empirical cdf $\widehat{F}_{jY}$ among those given in Definition 2.3.1 and conclude*
$$q \approx \inf\left\{y \,\middle|\, \widehat{F}_{jY}(y) \geq \alpha\right\} = \widehat{q} \tag{2.47}$$

Requirements are in any case the same as Algorithm 2.1.1.

2.4 Conclusion

Importance Sampling seems convenient as it keeps most of the CMC formalism. However, defining the appropriate change of measure is a hard task, especially when the transfer function ϕ is a black box and little to nothing is known about its mathematical properties. Cross-entropy searches an appropriate candidate within a parametric probability density family, while Non-parametric Importance Sampling builds its as it goes *via* probability kernels. Both require very important sound decisions from the user. The Splitting Technique dodges most of them.

Chapter 3

The Splitting Technique

The Splitting Technique (ST) is another attempt at modifying Crude Monte Carlo so as to deal with rare events. While CMC uses sheer calculation power to perform an expectation estimation, Importance Sampling tops or partially trades brute force with an educated change of probability measure so as to reduce variance. ST, on its part, chops the problem in easier to handle ones, solves them iteratively and eventually combines their solutions into that of the original problem.

For instance, in a typical industrial probability estimation problem, the system of interest is a deterministic black box mapping $\phi : \mathbb{X} \to \mathbb{R}^d$ with a random input X. The black box mapping ϕ has no analytical expression for it stands for a complex simulation software. Inputs X are random due to uncertainties on the actually physical quantities or because the environment the system will evolve in is random. Besides, ϕ can be composed with a satisfaction criterion $\zeta : \mathbb{R}^d \to \mathbb{R}$ into a score $\vartheta = \zeta \circ \phi$ in order to quantify the system's performance and check its compliance to either security norms or behaviour objectives *via* a target score set $\mathbf{C}$. The question is then the probability with which the target score set is hit by $Y = \vartheta(X) \sim \eta$.

$$\mathbb{P}[Y \in \mathbf{C}] = \mathbb{P}[\vartheta(X) \in \mathbf{C}] = \mathbb{P}[X \in \mathbf{A}] \text{ where } \mathbf{A} = \vartheta^{-1}(\mathbf{C}) \ ? \tag{3.1}$$

To cope with this, ST teams the old saying *divide et impera* and conditional probabilities $\mathbb{P}[a, b] = \mathbb{P}[b]\,\mathbb{P}[a|\, b]$.

ST first rephrases the target set as the smallest and ultimate element of a sequence of nested sets

$$\left\{ \begin{array}{ccccccccc} \mathbb{R} & = & \mathbf{C}_0 & \supset & \cdots & \supset & \mathbf{C}_{\kappa+1} & = & \mathbf{C} \\ \mathbb{X} & = & \mathbf{A}_0 & \supset & \cdots & \supset & \mathbf{A}_\kappa & = & \mathbf{A} \end{array} \right. \text{ where } \forall i, \mathbf{A}_i = \vartheta^{-1}(\mathbf{C}_i) \tag{3.2}$$

then the desired probability *via* a product of conditional probabilities

$$\mathbb{P}[X \in \mathbf{A}] = \prod_{i=1}^{\kappa} \mathbb{P}[X \in \mathbf{A}_i |\, X \in \mathbf{A}_{i-1}] \tag{3.3}$$

and eventually estimates them one after the other. This way, a difficult problem can be swapped against many seemingly easier ones.

Though based on simple ideas, implementing ST requires insight and shrewd tools.

1. How to define the nested set sequence?

2. How to estimate the conditional probabilities?

3. How does this estimator variance compare with that of CMC?

Though the two first questions are answered in the next section, the last one will be dealt with separately.

For simplicity sake and because it is a very typical form, $\mathbf{C}$ is henceforth assumed to be

$$\mathbf{C} = [\, s, \infty \,[\tag{3.4}$$

Besides, if $\mathbf{C}$ is an interval collection, it can be seen as such through any function such that $[\, s, \infty \,[$ antecedent is $\mathbf{C}$. Actually,

$$\text{If } g^{-1}\left([\, s, \infty \,[\right) = \mathbf{C}, \text{ then } \mathbf{1}_{[\, s,\infty[} \circ g = \mathbf{1}_{\mathbf{C}} \tag{3.5}$$

and the following equality holds.

$$\mathbb{P}\left[g\left(Y\right) \in [\, s, \infty \,[\right] = \int_{\mathbb{Y}} \mathbf{1}_{[\, s,\infty[} \circ g\left(Y\right) \eta\left(\,\mathrm{d}y\right) = \int_{\mathbb{Y}} \mathbf{1}_{\mathbf{C}}\left(Y\right) \eta\left(\,\mathrm{d}y\right) = \mathbb{P}\left[Y \in \mathbf{C}\right] \tag{3.6}$$

This transformation makes the ST algorithm more convenient to implement but, for it basically redefines the score function $\vartheta \equiv g \circ \vartheta$, modifies the estimator variance as explained in section 3.3.

3.1 The theoretical framework

The Splitting Technique is here introduced as a rare event dedicated technique, a good introduction to which can be found in [18, 52, 54], albeit their focus is mainly on random processes $(X_t)_{t\geq 0}$. As the black box mapping ϕ we will consider does not involve time, [19, 67] is more appropriate for the static case at hands.

Just as Importance Sampling, ST aims at approximating the optimal probability measure change with respect to the sought probability. The former requires *a priori* knowledge about the measure change to be designed before starting the estimation, as detailed in chapter 2. The latter gathers valued information on its way and builds along a proxy to the change of measure in the form of a set of points with appropriate distribution called particle cloud.

3.1.1 An intuitive approach

A Splitting Technique iteration ideally starts with a sample of *iid* throws of X known to be in $\mathbf{A}_{i-1}$. With $i = 1$, this is only performing a plain CMC simulation. Then, and until the subset $\mathbf{C}$ is reached, it is done as follows:

1. Estimate $\mathbb{P}\left[X \in \mathbf{A}_i \middle| X \in \mathbf{A}_{i-1}\right]$.

2. Select particles in $\mathbf{A}_i$ and replace those outside *via* uniform resampling among the selected ones.

3. Use the whole particle cloud to generate throws in $\mathbf{A}_i$ that are *iid* according to $\left(X \,|X \in \mathbf{A}_i\right)$ such that their scores *via* ϑ are *iid* according to $\left(Y \,|Y \in \mathbf{C}_i\right)$.

When $\mathbf{C}$ is reached, conclude. Points 1 and 3 now have to be formally detailed.

3.1.2 Designing the subset sequence

Formally one can either define the $\mathbf{C}_i$ or the $\mathbf{A}_i$. In practice however, ϕ is too complicated to estimate $\vartheta(x)$ beforehand for any given $x \in \mathbb{X}$ and design the $\mathbf{A}_i$ sequence. Therefore the $\mathbf{C}_i$ sequence is designed according to target score set $\mathbf{C}$ and scalar satisfaction criterion function ζ, which is usually easier to handle than ϑ, as $\zeta\left(\mathbb{R}^d\right)$ subsets.

As the problem at hands is a probability of exceedance, an increasing sequence of threshold is defined

$$-\infty = s_0 < \cdots < s_\kappa = s \text{ and } \left\{ \begin{array}{cccc} \forall i, & \mathbf{C}_i & = & [\, s_i, \infty\, [\\ \forall x \in \mathbb{X}, & [x \in \mathbf{A}_i] & \Leftrightarrow & [\vartheta(x) \in \mathbf{C}_i] \end{array} \right. \tag{3.7}$$

with the associated $\mathbf{A}_i$ sequence, which does not need explicit definition anymore. The thresholds do.

Formally, thresholds can be chosen arbitrarily. However, if $\mathbb{P}[X \in \mathbf{A}_{i+1} | X \in \mathbf{A}_i]$ is almost one *i.e.* the s_i and s_{i+1} thresholds are too close, virtually nothing is learned during the iteration and the simulation makes not headway and the estimation is needlessly costly. One the other hand, if $\mathbb{P}[X \in \mathbf{A}_{i+1} | X \in \mathbf{A}_i]$ is too small *i.e.* the event $\{X \in \mathbf{A}_{i+1} | X \in \mathbf{A}_i\}$ is rare and the whole scheme both is pointless and can collapse as the particle cloud vanishes. If no particle of the original swarm reaches the next stage, [53] suggests to keep simulating until a given number does, though it can be simulation and time consuming.

There is no proven optimal rule regarding how to choose thresholds, but [51] calculates that in an ideal case all conditional probabilities should be equal to unknown value $\alpha^* = \mathbb{P}[Y \in \mathbf{C}]^{\frac{1}{\kappa}}$ in order to reduce variance. According to [17], $0.75 \leq \alpha \leq 0.8$ works very well in practice.

Fixing the threshold s_{i+1} so that $\mathbb{P}[Y \geq s_{i+1} | Y \geq s_i]$ exactly has the decided value is impossible as F_Y is unknown. Though, this can be approximated on the fly defining the next threshold as the empirical α-quantile of a $(Y | Y \in \mathbf{C}_i)$-distributed sample set, making it a random variable S_{i+1}. This is the Adaptive Splitting Technique (AST), which helps coping with the lack of information.

The Adaptive Splitting Technique costs two extras with respect to ST. The first is that the number of steps κ is now unknown and random. The second is a little bias which vanishes according to Section 3.3. Both are due to the thresholds being empirical quantiles. This is overall not expensive when compared to the provided flexibility. Another way is running the AST first to learn the thresholds and use them performing the native Splitting Technique.

3.1.3 Sampling according to the conditional laws

To replenish the sample set after discarding points outside $\mathbf{A}_i$, there is no perfect solution, unless one can generate directly an *iid* sample set according to the conditional law, which is unlikely. Right after the step 2 duplication however, particles are identically distributed according to the conditional law but not independent as there are doubloons: they are correlated identically distributed (*cid*). The step 3 objective is hence triple:

1. Increase variety in the set: there must be no pair of identical points.

2. Respect the probability law: it must be preserved through the transformation to be done.

3. Decrease interdependence: the correlation between particles with shared genealogy must be reduced to a minimum.

To this purpose, a μ-reversible transition kernel $\mathrm{M}(\cdot,\cdot)$ is used.

Definition 3.1.1 (Transition kernel)**.** *A transition kernel* $\mathrm{M}(\cdot,\cdot) : \mathbb{X} \times \mathcal{X} \to \mathbb{R}$ *is a mapping such that*

$$\begin{cases} \forall \mathbf{A} \in \mathcal{X} \quad , \quad \mathrm{M}(\cdot, \mathbf{A}) \quad : \quad \mathbb{X} \to \mathbb{R} & \text{is measurable.} \\ \forall x \in \mathbb{X} \quad , \quad \mathrm{M}(x, \cdot) \quad : \quad \mathcal{X} \to \mathbb{R} & \text{is a probability measure.} \end{cases} \tag{3.8}$$

$\forall x \in \mathbb{X}$, $\mathrm{M}(x,\cdot)$ *stands as a* x*-specific random way to propose another point* Z *in* $\mathbb{X}$. $\mathrm{M}(\cdot, \mathbf{A})$ *is measurable so that is integrable and Markov kernels can be composed.*

A transition kernel M *can be composed with another transition kernel* Q *into a third transition kernel denoted* MQ *and defined as follows.*

$$\forall x \in \mathbb{X}, \mathrm{MQ}(x, \mathrm{d}z) = \int_{\mathbb{X}} \mathrm{M}(x, \mathrm{d}u)\, \mathrm{Q}(u, \mathrm{d}z) \tag{3.9}$$

This means the first transition is chosen according to M *and the second according to* Q.

M *can be iterated* i.e. *composed with itself.*

$$\forall n \in \mathbb{N}, \forall x \in \mathbb{X}, \begin{array}{ccc} \mathrm{M}^0(x, \mathrm{d}z) & = & \delta_x(\mathrm{d}z) \\ \mathrm{M}^{n+1}(x, \mathrm{d}z) & = & \int_{\mathbb{X}} \mathrm{M}^n(x, \mathrm{d}u)\, \mathrm{M}(u, \mathrm{d}z) \end{array} \tag{3.10}$$

M^n *means that* n *transitions are made according to the policy defined by* M.

For convenience sake, M will denote both the probability measure $\mathrm{M}(x, \mathrm{d}z)$ and the density $\mathrm{M}(x, z)$ w.r.t. the Lebesgue measure, when it exists.

A transition kernel can be used to remove doubloons within a set replacing two points located in x with two *iid* throws according to $\mathrm{M}(x, \mathrm{d}z)$. However, to maintain the set's probability distribution a specific property is needed.

Definition 3.1.2 (Transition kernel invariance)**.** *A transition kernel* M *is said to be* μ*–invariant if*

$$\forall x, z \in \mathbb{X}, \mu(\mathrm{d}x)\, \mathrm{M}(x, \mathrm{d}z) = \mu(\mathrm{d}z) \;\; \textit{i.e.} \;\; \mu\mathrm{M} = \mu \tag{3.11}$$

This implies that if from a μ set, M is used to generate another set, the new set is distributed according to μ as well. Such a transition kernel removes μ–doubloons and keeps their distribution. This is almost the actual target: kernels that are invariant with respect to the conditional probability measures of X.

$$\forall \mathbf{A} \in \mathcal{X}, \forall i, \mu_i \mathrm{M}_i = \mu_i \text{ where } \mu_i = \frac{\mathbf{1}_{\mathbf{A}_i} \mu}{\mu(\mathbf{A}_i)} \tag{3.12}$$

A collection of such invariant kernels is scarce. However, it can be built from M given that M does not only leave μ invariant but is also μ–reversible.

Definition 3.1.3 (Transition kernel reversibility)**.** M *is said to be a* μ*-reversible transition kernel if*

$$\forall x, z \in \mathbb{X}, \mu(\mathrm{d}x)\, \mathrm{M}(x, \mathrm{d}z) = \mu(\mathrm{d}z)\, \mathrm{M}(z, \mathrm{d}x) \tag{3.13}$$

This equation is known to physicists as the detailed balance equation [48]. Atop invariance, the reversibility property induces that when faced with a μ set and another generated from it through M, there is no statistical way to tell them apart and know which one generated the other [52].

Property 3.1.1 (Reversibility property)**.** *Let the transition kernel* M *be reversible with respect to the probability measure* μ*, that is to say* (3.13) *holds. Then*

1. μ *is invariant w.r.t.* M*:* $\mu\mathrm{M} = \mu$*.*
2. *If functions* g *and* h *are measurable, then* $\mu\left(g\mathrm{M}\left(h\right)\right) = \mu\left(h\mathrm{M}\left(g\right)\right)$*.*

Proof. (1) comes directly when integrating (3.13) w.r.t. variable x.

$$\int_{\mathbb{X}} \mu\left(\mathrm{d}x\right)\mathrm{M}\left(x,\,\mathrm{d}z\right) = \mu\left(\mathrm{d}z\right) \tag{3.14}$$

(2) comes through calculation.

$$\begin{aligned}
\mu\left(g\mathrm{M}\left(h\right)\right) &= \int_{\mathbb{X}}\left[g\left(x\right)\left(\int_{\mathbb{X}} h\left(z\right)\mathrm{M}\left(x,\,\mathrm{d}z\right)\right)\right]\mu\left(\mathrm{d}x\right) && (3.15)\\
&= \int\int_{\mathbb{X}}\left[g\left(x\right)h\left(z\right)\right]\left(\mu\left(\mathrm{d}x\right)\mathrm{M}\left(x,\,\mathrm{d}z\right)\right) && (3.16)\\
&= \int\int_{\mathbb{X}}\left[g\left(x\right)h\left(z\right)\right]\left(\mu\left(\mathrm{d}z\right)\mathrm{M}\left(z,\,\mathrm{d}x\right)\right) \qquad \text{by (3.13)} && (3.17)\\
&= \int_{\mathbb{X}}\left[h\left(z\right)\left(\int_{\mathbb{X}} g\left(x\right)\mathrm{M}\left(z,\,\mathrm{d}x\right)\right)\right]\mu\left(\mathrm{d}z\right) && (3.18)\\
&= \mu\left(h\mathrm{M}\left(g\right)\right) && (3.19)
\end{aligned}$$

□

Thanks to this property, with a μ–reversible kernel M, a kernel M_i that leaves μ_i invariant can be constructed.

Property 3.1.2 (Invariance of conditionals based on a reversible kernel)**.** *Let the transition kernel* M *be reversible with respect to the probability measure* μ*. If the positive mapping* $g : \mathbb{X} \to \mathbb{R}^+$ *is bounded, then the transition kernel* $\mathrm{M}^\bullet$

$$\mathrm{M}^\bullet\left(x,\,\mathrm{d}z\right) = \mathrm{M}\left(x,\,\mathrm{d}z\right)\frac{g\left(z\right)}{c} + \left(1 - \frac{\mathrm{M}\left(g\right).\left(x\right)}{c}\right)\delta_x\left(\mathrm{d}z\right) \quad \textit{where } c = \sup_{\mathbb{X}} g \tag{3.20}$$

leaves the probability measure $\mu^\bullet$

$$\mu^\bullet = \frac{g\mu}{\mu\left(g\right)} \;\; \textit{i.e.} \;\; \forall x \in \mathbb{X}, \mu^\bullet\left(x\right) = \frac{g\left(x\right)\mu\left(\mathrm{d}x\right)}{\int_{\mathbb{X}} g\left(z\right)\mu\left(\mathrm{d}z\right)} \;\; \textit{invariant.} \tag{3.21}$$

In particular, plugging $g = \mathbf{1}_{\mathbf{A}_i}$ *in the previous equation,* M_i

$$\begin{gathered}
\mathrm{M}_i\left(x,\,\mathrm{d}z\right) = \mathrm{M}\left(x,\,\mathrm{d}z\right)\mathbf{1}_{\mathbf{A}_i}\left(z\right) + \left(1 - \mathrm{M}\left(x,\mathbf{A}_i\right)\right)\delta_x\left(\mathrm{d}z\right)\\
\textit{where } \mathrm{M}\left(x,\mathbf{A}_i\right) = \mathbb{P}\left[Z \in \mathbf{A}_i\right] \;\; \textit{with } Z \sim \mathrm{M}\left(x,\,\mathrm{d}z\right)
\end{gathered} \tag{3.22}$$

leaves μ_i *invariant.*

Proof. Let ϕ be bounded and measurable. By definition

$$\mathrm{M}^{\bullet}(\phi)(x) = \int_{\mathbb{X}} \mathrm{M}^{\bullet}(x,z)\,\phi(z)\,\mathrm{d}z = \frac{\mathrm{M}(g\phi)(x)}{c} + \left(1 - \frac{\mathrm{M}(g)(x)}{c}\right)\phi(x) \tag{3.23}$$

thus, using reversibility property 3.1.1,

$$\mu\left(g\mathrm{M}^{\bullet}(\phi)\right) = \mu\left(g\mathrm{M}\left(\frac{g\phi}{c}\right)\right) + \mu(g\phi) - \mu\left(\frac{g\phi}{c}\mathrm{M}(g)\right) = \mu(g\phi) \tag{3.24}$$

and eventually

$$\mu^{\bullet}\left(\mathrm{M}^{\bullet}\phi\right) = \frac{\mu\left(g\mathrm{M}^{\bullet}(\phi)\right)}{\mu(g)} = \frac{\mu(g\phi)}{\mu(g)} = \mu^{\bullet}(\phi) \;\; i.e. \;\; \mu^{\bullet}\mathrm{M}^{\bullet} = \mu^{\bullet} \tag{3.25}$$

□

It is therefore possible to remove doubloons and grow a $\mu^{\bullet}$–distributed particle cloud thanks to a μ–reversible transition kernel M.

Property 3.1.3 (Sampling from the $\mu^{\bullet}$–reversible transition kernel)**.** *To sample $X^{\bullet}$ according to $\mathrm{M}^{\bullet}(x, \mathrm{d}x^{\bullet})$, where $x \in \mathbb{X}$*

1. *sample $Z^{\bullet}$ according to $\mathrm{M}(x, \mathrm{d}x^{\bullet})$*
2. *set* $X^{\bullet} = \begin{cases} Z^{\bullet} & \text{with probability} \quad \frac{g(Z^{\bullet})}{c} \\ x & \text{with probability} \quad 1 - \frac{g(Z^{\bullet})}{c} \end{cases}$

Proof. Let ϕ be bounded and measurable.

$$\begin{aligned}
\mathbb{E}\left[\phi(X^{\bullet})\right] &= \mathbb{E}\left[\mathbb{E}\left[\phi(X^{\bullet})\middle|\, Z^{\bullet}\right]\right] && (3.26)\\
&= \mathbb{E}\left[\phi(Z^{\bullet})\frac{g(Z^{\bullet})}{c} + \phi(x)\left(1 - \frac{g(Z^{\bullet})}{c}\right)\right] && (3.27)\\
&= \int_{\mathbb{X}} \phi(u)\frac{g(u)}{c}\mathrm{M}(x, \mathrm{d}u) + \phi(x)\left(1 - \int_{\mathbb{X}} \frac{g(u)}{c}\mathrm{M}(x, \mathrm{d}u)\right) && (3.28)\\
&= \int_{\mathbb{X}} \phi(u)\left[\frac{g(u)}{c}\mathrm{M}(x, \mathrm{d}u) + \left(1 - \frac{\mathrm{M}g(x)}{c}\right)\delta_x(\mathrm{d}u)\right] && (3.29)\\
&= \int_{\mathbb{X}} \phi(u)\,\mathrm{M}^{\bullet}(x, \mathrm{d}u) && (3.30)\\
&= \mathrm{M}^{\bullet}\phi(x) && (3.31)
\end{aligned}$$

□

Sampling X' according to μ_i from any of its realisation x can be done setting $g = \mathbf{1}_{\mathbf{A}_i}$ in the previous proposition. It is merely generating Z according to $\mathrm{M}(x, \mathrm{d}z)$ and setting $X' = Z$ if $Z \in \mathbf{A}$ and $X' = x$ otherwise.

Algorithm 3.1.1 (From a μ_i–particle to two *via* μ–reversible kernel M)**.** *Let X be distributed according to μ_i and let the transition kernel* M *be μ–reversible. To create another point distributed according to μ_i, do as follows.*

1. *Generate a throw* Z *according to probability measure* $\mathrm{M}(X, \mathrm{d}z)$.
2. *Set* $X' = \begin{cases} Z & \text{if } \vartheta(Z) \geq s_i \text{ i.e.} Z \in \mathbf{A}_i \\ x & \text{otherwise} \end{cases}$.

The only requirement is being able to generate throws according to $\mathrm{M}(x, \mathrm{d}z)$ *for any* $x \in \mathbf{A}_i$.

The last remaining issue is correlation.

The procedure described in algorithm 3.1.1 enables to inflate a μ_i–sample set but the sample set is correlated. Actually, the iterative nature of the algorithm is such that points have a common genealogy. That, at the end of the day, can translate into increased estimator variance and there is no theoretically proven way to avoid that yet. It is however shown in [62, 90] that under mild conditions, iterating algorithm 3.1.1 *ad libitum* cannot increase variance and might even reduce it. One can hence use M_i^ω in *lieu* of M_i where the integer ω is empirically fixed or decided according to the simulation budget or defined as a stopping time *e.g.* the first time when 90% of the points have moved from their original positions.

This whole sampling scheme offers the desired feature as it allows to grow a μ_i distributed set, tough in a correlated identically distributed way. However, it fully depends on a reversible transition kernel M, which still has to be found.

3.1.4 Reversibility or Metropolis-Hastings

The sampling step described in subsection 3.1.3 is of major importance for the whole ST algorithm and heavily depends on the availability of a μ–reversible transition kernel M or at least a collection of μ_i–invariant transition kernels. There are two ways to fill these gaps: solving the reversibility equation (3.13) or building a Metropolis-Hastings kernel.

3.1.4.1 An explicit solution of the reversibility equation

Finding an explicit solution of the reversibility equation, though the most straightforward solution, is often in practice very difficult.

$$\forall x, z \in \mathbb{X}, \mu(\mathrm{d}x)\,\mathrm{M}(x, \mathrm{d}z) = \mu(\mathrm{d}z)\,\mathrm{M}(z, \mathrm{d}x) \tag{3.13}$$

Therefore the equation is to be dealt with on a case by case basis.

Even the usually very convenient centered reduced Gaussian density case can be only partially sorted out: there is an available solution [19].

Property 3.1.4 (A reversible transition kernel in the Gaussian case)**.**

$$\begin{array}{lc} \textit{If} & X \sim \mathcal{N}(0_d, 1_d) \text{ i.e. } f_X(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{\|x\|^2}{2}\right) \\ \textit{then} & Z \sim \mathcal{N}\left(\frac{X}{\sqrt{1+\theta^2}}, \frac{\theta^2}{1+\theta^2} 1_d\right) \text{ i.e. } \mathrm{M}(x, z) = \frac{1}{(2\pi)^{\frac{d}{2}} \frac{\theta}{\sqrt{1+\theta^2}}} \exp\left(-\frac{1}{2}\frac{1+\theta^2}{\theta^2}\left\|z - \frac{x}{\sqrt{1+\theta^2}}\right\|^2\right) \\ \textit{is} & \textit{solution of reversibility equation (3.13)} \ \forall \theta \geq 0. \end{array} \tag{3.32}$$

In the happy very few cases where an explicit solution to Equation 3.13 can be found, tuning the solution is a heuristic handicraft. It can actually lead to a collection of μ–reversible transition kernels to pick from and adapt into μ_i–invariant kernels, such as in the Gaussian case. In a broader perspective however, only the latter are wanted. A direct other way to do it is the Metropolis-Hastings algorithm.

3.1.4.2 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm (MH) is currently one of the most powerful and widespread sampling algorithms. All along its origin [59], generalisation [40] and diffusion [90], the following bold promise accounted for its success: sampling according to any given probability density, which only needs be known up to a multiplicative constant.

Actually, for any given density f_Z that is not concentrated on a single point, MH builds a transition kernel M such that

$$\text{for } f_Z \text{ almost all } x, \forall \mathbf{A} \in \mathcal{X}, \lim_{l \to \infty} \mathrm{M}^l(x, \mathbf{A}) = \int_{\mathbf{A}} f_Z(z)\, \mathrm{d}z \tag{3.33}$$

based on another transition kernel Q chosen by the user in three steps.

Definition 3.1.4 (Metropolis-Hastings transition kernel)**.** *A transition kernel* M *is said to be of Metropolis-Hastings type with respect to a density f_Z if it is built as follows, using transition kernel* Q*.*

1. *Define* $g(x,z) = \begin{cases} \min\left\{\frac{f_Z(z)\mathrm{Q}(z,x)}{f_Z(x)\mathrm{Q}(x,z)}, 1\right\} & \text{if} \quad f_Z(x)\,\mathrm{Q}(x,z) > 0 \\ 1 & \text{if} \quad f_Z(x)\,\mathrm{Q}(x,z) = 0 \end{cases}$.
2. *Define* $h(x,z) = \begin{cases} \mathrm{Q}(x,z)\, g(x,z) & \text{if} \quad x \neq z \\ 0 & \text{if} \quad x = z \end{cases}$.
3. *Set* $\tau(x) = 1 - \int h(x,z)\, \lambda(\mathrm{d}z)$

The Metropolis kernel M *can be written as*

$$\mathrm{M}(x, \mathrm{d}z) = h(x,z)\, \lambda(\mathrm{d}z) + \tau(x)\, \delta_x(\mathrm{d}z) \tag{3.34}$$

There is much to say about a transition kernel built in that fashion. The main questions are how Q should be designed [2, 78] so "not too many" transitions should be done before one can claim the output to be distributed according to f_Z [4, 79] and how convergence can be checked [25] and even when MH should not be used at all [29, 88]. However important these features to the MH algorithm, there is a shortcut to the purpose at hands: building a μ_i–invariant transition kernel.

The good news is that the MH transition kernel M happens to be reversible with respect to its target density f_Z.

Property 3.1.5 (f_Z–invariance of the MH kernel)**.** *With respect to definition 3.1.4, the following holds.*

1. $\forall x, z, f_Z(x)\, h(x,z) = f_Z(z)\, h(z,x)$.
2. M *is f_Z–reversible.*

Assuming μ_i has density $f_{X|i}$ w.r.t. the Lebesgue measure λ, which is in practice very often the case, one can build a Metropolis-Hastings kernel that leaves $f_{X|i}$ invariant. Besides the user has much freedom in its design.

Sampling according to the MH kernel is similar with and as easy as sampling from the $\mu^\bullet$–reversible transition kernel as described in Proposition 3.1.3.

Property 3.1.6 (Sampling w.r.t. a MH transition kernel)**.** *To sample Z according to* $\mathrm{M}(x,\,\mathrm{d}z)$ *as defined in 3.1.4, where* $x \in \mathbb{X}$

1. *sample V according to* $\mathrm{Q}(x,\,\mathrm{d}v)$
2. *set* $Z = \begin{cases} V & \text{with probability} \quad g(x,V) \\ x & \text{with probability} \quad 1-g(x,V) \end{cases}$

Proof. Let ϕ be bounded and measurable.

$$\begin{aligned}
\mathbb{E}[\phi(Z)] &= \mathbb{E}[\mathbb{E}[\phi(Z)|\,V]] && (3.35)\\
&= \mathbb{E}[\phi(V)\,g(x,V) + \phi(x)\,(1-g(x,V))] && (3.36)\\
&= \int_{\mathbb{X}} \phi(v)\,g(x,v)\,\mathrm{Q}(x,\,\mathrm{d}v) + \left(1 - \int_{\mathbb{X}} g(x,v)\,\mathrm{Q}(x,\,\mathrm{d}v)\right)\phi(x) && (3.37)\\
&= \int_{\mathbb{X}} \phi(v)\,[g(x,v)\,\mathrm{Q}(x,\,\mathrm{d}v) + \tau(x)\,\delta_x(\,\mathrm{d}v)] && (3.38)\\
&= \mathrm{M}\phi(x) && (3.39)
\end{aligned}$$

□

In the particular case where the mapping g is an indicator function, the accepting test is deterministic.

In the Gaussian case there is an explicit solution of the reversibility equation. In other cases, one has to use a Metropolis-Hastings type kernel to run the ST algorithm as in [3] or solve the reversibility equation.

3.2 Four Splitting Technique based algorithms

ST requires *a priori* knowledge about the transfer function ϕ to design the threshold sequence, already described in Subsection 3.1.2. However, it can be analysed more easily than its adaptive counterpart, still providing insight about how to run it, as detailed in section 3.3.

3.2.1 The Splitting Technique reference algorithm

The following algorithm is the basis of Splitting Technique.

Algorithm 3.2.1 (ST probability of exceedance estimator)**.** *So as to estimate the probability with which a $y = \vartheta(x)$ system score lays in a given target set $\mathbf{C} = [\,s, \infty\,[$ when the input is random,*

$$\mathbb{P}[X \in \mathbf{A}] = \mathbb{P}[\vartheta(X) \in [\,s, \infty\,[] = \mathbb{P}[Y \in \mathbf{C}]$$

proceed as follows.

1. *Define a finite increasing sequence of thresholds $-\infty = s_0 < \cdots < s_i < \cdots < s_l = s$.*
2. *Set $i = 0$.*
3. *Generate* via *CMC the* iid *sample set $(X_1^i, \cdots, X_n^i)$, where $X_1^i \sim \mu$.*

4. *Apply* ϑ *to its elements to form* $(Y_1^i, \cdots, Y_n^i)$*, where* $Y_k^i = \vartheta\left(X_k^i\right)$.

5. *Until* $i = l - 1$*, do:*

   (a) *Approximate* $\mathbb{P}\left[X \in \mathbf{A}_{i+1} \middle| X \in \mathbf{A}_i\right]$ *with* $\dot{\alpha}_{i+1} = \frac{\operatorname{card}\left\{X_k^i \middle| \vartheta\left(X_k^i\right) \geq s_{i+1}\right\}}{n}$.

   (b) *Replace* $\left\{X_k^i \middle| \vartheta\left(X_k^i\right) < s_{i+1}\right\}$ *points with points uniformly selected with replacement within* $\left\{X_k^i \middle| \vartheta\left(X_k^i\right) \geq s_{i+1}\right\}$ *and call the sample set* $\left(Z_1^{i,0}, \cdots, Z_n^{i,0}\right)$.

   (c) *Set* $j = 0$ *and do as follows until* $j = \omega$*:*

      i. *Generate the proposal set* $\left(Z'^{i,j}_1, \cdots, Z'^{i,j}_n\right)$ *such that* $Z'^{i,j}_k \sim \mathrm{M}\left(Z_k^{i,j}, \mathrm{d}z'^{i,j}_k\right)$.

      ii. *Form* $\left(Z_1^{i,j+1}, \cdots, Z_n^{i,j+1}\right)$ *such that* $Z_k^{i,j+1} = \begin{cases} Z'^{i,j}_k & \text{if } \vartheta\left(Z'^{i,j}_k\right) \geq s_i \\ Z_k^{i,j} & \text{otherwise} \end{cases}$.

      iii. *Increment* j *by one:* $j = j + 1$.

   (d) *Set* $\left(X_1^{i+1}, \cdots, X_n^{i+1}\right) = \left(Z_1^{i,\omega}, \cdots, Z_n^{i,\omega}\right)$ *and* $\left(Y_1^{i+1}, \cdots, Y_n^{i+1}\right)$ *such that* $\forall k, Y_k^{i+1} = \vartheta\left(Z_k^{i,\omega}\right)$.

   (e) *Increment* i *by one:* $i = i + 1$.

6. *Approximate* $\mathbb{P}\left[X \in \mathbf{A}_l \middle| X \in \mathbf{A}_{l-1}\right]$ *with* $\dot{\alpha}_l = \frac{\operatorname{card}\left\{X_k^{l-1} \middle| \vartheta\left(X_k^{l-1}\right) \geq s_l\right\}}{n}$.

7. *Conclude that* $\mathbb{P}\left[Y \in \mathbf{C}\right] \approx \prod_{i=0}^{l-1} \dot{\alpha}_i$

The ST exceedance probability estimator will be denoted

$$\breve{Y} = \breve{Y}\left(n, \mathrm{M}, \omega, s_1, \cdots, s_l, s\right) \tag{3.40}$$

and the total number of generated points equals

$$\gamma = n \times (1 + (l - 1) \times \omega) \tag{3.41}$$

The requirements are the capacity of evaluating ϑ *for any* $\mathbb{X}$ *point and a* μ*–reversible transition kernel* M.

The important role played by the threshold sequence makes this algorithm tricky to implement efficiently, as already described in Subsection 3.1.2. Its adaptive form copes with this issue.

3.2.2 Three Adaptive Splitting Technique algorithms

An AST probability of exceedance estimator algorithm can now be stated. The major difference from the previous ST algorithm is that the threshold sequence is not known beforehand but constructed on the fly *via* random empirical quantiles.

Algorithm 3.2.2 (AST probability of exceedance estimator)**.** *So as to estimate the probability with which a* $y = \vartheta(x)$ *system score lays in a given target set* $\mathbf{C} = [\, s, \infty\, [$ *when the input is random,*

$$\mathbb{P}\left[X \in \mathbf{A}\right] = \mathbb{P}\left[\vartheta\left(X\right) \in [\, s, \infty\, [\right] = \mathbb{P}\left[Y \in \mathbf{C}\right]$$

proceed as follows.

1. *Choose* $\alpha \in]\,0,1\,[$, *based on Subsection 3.1.2 for instance, and set* $i = 0$.
2. *Generate* via *CMC the* iid *sample set* $(X_1^i, \cdots, X_n^i)$, *where* $X_1^i \sim \mu$.
3. *Apply* ϑ *to its elements to form* $(Y_1^i, \cdots, Y_n^i)$, *where* $Y_k^i = \vartheta\left(X_k^i\right)$.
4. *Calculate* $(Y_1^i, \cdots, Y_n^i)$ *empirical* α*-quantile* S_1.
5. *While* $S_{i+1} < s$ *and* $i < \kappa_{max}$, *do:*
   (a) *Replace* $\left\{X_k^i \,|\vartheta\left(X_k^i\right) < S_{i+1}\right\}$ *points with points uniformly selected with replacement within* $\left\{X_k^i \,|\vartheta\left(X_k^i\right) \geq S_{i+1}\right\}$ *and call the sample set* $\left(Z_1^{i,0}, \cdots, Z_n^{i,0}\right)$.
   (b) *Set* $j = 0$ *and do as follows until* $j = \omega$*:*
      i. *Generate the proposal set* $\left(Z'^{i,j}_1, \cdots, Z'^{i,j}_n\right)$ *such that* $Z'^{i,j}_k \sim \mathrm{M}\left(Z_k^{i,j}, \mathrm{d}z'^{i,j}_k\right)$.
      ii. *Form* $\left(Z_1^{i,j+1}, \cdots, Z_n^{i,j+1}\right)$ *such that* $Z_k^{i,j+1} = \begin{cases} Z'^{i,j}_k & \text{if } \vartheta\left(Z'^{i,j}_k\right) \geq S_i \\ Z_k^{i,j} & \text{otherwise} \end{cases}$.
      iii. *Increment* j *by one:* $j = j + 1$.
   (c) *Set* $\left(X_1^{i+1}, \cdots, X_n^{i+1}\right) = \left(Z_1^{i,\omega}, \cdots, Z_n^{i,\omega}\right)$ *and* $\left(Y_1^{i+1}, \cdots, Y_n^{i+1}\right)$ *such that* $\forall k, Y_k^{i+1} = \vartheta\left(Z_k^{i,\omega}\right)$.
   (d) *Increment* i *by one:* $i = i + 1$.
   (e) *Calculate* $(Y_1^i, \cdots, Y_n^i)$ *empirical* α*-quantile* S_{i+1}.
6. *Set* $\kappa = i$ *and* $S_{\kappa+1} = s$.
7. *Approximate* $\mathbb{P}\left[X \in \mathbf{A} \,|X \in \mathbf{A}_\kappa\right]$ *with* $\tau_\kappa = \tau_\kappa(s) = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{[\,s,\infty[}\left(Y_k^\kappa\right)$.
8. *Conclude that* $\mathbb{P}\left[Y \in \mathbf{C}\right] \approx \alpha^\kappa \times \tau_\kappa$

The number of crossed intermediate thresholds i.e. κ *is random. The AST estimator will be denoted*

$$\widetilde{Y}_n = \widetilde{Y}\left(n, \kappa, \alpha, \mathrm{M}, \omega\right) \tag{3.42}$$

and the total number of generated points is random and equals

$$\Gamma = n \times (1 + \kappa \times \omega) \tag{3.43}$$

The requirements are the capacity of evaluating ϑ *for any* $\mathbb{X}$ *point and a* μ*–reversible transition kernel* M.

Atop providing an estimation of $\mathbb{P}\left[Y \in \mathbf{C}\right]$, this algorithm comes with two extras: conditional expectation estimation and quantile estimation.

The first bonus of this algorithm is that it can conveniently be used to generate points distributed according to $\mu_{\kappa+1}$. An appropriate definition of satisfaction criterion ζ can be used[1] to approximate $\mathbb{E}\left[\phi\left(X\right)|\,\mathbf{B}\right]$, for any $\mathbf{B} \in \mathcal{Y}$. For instance, one can approximate

$$\mathbb{E}\left[\phi\left(X\right)|\,\vartheta\left(X\right) > s_i\right] \approx \frac{1}{n}\sum_{k=1}^n \phi\left(X_k^i\right) \tag{3.44}$$

the mean system output given that the score is above threshold s_i.

[1] Remember $\vartheta = \zeta \circ \phi$.

Algorithm 3.2.3 (AST conditional expectation estimator)**.** *So as to estimate the conditional expectation*

$$\mathbb{E}\left[\phi\left(X\right)|\,\vartheta\left(X\right)>s\right]$$

proceed as described in 3.2.2 until 6 then as follows.

6. *Set* $\kappa=i$ *and* $s_\kappa=s$.
7. *Set* $j=0$ *and do as follows until* $j=\omega$*:*
   (a) *Generate the proposal set* $\left(Z'^{\kappa,j}_1,\cdots,Z'^{\kappa,j}_n\right)$ *such that* $Z'^{\kappa,j}_k\sim\mathrm{M}\left(Z^{\kappa,j}_k,\,\mathrm{d}z'^{\kappa,j}_k\right)$.
   (b) *Form* $\left(Z^{\kappa,j+1}_1,\cdots,Z^{\kappa,j+1}_n\right)$ *such that* $Z^{\kappa,j+1}_k=\begin{cases}Z'^{\kappa,j}_k & \text{if } \vartheta\left(Z'^{\kappa,j}_k\right)\geq s_\kappa\\ Z^{\kappa,j}_k & \text{otherwise}\end{cases}$.
   (c) *Increment* j *by one:* $j=j+1$.
8. *Set* $\left(X^\kappa_1,\cdots,X^\kappa_n\right)=\left(Z^{\kappa,\omega}_1,\cdots,Z^{\kappa,\omega}_n\right)$ *and* $\left(Y^\kappa_1,\cdots,Y^\kappa_n\right)$ *such that* $\forall k, Y^\kappa_k=\vartheta\left(Z^{\kappa,\omega}_k\right)$.
9. *Conclude that* $\mathbb{E}\left[\phi\left(X\right)|\,\vartheta\left(X\right)>s\right]\approx\frac{1}{n}\sum_{k=1}^n\phi\left(X^\kappa_k\right)$.

Requirements are the same as 3.2.2.

The second extra benefit is that the algorithm can be used to estimate extreme quantiles as well. Actually, if $s'\in[\,s_i,si+1\,[$, then s' is approximately the $\left(\alpha^i\left(1-\tau_i\left(s'\right)\right)\right)$–quantile of $Y=\vartheta\left(X\right)$.

Algorithm 3.2.4 (AST quantile estimator)**.** *So as to estimate the* $(\alpha^\kappa\beta)$*–quantile* q *of* Y *start as described in 3.2.2. Replace stopping criterion 5 with*

5. *Do* κ *times:*

and conclude

6. *Approximate* q *with* $\widetilde{q}$*, the empirical* β*–quantile of* $\left(X^\kappa_1,\cdots,X^\kappa_n\right)$.

Once both the number of thresholds κ *and the number of kernel application* ω *are fixed, the algorithm cost is fixed as well according to Equation 3.43. Requirements are the same as in Algorithm 3.2.2.*

For rare event probabilities and extreme quantiles are the target, estimators $\widetilde{Y}_n$ and $\widetilde{q}$ will be further studied.

3.3 Asymptotic behaviour

AST expectation estimator $\widetilde{Y}_n$ and quantile estimator $\widetilde{q}$ are random variables. One naturally wants to know their distributions, variance and asymptotic behaviour when the number n of particles goes to infinity. Unfortunately, only asymptotic results in the probability estimation are available. As far as $\widetilde{q}$ is concerned, the practical strategy is tuning the algorithm as if estimating $\mathbb{P}\left[Y>\widetilde{q}\right]$, drawing one's inspiration from the coming results.

Both the Splitting Technique (ST) and the Adaptive Splitting Technique (AST) probability estimator come with a fluctuation analysis in [17]. The results are reproduced below and proposition 3.3.1 proof can be found in [17, proposition 3].

3.3.1 The Splitting Technique

Though in practice only AST will be performed, ST gives valuable information about the usefulness of kernel iteration and the impact of the score function.

Property 3.3.1 (ST exceedance probability estimator asymptotic behaviour)**.** *Let $\breve{Y}$ be the ST exceedance probability estimator and n the number of particles in the cloud. Then,*

$$\sqrt{n}\frac{\breve{Y} - \mathbb{P}\left[Y \in \mathbf{C}\right]}{\mathbb{P}\left[Y \in \mathbf{C}\right]} \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \sigma^2\right) \tag{3.45}$$

with

$$\sigma^2 = \sum_{i=0}^{l-1} \frac{1-\alpha_i}{\alpha_i} + \sum_{i=1}^{l-1} \frac{1}{\alpha_i} \mathbb{E}\left[\left(\frac{\mathbb{P}\left[X^{l-1} \in \mathbf{A}_l \middle| X^i\right]}{\mathbb{P}\left[X^{l-1} \in \mathbf{A}_l \middle| X^{i-1} \in \mathbf{A}_i\right]} - 1\right)^2 \middle| X^{i-1} \in \mathbf{A}_i\right] \tag{3.46}$$

where $\forall i \in \{1, \cdots, l\}, \alpha_i = \mathbb{P}\left[X \in \mathbf{A}_i \middle| X \in \mathbf{A}_{i-1}\right]$.

As [17] authors put it, Proposition 3.3.1 " does not correspond exactly to algorithm 3.2.1. The difference is that the proposition assumes that the resampling is done using a multinomial procedure, which gives a higher variance than that of algorithm 3.2.1. This does not make much difference for the following discussion, as the best possible variance is the same."

The first comment on this result is that the asymptotic variance is lower bounded.

$$\sigma^2 \geq \sum_{i=1}^{l} \frac{1-\alpha_i}{\alpha_i} = \sigma_*^2 \tag{3.47}$$

Equality can be theoretically reached applying the transition kernel an infinite number of times, as already discussed at the end of subsection 3.1.3.

$$\left[\sigma^2 = \sigma_*^2\right] \Leftrightarrow \left[\forall i \in \{1, \cdots, l\} \text{ and knowing } X^{i-1} \in \mathbf{A}_i, \mathbb{P}\left[X^{l-1} \in \mathbf{A}_l \middle| X^i\right] \perp X^i\right] \tag{3.48}$$

The second comment is that for a fixed number of thresholds, say l, σ_*^2 is minimised if all α_i are equal to $\alpha_* = \mathbb{P}\left[Y \in \mathbf{C}\right]^{\frac{1}{l}}$. Indeed, this is the solution of this constrained minimisation problem.

$$\underset{\alpha_0, \cdots, \alpha_{l-1}}{\arg\min}\, \sigma_*^2 \text{ such that } \prod_{i=0}^{l-1} \alpha_i = \mathbb{P}\left[Y \in \mathbf{C}\right] \tag{3.49}$$

This result has already been discussed in Subsection 3.1.2 and gives an hint about how to tune AST.

3.3.2 The Adaptive Splitting Technique

Though it will never be achieved in practice, it is assumed that the transition kernel can be applied infinitely many times so that all particles are considered *iid*. Experiences made in [17] show that the following idealised asymptotic variance can be approximated in practice. Besides, the cumulative distribution function F of $Y = \vartheta\left(X\right)$ is assumed continuous.

Property 3.3.2 (Number of steps)**.** *In algorithm framework the rare event probability can be written*

$$\mathbb{P}\left[Y \in \mathbf{C}\right] = \tau_{\kappa_0}\alpha^{\kappa_0}, \text{ with } \kappa_0 = \left\lfloor \frac{\log\left(\mathbb{P}\left[Y \in \mathbf{C}\right]\right)}{\log\left(\alpha\right)} \right\rfloor \text{ and } \tau_{\kappa_0} = \mathbb{P}\left[Y \in \mathbf{C}\right]\alpha^{-\kappa_0}$$

so that $\tau_{\kappa_0} \in \left]\, \alpha, 1 \,\right]$*. Then, denoting* $c = \min\left(\alpha - \mathbb{P}\left[Y \in \mathbf{C}\right]^{\frac{1}{\kappa_0}}, \mathbb{P}\left[Y \in \mathbf{C}\right]^{\frac{1}{\kappa_0+1}} - \alpha\right)$,

$$\mathbb{P}\left[\kappa \neq \kappa_0\right] \leq 2\left(\kappa_0 + 1\right)e^{-2nc^2} \text{ and as a consequence } \kappa \xrightarrow[n\to\infty]{a.s.} \kappa_0 \tag{3.50}$$

The number of steps is κ_0 with a probability very close to 1 given that n is large enough. Though this probability upper bound depends on the sought value, if the later can be lower bounded, one can forecast an upper bound for the needed simulation budget. Said budget has to be compared with the estimator asymptotic variance.

The following result gives AST's estimator variance in the aforementioned ideal cased.

Theorem 3.3.1 (AST probability estimator's optimal asymptotic behaviour)**.** *If* F_Y*, the cumulative distribution function of* $Y = \vartheta\left(X\right)$ *is continuous, then*

$$\sqrt{n}\frac{\widetilde{Y} - \mathbb{P}\left[Y \in \mathbf{C}\right]}{\mathbb{P}\left[Y \in \mathbf{C}\right]} \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \varsigma^2\right) \text{ where } \varsigma^2 = \kappa_0\frac{1-\alpha_0}{\alpha_0} + \frac{1-\tau_0}{\tau_0} \tag{3.51}$$

Besides, $\widetilde{Y}$ is a biased estimator as stated in the next property. This bias is always positive, conservatively overestimating the probability of dreadful events. It decreases with a $\frac{1}{n}$ rate, being thus negligible with respect to the variance.

Property 3.3.3 (AST probability estimator's bias)**.** *If* F_Y*, the cumulative distribution function of* $Y = \vartheta\left(X\right)$ *is continuous, then*

$$n\frac{\mathbb{E}\left[\widetilde{Y}\right] - \mathbb{P}\left[Y \in \mathbf{C}\right]}{\mathbb{P}\left[Y \in \mathbf{C}\right]} \xrightarrow[n\to\infty]{} \kappa_0\frac{1-\alpha_0}{\alpha_0} \tag{3.52}$$

Eventually, one can write the following expansion

$$\widetilde{Y} = \mathbb{P}\left[Y \in \mathbf{C}\right]\left(1 + \frac{1}{\sqrt{n}}\sqrt{\kappa_0\frac{1-\alpha_0}{\alpha_0} + \frac{1-\tau_0}{\tau_0}}Z + \frac{1}{n}\kappa_0\frac{1-\alpha_0}{\alpha_0} + o_{\mathbb{P}}\left(\frac{1}{n}\right)\right) \text{ with } Z \sim \mathcal{N}\left(0,1\right) \tag{3.53}$$

and notice $\widetilde{Y}$ has a smaller mean square error than $\breve{Y}$ in spite of its bias.

3.4 Conclusion

The Adaptive Splitting Technique rephrases the rare event probability as the product of not so rare events that are defined on the fly. This limits the necessary beforehand user decision to tuning the intermediate quantile level and the transition kernel's parameters. Given the former is fixed, the later can even be constructed adaptively. This way, AST does not need *a priori* knowledge about ϕ. This compensates the extra technicality of the method with respect to Crude Monte Carlo or Importance Sampling. Experiments, however, are needed to develop a real experience of the different methods.

Conclusion

Crude Monte Carlo (CMC), the most used probability or quantile estimation technique, is reviewed in Chapter 1. CMC is well known theoretically, easy to implement and of very convenient use. However, it is essentially costly when it comes to rare events. This is why we have review rare event dedicated probability techniques: Importance Sampling (IS) and more specifically Non-parametric Adaptive Importance Sampling (NAIS) and Cross-Entropy (CE) in Chapter 2 and the Adaptive Splitting Technique (AST), the adaptive form of the Splitting Technique, in Chapter 3.

Importance Sampling seems formally and practically similar to CMC as its fundamental idea is performing convenient CMC after a *well-chosen* change of measure. Designing this change of measure is the tricky part. CE suggests choose it within a parametric density family as the one which minimizes the Kullback-Leibler divergence w.r.t. the optimal change of measure. NAIS proposes to use an educated kernel based auxiliary density to approximate the optimal change of measure. In any cases, this optimal change of measure has to be learnt *via* costly simulations. The idea of [5, 15] is training the candidate auxiliary density with a surrogate of the real system built, *via* kriging [22] for instance, at a low simulation cost. Doing so involves an extra technical layer, building a representative surrogate, but implies better chosen simulations of the real system. This strategy can be used with any IS algorithm so we will focus on growing some practical experience of CE and NAIS.

The Adaptive Splitting Technique rephrases the target very small probability as the product of conditional not larger probabilities and estimates them iteratively. This new formulation makes the algorithm slightly more complex than CMC. However, AST requires a Markov kernel that is reversible w.r.t. the original random variable and this is the first difficult feature. Such a kernel is known in the Gaussian case but in most real life situations, one will have to use a Metropolis-Hastings type algorithm to build one. Such an algorithm seems delicate to handle, mostly because of extra parameters that add atop the numerous ones of AST. Tuning therefore appears as the main difficulty of this algorithm as most parameters have to be fixed almost arbitrarily and with little rational. In the absence of theoretical answers, we will use experiments to find out what can be expected from AST as is.

As a consequence of this review, we tried CE, NAIS and AST out on toy cases first then on a real case.

Part II

Adaptive Non-parametric Adaptive Importance Sampling, Robustness and the Iridium-Cosmos case

Introduction

Crude Monte Carlo (CMC) – Chapter 1–, though easy to implement and of very convenient use, can estimate neither minute probabilities nor extreme quantiles at a reasonable cost. In order to deal with these rare events, three methods had been reviewed.

1. Cross-Entropy (CE) : Algorithm 2.2.1.

2. Importance Sampling (IS) : Algorithm 2.2.2.

3. Adaptive Splitting Technique (AST) : Algorithm 3.2.2 and Algorithm 3.2.4.

We then needed to decide which methods we would focus our attention on in order to cope with our specific aim: providing the lab with a user-friendly rare event dedicated probability and quantile estimator that operates on black-box complex systems. Therefore, we specified what is expected from the estimator, built tests and described how aforementioned techniques responded to them.

We do not claim our tests were thorough nor think all meaningful aspects of the estimators were considered. With the crash test, we wanted to grow some experience of the estimators. With the robustness tests, we saw how they behaved with high dimension inputs and how rare an event they could handle so as to mimic the practical situations of interest. The final aim was an educated choice of method when facing a given task.

Evaluation criteria

The methods were compared according to four aspects.

- Cost: the method simulation cost should be as small as possible,

- Applicability: the method must be able to operate on a black box mapping and without prior,

- User-friendliness: the method should be as self-tuning as possible,

- Estimation duration: the swifter, the better.

Let us detail a bit.

Cost

An estimator requires time and calculation power to deliver but both are limited and hence expensive. In practice, there is a trade-off to be made between estimation accuracy and cost as the bigger the budget, the more accurate the estimation can be. With this respect, the estimator choice is of major importance.

In the situations the lab faces, the calls to the transfer function ϕ consume almost all the simulation budget. The later will be modeled as the number N of available calls to ϕ. We compared the variance yielded by the estimators when allocated the same simulation budget and mitigate them with their bias, if any.

Applicability

The transfer function ϕ is a black-box as it stands for a simulation software patchwork with no tractable analytical counterpart. Therefore, there is no insight about which part of the input space is of sampling interest. Capability to deliver even when dealing with a black box was of major importance.

User-friendliness

Rare events are a concern in various aspects of system performance and safety assessment and people without probability expertise will have to deal with them as well. The selected estimator(s) should be as easy to use as possible so as to spread efficiently. Therefore, complex *ad hoc* tuning was avoided as much as possible.

Estimation duration

Though a small variance is the main objective, out of two otherwise equivalent algorithm, the faster was preferred. Actually, the algorithm duration will in most practical cases be negligible with respect to the system simulation time. Algorithm durations are nonetheless given for the sake of the cases when a fast estimation is required.

Chapter 4

Two crash tests

After reviewing the methods, we needed some practical experience so as to devise a selection rule. We decided to start with very simple toy cases in order to really focus on the algorithms and better identify their main strengths and weaknesses.

4.1 Rayleigh law in dimension two

First was a toy case to build some experience of the methods. To keep things simple, we chose a Gaussian input transferred into a Rayleigh random variable *via* the Euclidean norm.

$$X \sim \mathcal{N}(0_2, I_2) \qquad \phi : \left| \begin{array}{ccc} \mathbb{R}^2 & \to & \mathbb{R} \\ x & \mapsto & \|x\| \end{array} \right. \qquad Y = \phi(X) \sim \mathcal{R}_2 \tag{4.1}$$

This way, we had a complete theoretical knowledge of all random variables at hands.

$$\forall x \in \mathbb{R}^2, \forall (r,a) \in \mathbb{R}^+ \times [\,0, 2\pi\,[\qquad \forall y \in \mathbb{R}, \forall \alpha \in \,]\,0, 1\,[\tag{4.2}$$

$$f_X(x) = \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right) \qquad f_Y(y) = \underset{\mathbb{R}^+}{\mathbf{1}}(y)\, y e^{-\frac{y^2}{2}} \tag{4.3}$$

$$f_X(r,a) = \frac{1}{2\pi} r \exp\left(-\frac{r^2}{2}\right) \qquad F_Y(y) = \underset{\mathbb{R}^+}{\mathbf{1}}(y)\left(1 - e^{-\frac{y^2}{2}}\right) \tag{4.4}$$

$$\{(r,a)\,|\phi(r,a) \geq q > 0\} = [\,q, \infty\,[\times [\,0, 2\pi\,[\qquad F_Y^{-1}(\alpha) = \sqrt{-2\ln(1-\alpha)} \tag{4.5}$$

We then set targets, simulation budget, and number of estimation to be done with each estimator

$$\alpha = 1 - 10^{-5} \qquad q = F_Y^{-1}(\alpha) \approx 4.79 \qquad N = 5 \cdot 10^4 \qquad m = 100 \tag{4.6}$$

and tried to estimate $\mathbb{P}[Y \geq q] = \mathbb{E}[V] = 1 - \alpha = 10^{-5}$, where $V = \mathbf{1}_{\{\phi(X) \geq q\}}$, and random variable Y α-quantile q with given simulation budget N *via* the methods to be tested.

4.1.1 Cross-Entropy

4.1.1.1 Algorithm 2.2.1 settings

At first, we failed to use Cross-Entropy Algorithm 2.2.1.

The first issue arose when we had to choose the family of parametric densities to search. Without knowledge of the geometry of $\{\phi \geq q\}$, one is clueless and can only rely on wild guesses or make the convenient choice of Gaussian distributions. That reduces CE user-friendliness. However, we had extra information and wondered what CE is capable of when it does work. Given $\{\phi \geq q\}$, a distribution that is invariant with respect to rotations around the origin and mainly loads outside the circle whose center is the origin and radius is q seemed a good choice. So we chose the set of bidimensional Gaussian random variables. Hopefully, CE would return one with mean 0_2 and a very large isotropic variance.

Though [9] advices to chose the auxiliary density function the Natural Exponential Family (NEF) because this makes the optimization step analytically tractable, we wanted to see if one can choose outside the NEF. Based on this test, we considered sticking to NEF is the safest way.

The Gaussian distribution with fixed variance is within the NEF. However, to take advantage of all the available distribution and we chose the centered Laplace distribution because its tails are heavier than those of the Gaussian. Having so much information about the optimal change of measure would not be possible in real cases.

$$\theta \in \mathbb{R}_+^* \qquad f_Z^\theta(z) = \frac{1}{4\theta^2} \exp\left(-\frac{|z_1| + |z_2|}{\theta}\right) \tag{4.7}$$

$$\theta_{k+1} = \frac{\sum_{i=1}^n \mathbf{1}_{\left\{\|Z_i^{\theta^k}\| \geq s_{k+1}\right\}} \mathrm{w}^{\theta^k}\left(Z_i^{\theta^k}\right)\left(\left|Z_{i,1}^{\theta^k}\right| + \left|Z_{i,2}^{\theta^k}\right|\right)}{\sum_{i=1}^n \mathbf{1}_{\left\{\|Z_i^{\theta^k}\| \geq s_{k+1}\right\}} \mathrm{w}^{\theta^k}\left(Z_i^{\theta^k}\right)} \tag{4.8}$$

We then set the initial auxiliary density parameter θ^0 so that it did not load the 0_2 region more than f_X and blanketed its tails, as can be seen on Figure 4.1. The empirical quantile level $\beta_\flat$ was chosen close to 1 so as to reach target threshold fast.

$$\theta = \sqrt{\frac{\pi}{2}} \qquad \beta_\flat = 0.9 \tag{4.9}$$

As for the quantile estimator, we chose the weight normalised cdf $\widehat{F}_{3Y}$ as *per* Definition 2.3.1.

4.1.1.2 Experimental results

Desired probability and quantile were estimated at the same time and the whole estimation budget was always exactly depleted at each estimation. The results are presented in Figure 4.2. They are summarized as follows, where T is the estimation duration in seconds.

$$\overline{\widehat{T_\flat}}_m = 0.04 \qquad \overline{\widehat{V_\flat}}_m = 9.96 \cdot 10^{-6} \qquad \overline{\widehat{q_\flat}}_m = 4.80 \tag{4.10}$$

$$\rho\left(\widehat{T_\flat}, m\right) = 0.50 \qquad \rho\left(\widehat{V_\flat}, m\right) = 0.03 \qquad \rho\left(\widehat{q_\flat}, m\right) = 0.00 \tag{4.11}$$

Estimations are accurate and show very little relative variances.

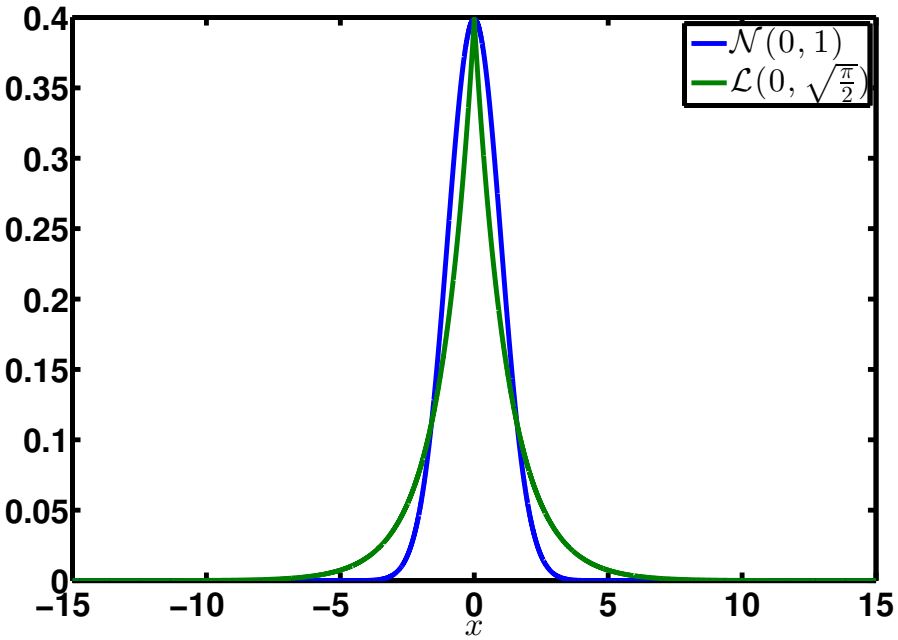


Figure 4.1: Gaussian and Laplacian distributions: $\mathcal{L}\left(0, \sqrt{\frac{\pi}{2}}\right)$ has heavier tails than $\mathcal{N}(0,1)$ and blankets it on its outermost region.

4.1.2 Non parametric Adaptive Importance Sampling

4.1.2.1 Algorithm 2.2.2 settings

Checking the bandwidth hypothesis described in [94] is impossible in practice. We therefore used the `Kernel Density Estimator (KDE) MATLAB toolbox` developed by Alexander Ihler and Mike Mandel[1] with Gaussian kernels and the Asymptotic Mean Integrated Square Error (`AMISE`) bandwidth selection algorithm described in [86, 93]. `AMISE` is a classical bandwidth selection estimator which aims at copying the sampling density as much as possible whereas our target is approximating the optimal auxiliary density. This led us to counter-intuitively use zero weight points when calculating the bandwidth. We kept AMISE nonetheless in the absence of a dedicated tool.

Sharing out the simulation budget was an issue as well. Budget has to be shared into c batches of u_i points so that $N = \sum_{i=1}^{c} u_i$. What should one do: big c and small u or the other way around or a well-informed trade-off? We decided arbitrarily to set $c = 50$ and $u = 10^3$. The main concern however was the algorithm's important dependence on its initial density.

We first chose $f_X = \mathcal{N}(0_2, I_2)$ as initial distribution as it seemed a natural choice in the absence of any other information. The algorithm stopped straight away as no point had been drawn in $\{\phi \geq q\}$ and no kernel density estimation could be done at all. As $\{\phi \geq q\}$ is the rare event we are dealing with, that was to be expected. This highlighted the initial density choice as crucial. When nothing will be known about the geometry of the input subset of interest, most likely, one will find NAIS difficult to use. The initial sampling distribution was

[1] See `http://www.ics.uci.edu/~ihler/code/kde.html` and contact ihler@alum.mit.edu.

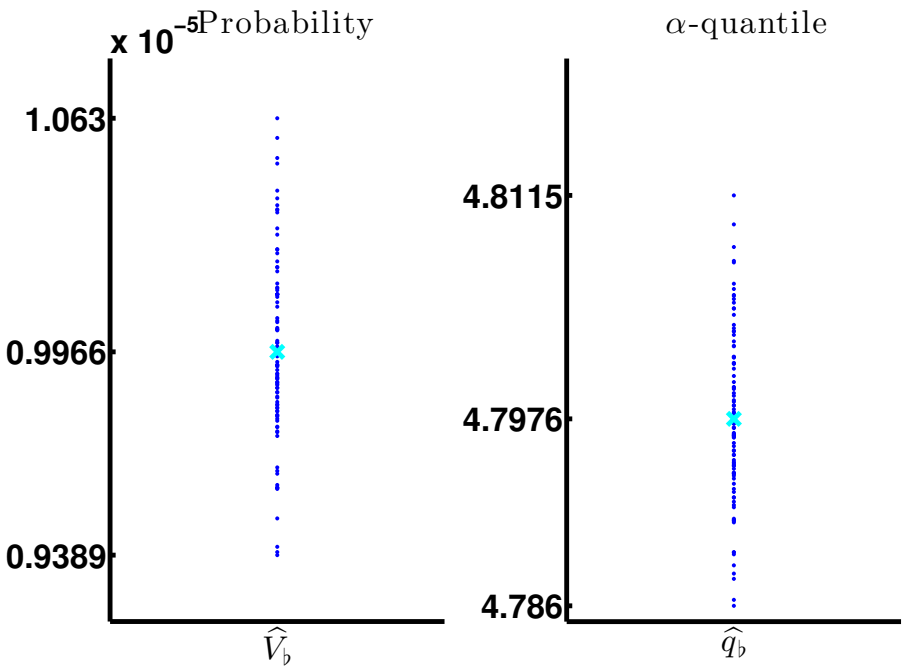


Figure 4.2: Cross-Entropy probability and quantile estimations as *per* Section 4.1. CE is tested against $\mathcal{R}_2$. Estimated values are represented with blue dots $\cdot$ and the mean estimate with a cyan cross $\times$.

eventually fixed as $\mathcal{N}(0_2, 10^2 I_2)$, taking benefit from the available extra information.

As for quantile estimation, we used the empirical Importance Sampling cdf renormalised by the sum of the weights $\widehat{F}_{3Y}$ as *per* Definition 2.3.1.

4.1.2.2 Experimental results

The results are presented in Figure 4.3.

$$\overline{\widehat{T}}_{\bullet m} = 87.90 \qquad \overline{\widehat{V}}_{\bullet m} = 1.00 \cdot 10^{-5} \qquad \overline{\widehat{q}}_{\bullet m} = 4.80 \tag{4.12}$$

$$\rho\left(\widehat{T}_\bullet, m\right) = 0.01 \qquad \rho\left(\widehat{V}_\bullet, m\right) = 0.04 \qquad \rho\left(\widehat{q}_\bullet, m\right) = 0.01 \tag{4.13}$$

Estimations are accurate and show little relative variances.

4.1.3 Adaptive Splitting Technique

We did not estimate the probability and the quantile at the same time. It is possible to do so, but tuning the algorithm is purpose dependent and we wanted to see how the technique responded in each case.

4.1.3.1 Algorithm 3.2.2 settings: probability of exceedance estimation

Tuning was difficult. A poor combination of parameters can lead to a higher variance. Besides, the simulation cost being random, to choose the parameters so that the simulation budget is

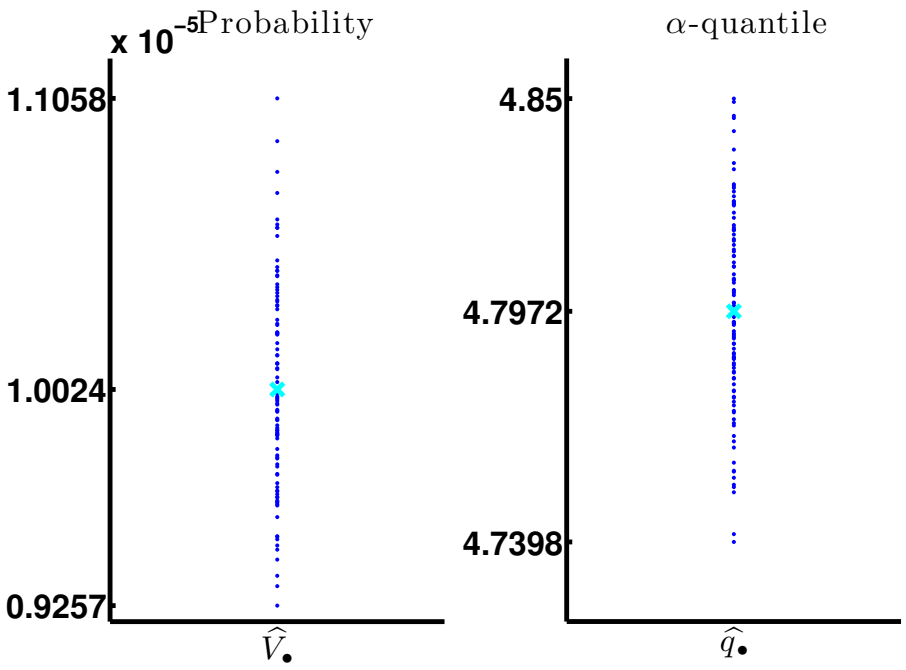


Figure 4.3: NAIS probability and quantile estimations as *per* Section 4.1. NAIS is tested against $\mathcal{R}_2$. Estimated values are represented with blue dots $\bullet$ and the mean estimate with a cyan cross $\times$.

rarely exceeded we based our choice on Proposition 3.3.2 aiming $\Gamma = 50 \cdot 10^3$. If the target probability is unknown this is not possible and the simulation budget can be depleted before the estimation is completed. A solution might then be to aim for a probability threshold under which all probabilities are assumed to be zero and set AST parameters so as to aim this threshold. If the event of interest is met before complete budget depletion, one can spend the remnant so as to better estimate the probability of interest.

$$\omega = 25 \qquad n = 220 \qquad \widetilde{\beta} = 0.29 \tag{4.14}$$

Sampling according to the conditional laws was easy because the input was Gaussian. We used the transition kernel proposed in Proposition 3.1.4. However, choosing and tuning θ requires sleight and insight about the system at hands. The heuristic adaptive tuning procedure we used follows.

Property 4.1.1 (Heuristic AST tuning)**.** *Start with θ^0 and multiply θ by ι if more than* 50% *of the particles have moved after the previous use of* M_i *and dividing is by ι otherwise.*

We used $\iota = 1.1$ and $\theta^0 = 1$. Had the input not been Gaussian, we would have had to design a Metropolis-Hastings transition kernel, which is not always straightforward and involves additional tuning parameters.

4.1.3.2 Experimental results: probability of exceedance estimation

The estimation cost Γ was random as expected and exceeded available budget N four times out of a hundred and always by 5220 points. This can be seen in Figure 4.4.

$$\overline{\Gamma}_m = 49445 \qquad \overline{\widetilde{T}}_m = 0.02 \qquad \overline{\widetilde{V}}_m = 9.56 \cdot 10^{-6} \tag{4.15}$$

$$\rho(\Gamma, m) = 0.04 \qquad \rho\left(\widetilde{T}, m\right) = 0.15 \qquad \rho\left(\widetilde{V}, m\right) = 0.37 \tag{4.16}$$

Simulations were fast, estimations were accurate though not really consistent.

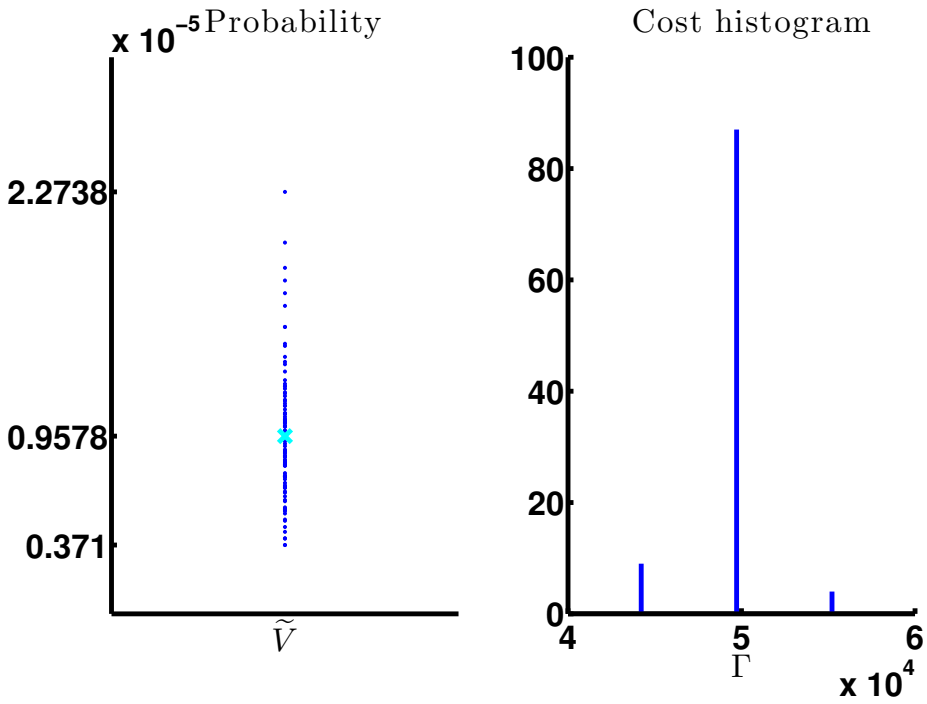


Figure 4.4: AST probability estimations as *per* Section 4.1. AST is tested against $\mathcal{R}_2$ to estimate a probability. Estimated values are represented with blue dots $\cdot$ and the mean estimate with a cyan cross $\times$.

4.1.3.3 Algorithm 3.2.4 settings: quantile

We used the same settings as when estimating the probability and set $\kappa = 9$.

4.1.3.4 Experimental results: quantile

As theoretically forecasted in Algorithm 3.2.4, the simulation budget was never exceeded as shown in Figure 4.5.

$$\overline{\Gamma}_m = 49720 \qquad \overline{\widetilde{T}}_m = 0.02 \qquad \overline{\widetilde{q}}_m = 4.94 \tag{4.17}$$

$$\rho(\Gamma, m) = 0 \qquad \rho\left(\widetilde{T}, m\right) = 0.02 \qquad \rho(\widetilde{q}, m) = 0.01 \tag{4.18}$$

calculations were fast, estimations accurate and consistent.

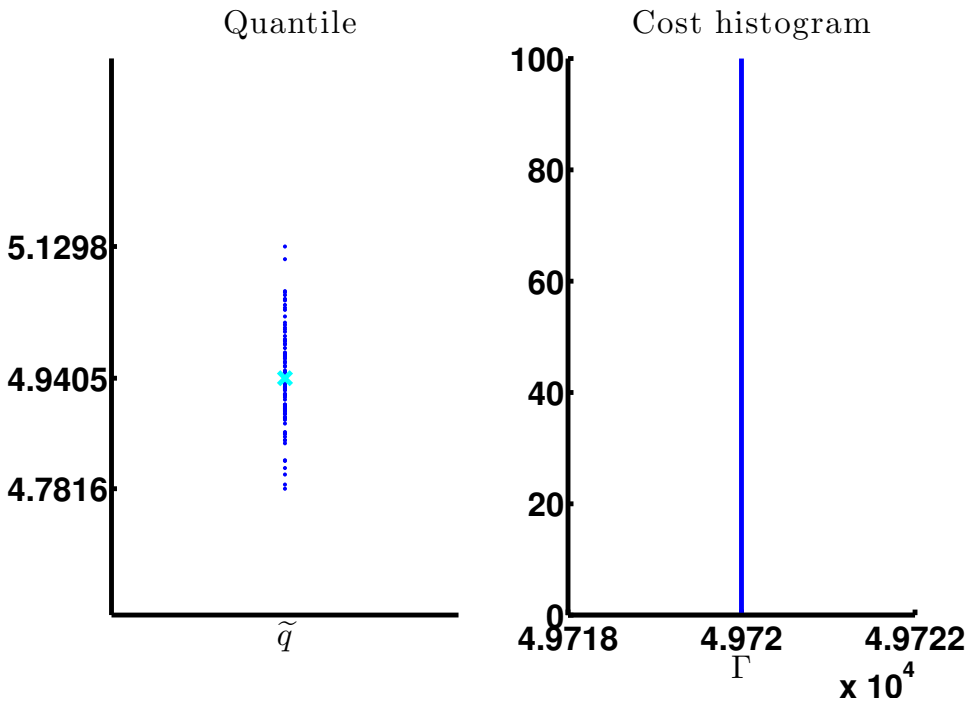


Figure 4.5: AST quantile estimations as *per* Section 4.1. AST is tested against $\mathcal{R}_2$ to estimate a quantile. Estimated values are represented with blue dots $\cdot$ and the mean estimate with a cyan cross $\times$.

4.1.4 Bidimensional Rayleigh toy case teachings

Cross-Entropy goes through an optimization step that can be a problem of its own. However, if one chooses the auxiliary density among the Natural Exponential Family , the optimization step is solved analytically and the estimations can be fast and accurate. If for some reason, the NEF is known to be a poor choice, if the subset of interest shows two modes for instance, CE can be very cumbersome to use.

Non parametric Importance Sampling suffers too high a dependence on the initial distribution as the later can cripple the quantile estimation if ill-chosen. The initial pdf should generate some points in the rare set so as to help the user design a pdf that more frequently generates point in the rare set. This lies between the vicious circle and the very slowly diverging spiral. NIS seems of limited use without *a priori*.

Adaptive Importance Splitting shows higher variance than the two previous methods and depends heavily on tunings. It does not seem usable without a good rational tuning rule or spending a significant amount of the budget in learning a good tuning choice.

4.2 Rayleigh law in dimension twenty

This second toy case was designed to find out the estimators behaviours when faced with an input of high dimension because complex systems have many inputs. We chose a $\mathcal{R}_{20}$ random

variable.

$$X \sim \mathcal{N}(0_{20}, I_{20}) \qquad \phi : \left| \begin{array}{ccc} \mathbb{R}^{20} & \to & \mathbb{R} \\ x & \mapsto & \|x\| \end{array} \right. \qquad Y = \phi(X) \sim \mathcal{R}_{20} \tag{4.19}$$

This time, we did not know Y's α-quantile beforehand and had to estimate it *via* CMC using a huge sample set.

$$\forall i \in \left\{1, \cdots, 3 \cdot 10^7\right\}, X_i \overset{\text{iid}}{\sim} \mathcal{N}(0_{20}, I_{20}) \quad Y_i = \phi(X_i) \quad \alpha = 1 - 10^{-5} \quad \overline{q_\alpha} = 7.6696 \tag{4.20}$$

We then set targets accordingly and kept the same available simulation budget as in round I.

$$\mathbb{P}[Y \geq \overline{q_\alpha}] =? \qquad F_Y^{-1}(\alpha) =? \qquad N = 5 \cdot 10^4 \qquad m = 100 \qquad V = \underset{\{\phi(X) \geq \overline{q_\alpha}\}}{\mathbf{1}} \tag{4.21}$$

We proceeded as during Round I and kept the same settings.

4.2.1 Cross-Entropy

Results can be seen on Figure 4.6.

$$\overline{\widehat{T}}_{\flat m} = 0.33 \qquad \overline{\widehat{V}}_{\flat m} = 1.10 \cdot 10^{-5} \qquad \overline{\widehat{q}}_{\flat m} = 7.69 \tag{4.22}$$

$$\rho\left(\widehat{T}_\flat, m\right) = 0.11 \qquad \rho\left(\widehat{V}_\flat, m\right) = 0.33 \qquad \rho\left(\widehat{q}_\flat, m\right) = 0.01 \tag{4.23}$$

Results shown on Figure 4.2, probability estimations exhibit more variance than those of Subsubsection 4.1.1.2 . On the other hand, the quantile estimation maintained a low relative variance.

4.2.2 Non parametric Adaptive Importance Sampling

The results are presented in Figure 4.7.

$$\overline{\widehat{T}}_{\bullet m} = 313.63 \qquad \overline{\widehat{V}}_{\bullet m} = 3.55 \cdot 10^{-54} \qquad \overline{\widehat{q}}_{\bullet m} = 23.83 \tag{4.24}$$

$$\rho\left(\widehat{T}_\bullet, m\right) = 0.06 \qquad \rho\left(\widehat{V}_\bullet, m\right) = 10.00 \qquad \rho\left(\widehat{q}_\bullet, m\right) = 0.08 \tag{4.25}$$

The estimation was a failure: it took a very long time and both the probability and the quantile estimator show great bias and empirical relative deviations. This may be due to the way `KDE` was implemented and the density being essentially costly to evaluate.

4.2.3 Adaptive Splitting Technique

4.2.3.1 Experimental results: probability

Results can be seen on Figure 4.8.

$$\overline{\Gamma}_m = 49280 \qquad \overline{\widetilde{T}}_m = 0.07 \qquad \overline{\widetilde{V}}_m = 9.94 \cdot 10^{-6} \tag{4.26}$$

$$\rho(\Gamma, m) = 0.04 \qquad \rho\left(\widetilde{T}, m\right) = 0.08 \qquad \rho\left(\widetilde{V}, m\right) = 0.31 \tag{4.27}$$

The estimator had the same behaviour as in Subsection 4.1.3, Figure 4.4.

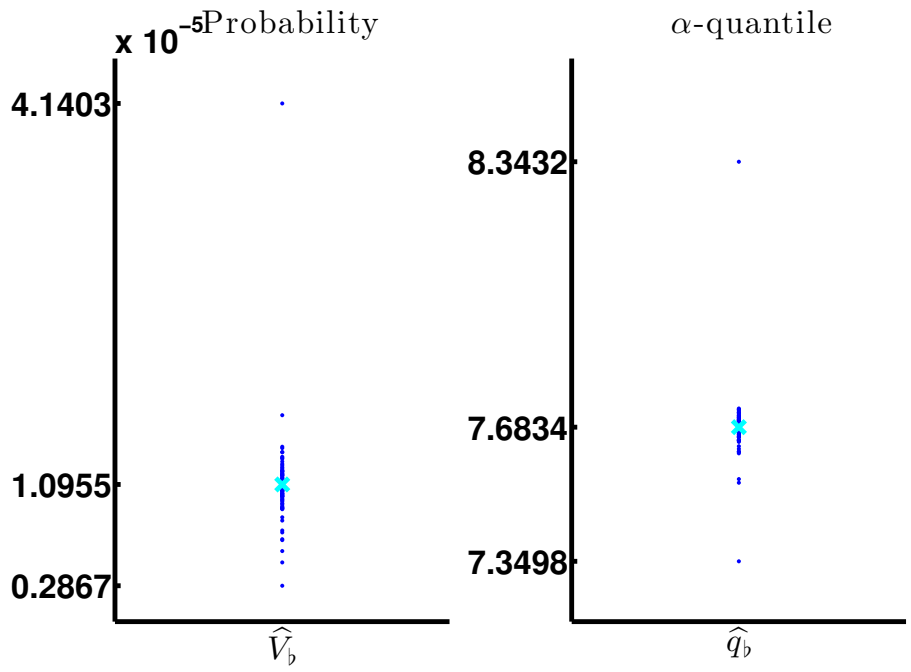


Figure 4.6: Cross-Entropy probability and quantile estimations as *per* Section 4.2. CE is tested against $\mathcal{R}_{20}$. Estimated values are represented with blue dots $\bullet$ and the mean estimate with a cyan cross $\times$.

4.2.3.2 Experimental results: quantile

Results can be seen on Figure 4.9.

$$\overline{\Gamma}_m = 49720 \qquad \overline{\widetilde{T}}_m = 0.08 \qquad \overline{\widetilde{q}}_m = 7.60 \tag{4.28}$$

$$\rho\left(\Gamma, m\right) = 0 \qquad \rho\left(\widetilde{T}, m\right) = 0.04 \qquad \rho\left(\widetilde{q}, m\right) = 0.01 \tag{4.29}$$

The estimator had the same behaviour as Subsection 4.1.3, Figure 4.5.

4.2.4 High dimension toy case teaching

The NAIS estimator fails in high dimension, victim of the curse of dimensionality cast on all kernel density estimators. CE and AST can estimate high dimension quantile estimations with little bias and empirical relative deviation. As for high dimension probability estimation, CE and AST show little bias and a higher empirical relative deviation than in little dimension.

4.3 Conclusion

We built some experience of the Cross-Entropy (CE), Non parametric Adaptive Importance Sampling (NAIS) and Adaptive Splitting Technique (AST) estimating a 10^{-5} probability and $(1 - 10^{-5})$-quantile twice.

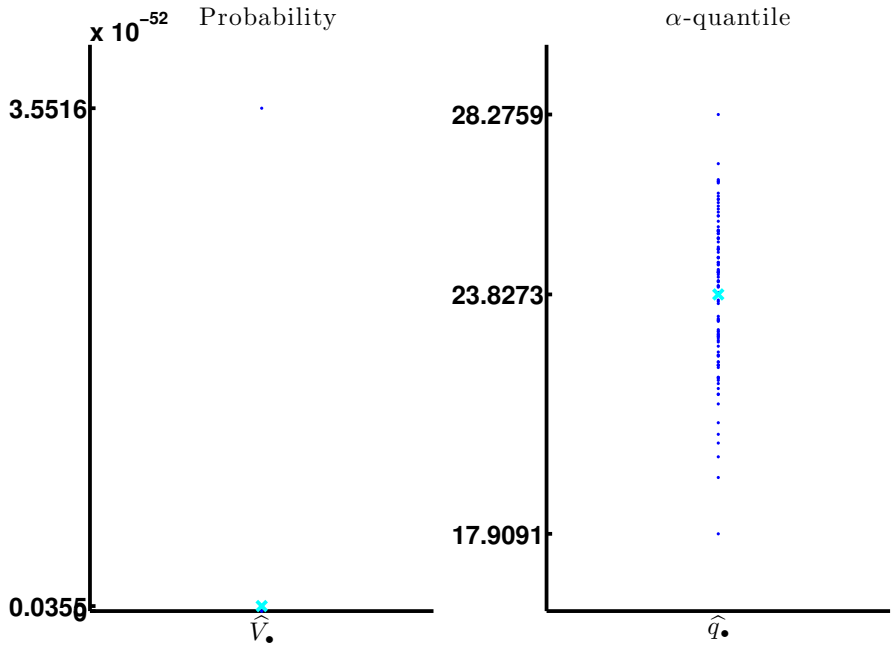


Figure 4.7: NAIS probability and quantile estimations as *per* Section 4.1. NAIS is tested against $\mathcal{R}_{20}$. Estimated values are represented with blue dots $\bullet$ and the mean estimate with a cyan cross $\times$.

We first used a $\mathcal{R}_2$ *i.e.* a dimension 2 Rayleigh random variable as this random variable is completely known theoretically, Section 4.1. It came out that though all estimators eventually delivered accurately and reasonably fast, they all come with their own drawbacks.

- CE's optimisation step is cumbersome as soon as the the auxiliary density is outside the Natural Exponential Family such as when the subset of interest is multi-modal.
- NAIS can not work until the initial distribution does generate rare events.
- AST requires a defter tuning that induces a random and hard to control cost, and that may result in a higher variance.

The second test involved a $\mathcal{R}_{20}$. This way, the estimator behaviours w.r.t. a high dimension input was observed, as described in Section 4.2.

- NAIS fails in high dimension and should not be used in such cases.
- CE and AST quantile estimations were not affected by the dimension increase.
- CE and AST estimated the probability without bias but with higher variance than with $\mathcal{R}_2$.

CE and AST appeared as stable with respect to high dimensions.

As a consequence of these first two tests, we decided not to use NAIS because it depends too much on its initial auxiliary density and seemed unable to deliver in high dimension. However,

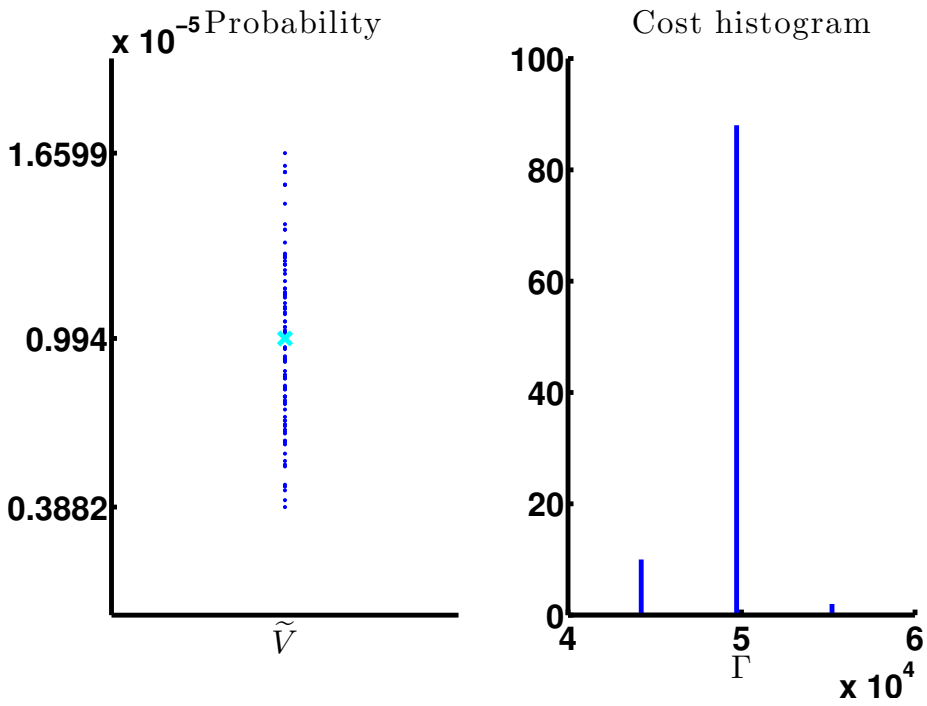


Figure 4.8: AST probability estimations as *per* Section 4.2. AST is tested against $\mathcal{R}_{20}$ to estimate a probability. Estimated values are represented with blue dots $\cdot$ and the mean estimate with a cyan cross $\times$.

out of consideration for its accuracy and consistence when it does work, we looked for a way to resolve NAIS issues keeping the promising idea of iteratively improve the auxiliary density in Chapter 5.

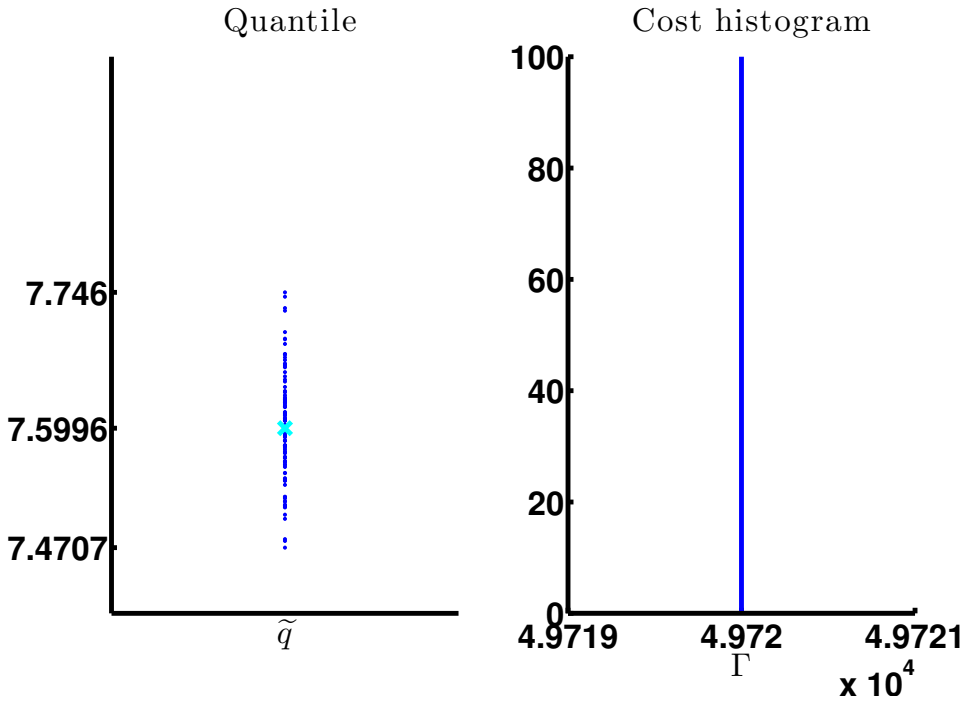


Figure 4.9: AST quantile estimations as *per* Section 4.2. AST is tested against $\mathcal{R}_{20}$ to estimate a quantile. Estimated values are represented with blue dots $\cdot$ and the mean estimate with a cyan cross $\times$.

Chapter 5

ANAIS: the natural sampling density as initial NAIS auxiliary density

When using the original NAIS Algorithm 2.2.2 to estimate $\mathbb{E}\left[\mathbf{1}_{\{\phi(X)\leq q\}}\right]$, the shoe pinches on the initial density f_Z°. As having points in $\{\phi \geq q\}$ is compulsory to start the iterative building of the auxiliary density, the set should not be rare with respect to pdf f_Z°. NAIS builds a better auxiliary density if given one that is already good. Without a good f_Z°, one is stuck in square one. We hence proposed a modification of NAIS to cope with this issue: the Adaptive Non parametric Adaptive Importance Sampling (ANAIS)[1] . First we built up the algorithm – Section 5.1–. Then, the new algorithm was put through the selection process – Section 5.2 –.

5.1 Equipping NAIS with adaptive thresholds

The seminal idea was the increasing threshold from both CE and AST – Subsection 5.1.1 –. Besides, we suggested not taking into account all batches of points to build the desired estimators – Subsection 5.1.2 –. The ANAIS algorithm could then be stated – Subsection 5.1.3 –.

5.1.1 Building the auxiliary probability density functions

The set of interest $\{\phi \geq q\}$ is defined by a threshold. In such cases, both CE and AST advice to iteratively build an increasing sequence of thresholds *via* empirical β-quantiles

$$\forall i \in \{1, \cdots, n_k\}, Z_i^{\dagger,k} \overset{\text{iid}}{\sim} f_Z^{\dagger,k} \qquad C_i^{\dagger,k} = \phi\left(Z_i^{\dagger,k}\right) \qquad S_{k+1} = \min\left\{C^{\dagger,k}_{(\lfloor \beta n \rfloor + 1)}, q\right\} \tag{5.1}$$

[1]The article transcribed here is in reviewing process, as of writing time.

and construct an Importance Sampling density that given a threshold aims at estimating the probability to be above the next one[2]

$$\mathrm{w}^{\dagger,k+1} = \frac{f_X}{f_Z^{\dagger,k+1}} \qquad f_Z^{\dagger,k+1}(z) = \frac{\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(\mathbf{1}_{\{\phi\geq S_{k+1}\}}\,\mathrm{w}^{\dagger,i}\right)\left(Z_j^{\dagger,i}\right)K_d\left(Z_j^{\dagger,i}-z,h^{\dagger,k+1}\right)}{|h^{\dagger,k+1}|\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(\mathbf{1}_{\{\phi\geq S_{k+1}\}}\,\mathrm{w}^{\dagger,i}\right)\left(Z_j^{\dagger,i}\right)} \tag{5.2}$$

$$\text{where } \forall x\in\mathbb{R}^d,\quad \forall h\in\left(\mathbb{R}^+\right)^d,\forall t\in\mathbb{R}, \tag{5.3}$$

$$|h| = \prod_{i=1}^{d} h_i \qquad K_d(x,h) = \prod_{i=1}^{d} K\left(\frac{x_i}{h_i}\right), \left(\underset{\{\phi\geq t\}}{\mathbf{1}}\,\mathrm{w}\right)(x) = \underset{\{\phi(x)\geq t\}}{\mathbf{1}}\,\mathrm{w}(x) \tag{5.4}$$

until iteration κ when the target threshold q is reached *i.e.* $S_\kappa = q$. At this point, one can start the second phase of ANAIS: the estimation itself, as described in Subsection 5.1.2.

The threshold sequence allows a looser choice of the initial density: the algorithm never stops unexpectedly because empirical quantile level $\beta\in\,]0,1[$ implies so. Actually, there is always at least a point in $\{\phi\geq S_{k+1}\}$. As a result, one can now go for the natural choice: $f_Z^{\dagger,0} = f_X$. Besides, as only the most useful points are used to build the kernel density, the calculation burden is lightened w.r.t. NAIS.

Once the step κ is reached, the user has various choices as for how to spend the remaining budget if any. The two main follow.

1. Set $f_Z^{\circ} = f_Z^{\dagger,\kappa}$ and perform a regular NAIS algorithm in order to further better the auxiliary sampling density.

2. Sample $\aleph = N - \sum_{i=0}^{\kappa} n_i$ points according to $f_Z^{\dagger,\kappa}$ and use them to build estimators.

We chose number 2. It is easier to implement and the whole purpose of the iteration process seems already reached. Besides, we advice to use only these last points as not using the first batches of points might help free from a potentially poor choice of $f_Z^{\dagger,0}$.

5.1.2 Building the estimator

One can use $f_Z^{\dagger,\kappa}$-generated points as he or she sees fit, based on the objective. One can decide to normalise the estimators with $\aleph$ as well. As stated in the *in extenso* algorithm, we advice one to use $\aleph$ when estimating an expectation and $\sum_{j=1}^{\aleph}\mathrm{w}^{\dagger,\kappa}\left(Z_j^{\dagger,\kappa}\right)$ when looking for a quantile.

- Probability of exceedance

$$\mathbb{P}\left[\phi(X)\geq q\right] \approx \frac{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{\phi\geq q\}}\,\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}{\sum_{j=1}^{\aleph}\mathrm{w}^{\dagger,\kappa}\left(Z_j^{\dagger,\kappa}\right)} \tag{5.5}$$

- Conditional probability of exceedance $(s\geq q)$

$$\mathbb{P}\left[\phi(X)\geq s\,|\,\phi(X)\geq q\right] \approx \begin{cases} \frac{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{\phi\geq s\}}\,\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{\phi\geq q\}}\,\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)} & \text{if } \sum_{j=1}^{\aleph}\mathbf{1}_{\{\phi\geq q\}}\left(Z_j^{\dagger,\kappa}\right)\geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

[2] $\forall x=(x_1,\cdots,x_d)\in\mathbb{R}^d,\forall h\in\left(\mathbb{R}^+\right)^d, K_d(x,h)=\prod_{i=1}^{d}K\left(\frac{x_i}{h_i}\right),|h|=\prod_{i=1}^{d}h_i$.

- Conditional expectation – an extra NAIS step can help in this case–

$$\mathbb{E}\left[\vartheta\left(X\right)\middle|\,\phi\left(X\right)\geq q\right]\approx\begin{cases}\frac{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{\phi\geq q\}}\,\vartheta\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{\phi\geq q\}}\,\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)} & \text{if } \sum_{j=1}^{\aleph}\mathbf{1}_{\{\phi\geq q\}}\left(Z_j^{\dagger,\kappa}\right)\geq 1\\ 0 & \text{otherwise}\end{cases}\tag{5.7}$$

- $\phi\left(X\right)$ α-quantile

$$q\approx\min_{\mathbb{R}}\left\{t\,\middle|\,\alpha\leq\frac{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{t\geq\phi\}}\,\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}{\sum_{j=1}^{\aleph}\mathrm{w}^{\dagger,\kappa}\left(Z_j^{\dagger,\kappa}\right)}\right\}\tag{5.8}$$

- $\phi\left(X\right)$ α-quantile conditionally to $\{\phi\geq q\}$

$$q\approx\begin{cases}\min_{\mathbb{R}}\left\{t\,\middle|\,\alpha\leq\frac{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{t\geq\phi\geq q\}}\,\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}{\sum_{j=1}^{\aleph}\left(\mathbf{1}_{\{\phi\geq q\}}\,\mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}\right\} & \text{if } \sum_{j=1}^{\aleph}\mathbf{1}_{\{\phi\geq q\}}\left(Z_j^{\dagger,\kappa}\right)\geq 1\\ \max\left\{\phi\left(Z_i^{\dagger,\kappa}\right)\right\} & \text{otherwise}\end{cases}\tag{5.9}$$

Actually, once $f_Z^{\dagger,\kappa}$ is formed, it is smooth sailing as this does not amount to much more than CMC. The next station sums everything up as an algorithm.

5.1.3 ANAIS algorithm

Algorithm 5.1.1 is stated in the car one want to estimate the probability to exceed a high threshold or a high level quantile.

Algorithm 5.1.1 (ANAIS estimator)**.** *If the set of sampling interest with respect to random variable X is $\{\phi\geq t\}$, Adaptive Non Parametric Adaptive Importance Sampling (ANAIS) builds its auxiliary probability density function as follows.*

1. *Set $k=0$ and $f_Z^{\dagger,0}=f_X$.*
2. *Until $S_k=t$ proceed as described:*
   (a) *Generate the new batch of points according to the newest sampling density.*

$$\forall i\in\{1,\cdots,n_k\},\,Z_i^{\dagger,k}\overset{\text{iid}}{\sim}f_Z^{\dagger,k}\tag{5.10}$$

   (b) *Apply ϕ to each point and define S_{k+1} as the minimum of t and the image set empirical β-quantile.*

$$C_i^{\dagger,k}=\phi\left(Z_i^{\dagger,k}\right)\qquad\qquad S_{k+1}=\min\left\{C_{(\lfloor\beta n\rfloor+1)}^{\dagger,k},t\right\}\tag{5.11}$$

   (c) *Build the new kernel based sampling density.*

$$f_Z^{\dagger,k+1}\left(z\right)=\frac{\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(\mathbf{1}_{\{\phi\geq S_{k+1}\}}\,\mathrm{w}^{\dagger,i}\right)\left(Z_j^{\dagger,i}\right)K_d\left(Z_j^{\dagger,i}-z,h^{\dagger,k+1}\right)}{|h^{\dagger,k+1}|\sum_{i=0}^{k}\sum_{j=1}^{n_i}\left(\mathbf{1}_{\{\phi\geq S_{k+1}\}}\,\mathrm{w}^{\dagger,i}\right)\left(Z_j^{\dagger,i}\right)}\qquad\mathrm{w}^{\dagger,k+1}=\frac{f_X}{f_Z^{\dagger,k+1}}\tag{5.12}$$

(d) Set $k = k + 1$.

3. *Define* $\kappa = k$.

4. *Use up all remaining simulation budget* $\aleph$ *sampling* iid *points according to* $f_Z^{\dagger,\kappa}$.

5. *Build the desired expectation estimator, picking from Subsection 5.1.2 list for instance.*

In particular, $V = \mathbf{1}_{\{\phi(X)\geq t\}}$ *expectation estimator* i.e. *the ANAIS probability of exceedence estimator will be denoted* $\widehat{V_\dagger}$.

$$\widehat{V_\dagger} = \frac{\sum_{j=1}^{\aleph} \left(\mathbf{1}_{\{\phi\geq t\}} \mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}{\aleph} \tag{5.13}$$

The α*-quantile of* ϕ *ANAIS estimator* $\widehat{q_\dagger}$ *is constructed using the same algorithm but setting* $\widehat{q_\dagger}_k = \infty$ *and changing item 2 into*

$$\textit{Until } S_k \geq \widehat{q_\dagger}_k = \inf\left\{t \in \mathbb{R} \middle| \alpha \leq 1 - \frac{\sum_{j=1}^{\aleph} \left(\mathbf{1}_{\{t\leq\phi\}} \mathrm{w}^{\dagger,\kappa}\right)\left(Z_j^{\dagger,\kappa}\right)}{\sum_{j=1}^{\aleph} \mathrm{w}^{\dagger,\kappa}\left(Z_j^{\dagger,\kappa}\right)}\right\}$$

If simulation budget is depleted before step κ *or very little then remains, one can use all the simulated points to make any other IS estimator.*

Algorithm 5.1.1 can be easily adapted to the cases when a low threshold or a low level is wanted. One then has just to flip the inequalities in the indicator functions and use $\widehat{F_{3X}}$ from Definition 2.3.1 to estimate the quantile. However, this algorithm does deserve some theoretical deepening.

- How to set the intermediary quantile level(s)?
- How to choose the density kernel?
- How to set the density kernel bandwidth?
- What about the convergence properties?

We made do without nonetheless and contented ourselves with the already presented general importance sampling theory.

5.2 ANAIS goes through the crash tests

We proposed the new ANAIS algorithm to cope with NAIS shortcomings but without theoretical study. To demonstrate its efficiency we put through the crash tests from Chapter 4. As for Algorithm 5.1.1 settings, `KDE` was used in the very same fashion as when testing NAIS in Subsection 4.1.2. Besides we made one thousand point batches and the empirical quantile level was set to 0.90.

$$f_Z^{\dagger,0} = f_X = \mathcal{N}\left(0_2, 1_2\right) \qquad \forall i \in \{0, \cdots, \kappa\}, n_i = 10^3 \qquad \beta_\dagger = 0.90 \tag{5.14}$$

5.2.1 Experimental results with Rayleigh law in dimension two

Desired probability and quantile were estimated at the same time. The whole estimation budget was always exactly depleted at each estimation and $(\kappa, \aleph)$ was always $(4, 36 \cdot 10^3)$. All estimates can be seen in Figure 5.1.

$$\overline{\widehat{T_\dagger}}_m = 2.34 \qquad \overline{\widehat{V_\dagger}}_m = 1.00 \cdot 10^{-5} \qquad \overline{\widehat{q_\dagger}}_m = 4.88 \tag{5.15}$$

$$\rho\left(\widehat{T_\dagger}, m\right) = 0.26 \qquad \rho\left(\widehat{V_\dagger}, m\right) = 0.02 \qquad \rho\left(\widehat{q_\dagger}, m\right) = 0.04 \tag{5.16}$$

Calculation were fast and results are consistent and probability estimates are biasless as expected. The empirical deviation is small. This can be seen on Figure 5.1.

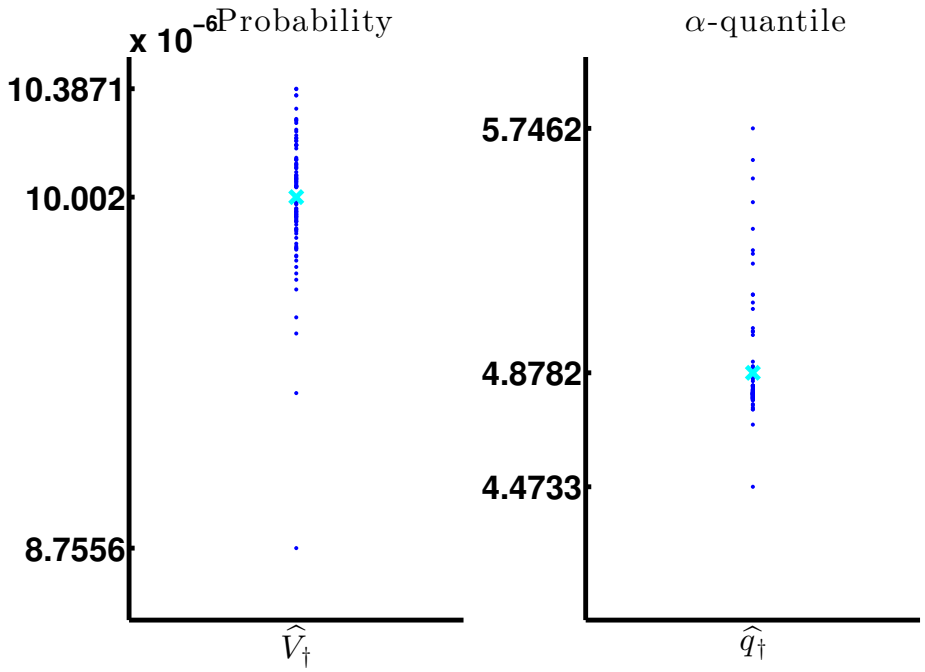


Figure 5.1: ANAIS probability and quantile estimations as *per* Section 4.1. ANAIS is tested against $\mathcal{R}_2$. Estimated values are represented with blue dots • and the mean estimate with a cyan cross ×.

5.2.2 Experimental results with Rayleigh law in dimension twenty

Results can be seen on Figure 5.2.

$$\overline{\widehat{T_\dagger}}_m = 18.04 \qquad \overline{\widehat{V_\dagger}}_m = 5.49 \cdot 10^{-6} \qquad \overline{\widehat{q_\dagger}}_m = 7.99 \tag{5.17}$$

$$\rho\left(\widehat{T_\dagger}, m\right) = 0.11 \qquad \rho\left(\widehat{V_\dagger}, m\right) = 2.38 \qquad \rho\left(\widehat{q_\dagger}, m\right) = 0.05 \tag{5.18}$$

Contrary to during the first test, Figure 5.1, occasional but massive probability overestimations were observed. Besides, the probability was often underestimated. This resulted in a tangible

increased variance w.r.t. test 1, when the input dimension was lesser. As for the quantile estimations, they maintained their low relative variance and bias. However, the simulation time change was important: ten times as long as before.

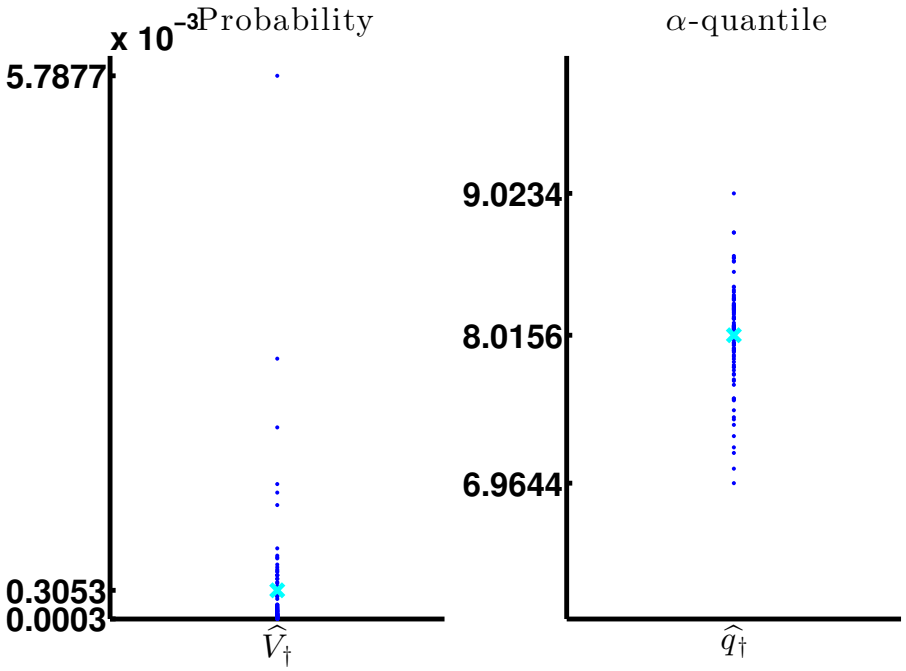


Figure 5.2: ANAIS probability and quantile estimations as *per* Section 4.2. ANAIS is tested against $\mathcal{R}_{20}$ to estimate and quantile and a probability. Estimated values are represented with blue dots $\cdot$ and the mean estimate with a cyan cross $\times$.

5.3 Conclusion

Comparing Section 5.2 results with Subsubsection 4.1.2.2 and Subsection 4.2.2, it can be seen that in both cases, ANAIS delivered results as accurate as NAIS. Besides ANAIS is way more user-friendly than NAIS because we did not have to look for an initial auxiliary distribution and could use f_X directly. Actually, though they both require the user to define a bandwidth, ANAIS does not require prior and its performance does not deteriorate as much as that of NAIS in high dimensions. This is why it was used in NAIS stead from then on. As a practical consequence, only CE, ANAIS and AST robustnesses to input dimension and rarity increase were tested in Chapter 6.

Chapter 6

Robustness to dimensionality and rarity increase

In practice, the systems of interest for ONERA have different numbers of variables. How do the CE, AST and ANAIS performances evolve when dimensionality increases ? The probabilities of failure and safety requirements ONERA faces increase faster than the available simulation budget. How rare an event and how extreme a quantile can CE, ANAIS and AST estimate with a given budget? To answer these questions, we decided to test the algorithms robustnesses with respect to dimensionality in Section 6.1 and to rarity in Section 6.2.

6.1 Robustness to dimensionality

According to Chapter 4 and Section 5.2 results, the input space dimension affects the algorithms performances. To have a clearer insight about this influence, we used the following experiment.

6.1.1 Reference values

We stuck to the Rayleigh law and observed how the estimators performed to the dimension d increase

$$d \in \{1, \cdots, 20\} \qquad \alpha = 1 - 10^{-5} \qquad X \sim \mathcal{N}(0_d, 1_d) \qquad Y = \|X\| \tag{6.1}$$

$$Y \sim \mathcal{R}_d \qquad \mathcal{R}_d \ \alpha\text{-quantile?} \qquad V = \underset{\{\phi(X) \geq q_\alpha(d)\}}{\mathbf{1}} \qquad \mathbb{E}[V] = ? \tag{6.2}$$

while maintaining the same simulation constraints

$$N = 50 \cdot 10^3 \text{ simulations } \textit{per} \text{ estimation}, \qquad m = 100 \text{ estimations}. \tag{6.3}$$

Actually, the Rayleigh law seemed a good choice for two reasons. First, it is simple enough to be sure that a rising issue would originate from the algorithm and not the transfer function. Second, it is nonetheless a scalar transformation of a multi-dimensional Gaussian as most industrial cases.

To avoid the cumbersomeness of calculating and inverting the cumulative distribution function of $\mathcal{R}_d$ for $d \in \{1, \cdots, 20\}$, we made a huge CMC simulation.

$$N^* = 10^7 \quad m^* = 100 \quad \forall i \in \{1, \cdots, N^*\}, \forall d \in \{1, \cdots, 20\} \quad Y_i \overset{\text{iid}}{\sim} \mathcal{R}_d \quad q_{ref}(d) = \overline{Y_{(\lfloor \alpha N \rfloor)}}_m \tag{6.4}$$

Results can be seen on Figure 6.1.

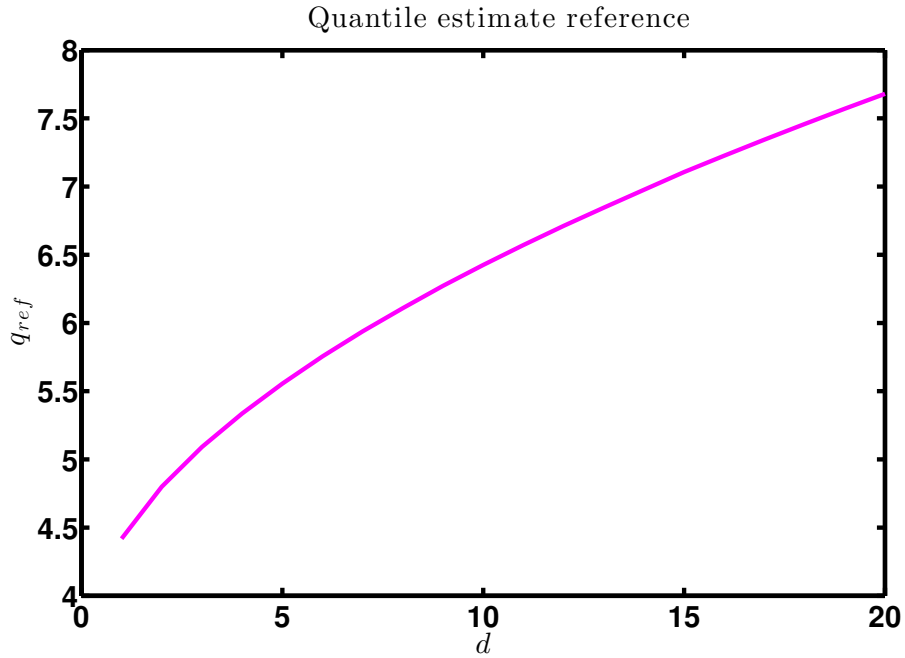


Figure 6.1: Reference quantiles used during the robustness w.r.t. dimension test.

6.1.2 Experimental results

The computer experiments led to the following results. The parameter settings are those of Subsubsection 4.1.1.1 for CE, Equation 5.14 for NAIS and Equation 4.14 for AST.

Estimation cost

CE and ANAIS never stray from the assigned N simulation budget. AST does not either when estimating a quantile but has a random probability estimation cost that can induce a small excess cost or remainder. However, as Figure 6.2 shows, this variance did not seem to depend on the dimension.

Probability estimation

Subfigure 6.3a shows that all three techniques almost always yielded a bias of $\pm 5\%$. CE seemed unaffected by the density and had little bias. AST seemed unaffected by the dimension but had a larger bias. As ANAIS, its bias seemed to increase with the dimension.

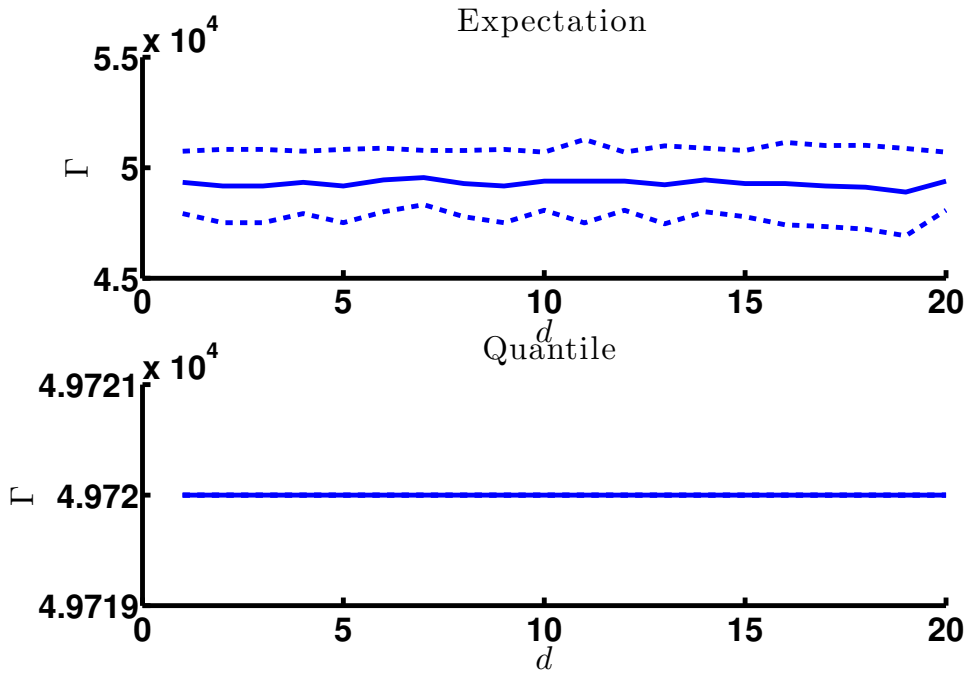


Figure 6.2: AST random cost evolution as the input dimension increases. Mean cost is represented with a straight line and the standard deviation with a dotted line.

It can be seen on Subfigure 6.3b that AST erd remained flat, though around 30%. ANAIS erd started closer to zero but grew steadily and outgrew AST's from $d = 14$ on and peaked to 2 for $d = 20$. This was an appearance of the well known curse of dimensionality cast upon kernel density based estimators. As for CE's empirical relative deviation, it showed a mild slope from almost 0 to 25% and always was the smallest.

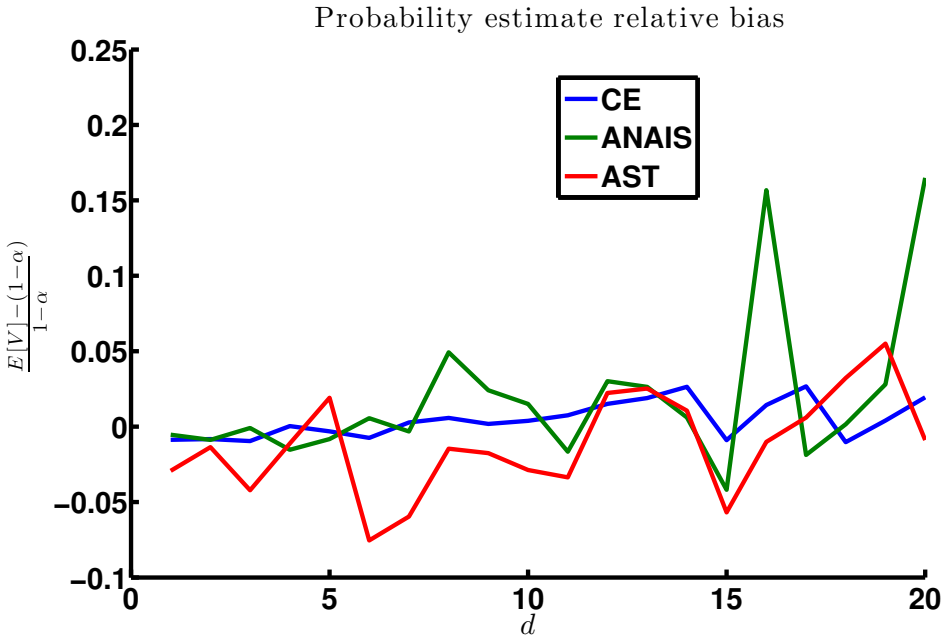
For all tested dimensions, CE performed very well due to the wealth of available information but cannot be expected to do the same in real case. AST never bested the estimations but looked more likely to perform the same way in real case. ANAIS proved reliable if the dimension is no more than 10, to use with caution for $d \in \{11, \cdots, 14\}$ and to be avoided for d greater or equal to 15.

Quantile estimation

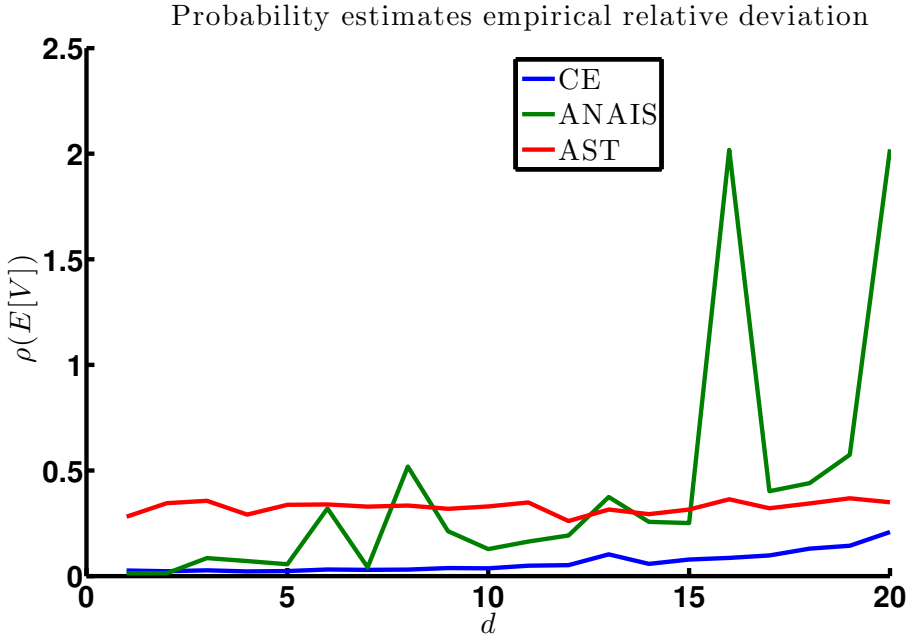
For all three estimators, the relative bias *i.e.* $\frac{q_{est} - q_{ref}}{q_{ref}}$, was always of 10^{-3} magnitude and could therefore be deemed accurate. CE and AST biases seemed insensitive to dimension whereas that of ANAIS decreased from $d = 1$ to $d = 10$ and from then on increases. CE's relative bias was almost zero for all values of d. ANAIS estimator bias was the biggest and it peaked for $d = 20$. As for AST estimator bias, it remained around -10^{-3}. This can be seen on Subfigure 6.4a.

AST erd decreased linearly from $1,7\%$ to 1% as d goes from 1 to 20. In the same time, CE's erd increased from 0,1% to 0,5%. As for ANAIS erd, it showed a right-skewed bowl shape: 5,5% for $d = 1$, 0,7% for $d = 7$ and 3% for $d = 20$. This can be seen on Subfigure 6.4b.

Though CE performed better than AST which itself performed better than CE, all three

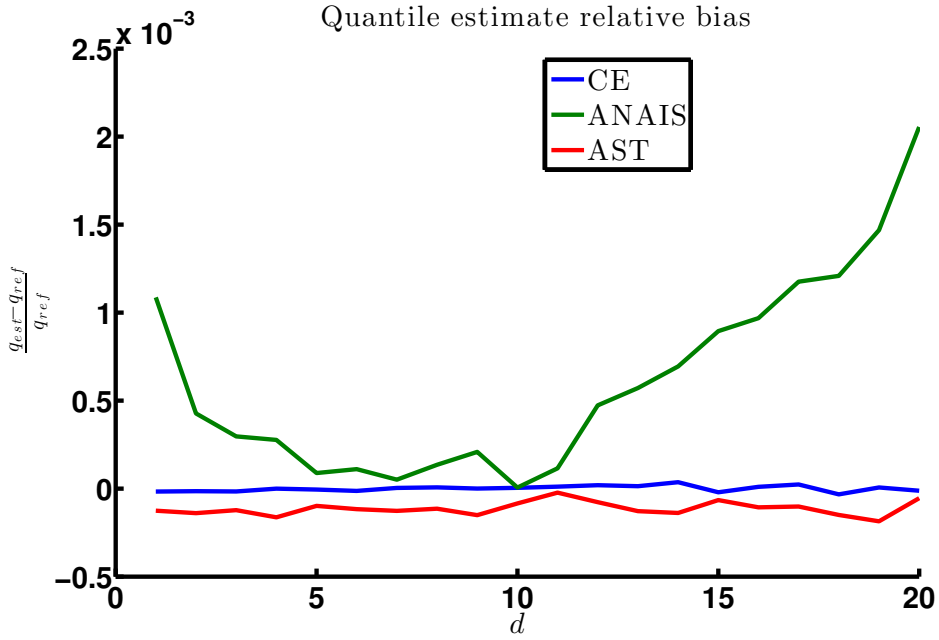


(a) Probability estimates relative bias.

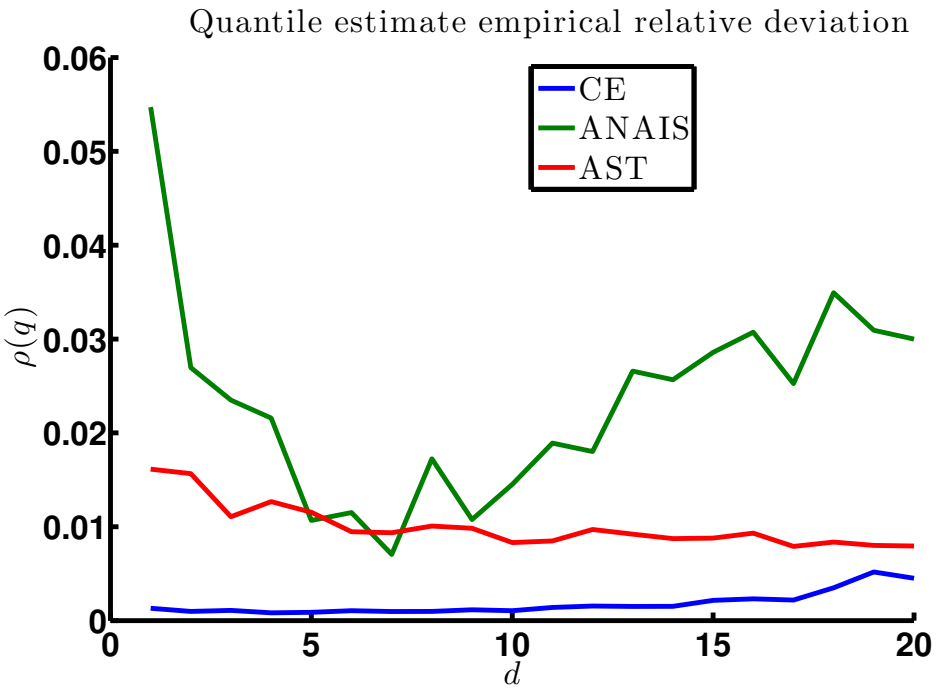


(b) Probability estimates empirical relative deviation.

Figure 6.3: CE, ANAIS and AST estimate the $\mathbb{P}\left[\mathcal{R}_d \geq q_{ref}\left(d\right)\right]$ for $d \in \{1, \cdots, 20\}$ and $q_{ref}\left(d\right)$ given by Section 6.1. The relative bias probability reference is 10^{-5}.



(a) Quantile estimates relative bias.



(b) Quantile estimates empirical relative deviation.

Figure 6.4: CE, ANAIS and AST estimate the $(1-10^{-5})$-quantile of a $\mathcal{R}_d$. The bias references are the $q_{ref}(d)$ thresholds for $d \in \{1, \cdots, 20\}$ derived according to Section 6.1 and given Section 6.1.

techniques delivered satisfactory quantile estimations. Therefore, quantile estimation should not be of major importance when choosing which estimator to use.

Estimation duration

One can see on Subfigure 6.5a that ANAIS takes way more time than the other methods. This may be due to the fact we used the `KDE` one size fits all `MATLAB toolbox` instead of a hand made *sur mesure* piece of code. Actually, evaluating $f_Z^\dagger$ is time consuming. As for CE and AST, it can be seen on Subfigure 6.5b that estimating a quantile and estimating a probability *via* AST are equally faster than performing CE, whatever the dimension. CE's simulation time seems to increase linearly with dimension.

No noticeable trend could be seen on Figure 6.6 as all empirical relative deviation swang between 0.05 et 0.3. Only ANAIS peak to 0.8 for $d = 10$ stood out. One hypothesis was that due to the calculation size increase `MATLAB` had changed its calculation strategy.

In most real case, the estimator self calculation time will be negligible with respect to the black-box simulation cost. But if time matters, AST is the best choice as ANAIS requires more calculation than the two other techniques and there will most likely be no time to devise a good CE auxiliary density.

6.1.3 Outcome: robustness to input dimension

CE seemed insensitive to the input dimension and, had very little bias and relative variances and time cost but actually benefited from a lot of *a priori* information that will not be available in a real case. AST seemed insensitive to the input dimension and had vey little bias and time cost as well but showed relative variances around a third. However, AST did without needing any prior that would unlikely be available in a real case. ANAIS is sensitive to dimensionality as could be expected from a density kernel based technique. ANAIS performed better than AST when dimension was no greater than 15 and showed a dramatic variance increase afterwards. Though it took longer, ANAIS needed neither prior nor deft tuning.

To balance these robustness to dimensionality results with robustness to rarity results, we made the Section 6.2 experiment.

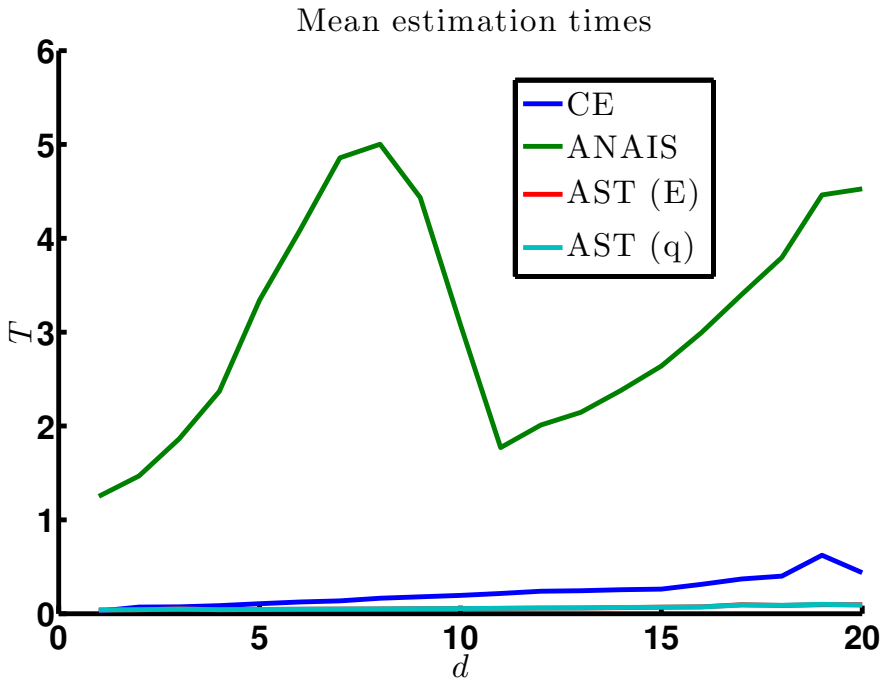
6.2 Robustness to rarity

In the following experiment, we try the estimation range of CE, AST and ANAIS *i.e.* how minute a probability and extreme a quantile they can estimate accurately with a given budget. This is a very important feature because this is a of major importance when choosing an estimator for a given task.

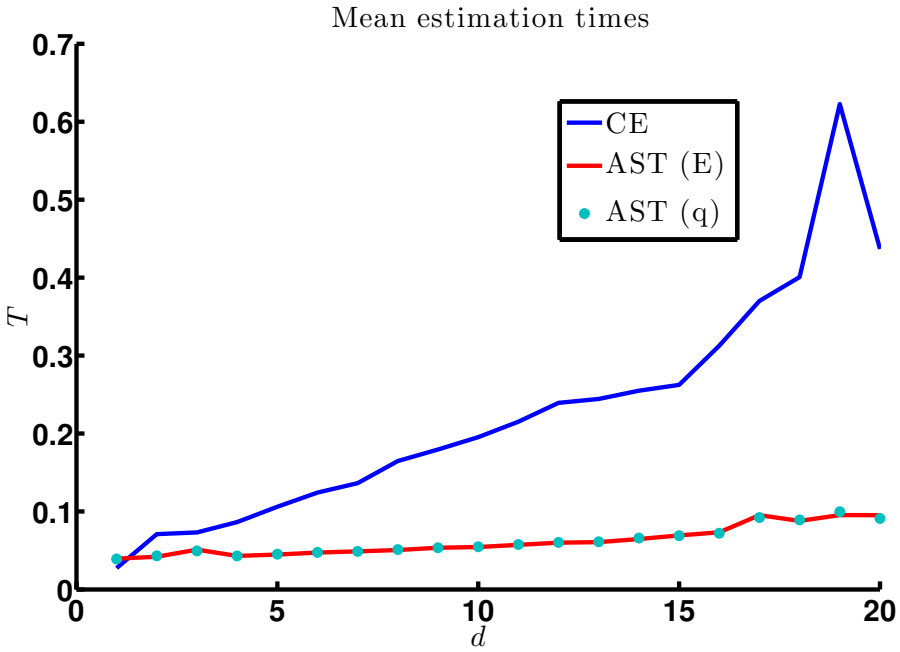
6.2.1 Settings and reference values

We came back to a $\mathcal{R}_2$ random variable generated *via* a bi-dimensional Gaussian.

$$X \sim \mathcal{N}(0_2, 1_2) \qquad Y = \|X\| \qquad Y \sim \mathcal{R}_2 \tag{6.5}$$



(a) ANAIS takes way more time than other methods.



(b) AST is faster than CE, be it for quantile or probability estimation, and seems less dimension sensitive.(Zoom from above plot, after removing ANAIS data.)

Figure 6.5: CE, ANAIS and AST estimation times as the input dimension increases.

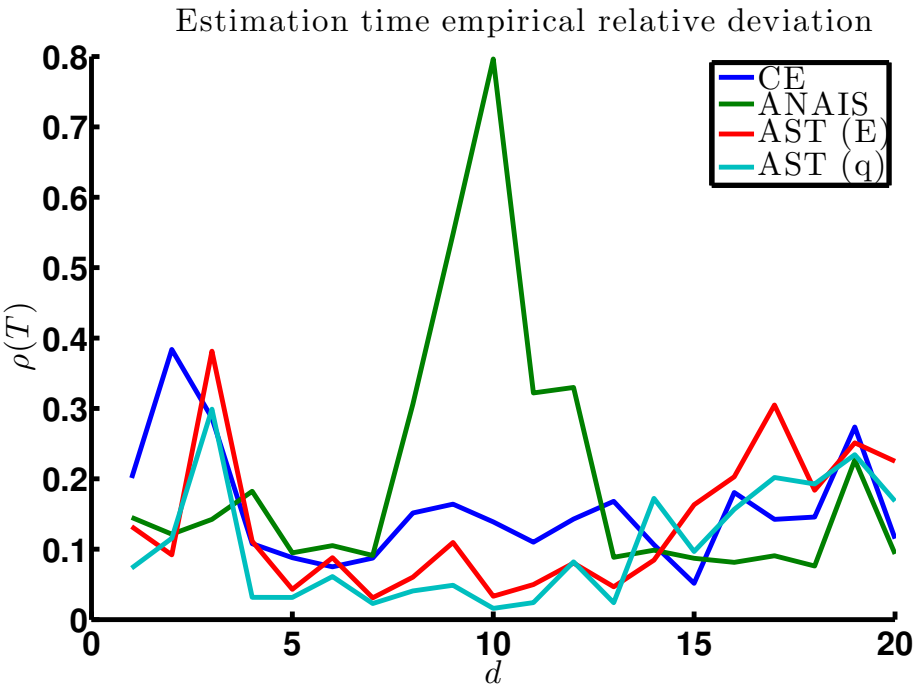


Figure 6.6: CE, ANAIS and AST estimation time empirical relative deviation as the input dimension increases.

This way, dimensionality was not an issue and all quantiles were theoretically known for any given level. We chose 17 levels equally spaced from $1-10^{-5}$ to $1-10^{-13}$ as shown on Figure 6.7.

$$\forall \alpha \in \left]0,1\right[, \qquad q(\alpha) = \sqrt{-2\ln(1-\alpha)} \tag{6.6}$$

$$\forall l \in \{1,\cdots,17\}, \qquad \alpha_l = 1 - 10^{-\frac{l+9}{2}} \tag{6.7}$$

With all three techniques, we estimated the probability to exceed the quantiles and the quantiles themselves.

$$\forall l \in \{1,\cdots,17\}, \qquad \mathcal{R}_d \ \alpha_l\text{-quantile?} \qquad V_l = \underset{\left\{\phi(X)\geq q_{\alpha_l}(d)\right\}}{\mathbf{1}}, \qquad \mathbb{E}[V_l] = ? \tag{6.8}$$

The constraint was to do it with a given simulation budget.

$$N = 50\cdot 10^3 \text{ simulations } \textit{per} \text{ estimation}, \qquad m = 100 \text{ estimations}. \tag{6.9}$$

CE and ANAIS were tuned as for in Section 4.2. Tuning AST proved harder.

6.2.2 Experimental results

The computer experiments led to the following results. The parameter settings are the same as those of Subsection 6.1.2 dimension robustness test, except for AST.

Three parameters define AST's random cost $\Gamma = n \times (1 + \kappa \times \omega)$ as explained in Equation 3.43. n and ω are given by the user but κ is random and only controlled *via* β, the

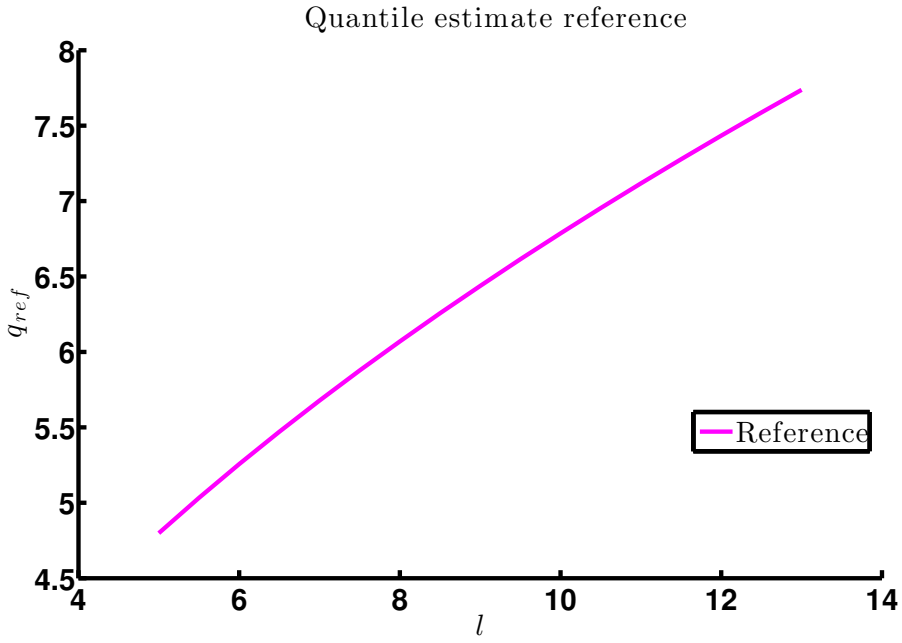


Figure 6.7: Reference quantiles used during the robustness w.r.t. dimension test.

intermediate quantile-threshold level which is user defined. Besides, it depends essentially on the target probability or target quantile's level. We used Proposition 3.3.2 and κ deterministic limit given therein to try to respect the budget constraint while adapting to the different α_l.

$$N = 5 \cdot 10^5 \qquad n = 200 \qquad \beta = 0.29 \qquad \kappa_l = \lfloor \frac{\ln(\alpha_l)}{\ln(\beta)} \rfloor \qquad \omega_l = \lfloor \frac{N}{\kappa_l n} \rfloor \tag{6.10}$$

Essentially, this meant applying the Markovian transition kernel less times between each steps of the algorithms. We had mitigated budget and expected variance reduction in favor of budget respect.

Estimation cost

Despite the above described parameter settings, AST exceeded its N budget as represented on Figure 6.8. We accepted Γ essential randomness and expected it to at least outperform ANAIS and CE which did respect the budget limitation.

Probability estimation

Figure 6.9a shows that CE and ANAIS had more stable and smaller relative bias around $\pm 2\%$. AST bias swang around zero by $\pm 10\%$.

It can be seen on Subfigure 6.9b that AST empirical deviation increased from 30% to 55%. Meanwhile, ANAIS and CE empirical relative deviations grew respectively from 2% to 17% and from 2.5% to 4%.

If CE or ANAIS is well-tuned, it as very long estimation range. With our parameter settings, when rarity increases, AST budget is expected to increase as well so as not to increase the variance: its estimation range is smaller.

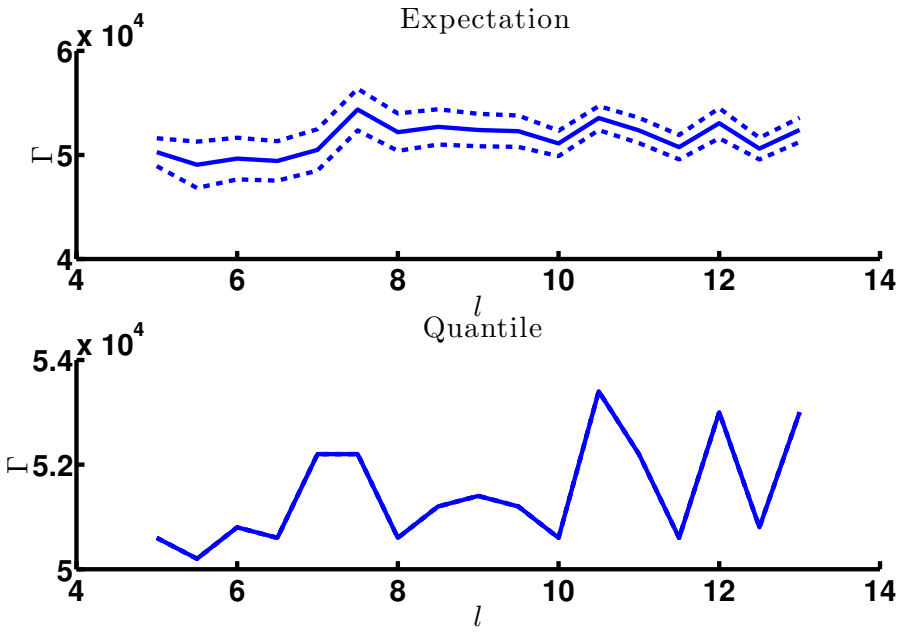


Figure 6.8: AST random cost evolution as the rarity increases. Mean cost is represented with a straight line and the standard deviation with a dotted line.

Quantile estimation

Figure 6.10a shows that ANAIS relative bias decreased as rarity increased from $+1,2\%$ to $-0,1\%$. Meanwhile AST relative bias swung between 0 and -0.6%. As for CE, its relative bias never was more 0.2% away from zero.

It can be seen on Subfigure 6.10b that AST empirical deviation decreased from 1.75% to 1%. Meanwhile, ANAIS empirical relative deviation decreased as well from 3% to 0.6%. CE's wavered around 0.1%.

All three techniques can estimate extreme quantile accurately and consistently. Quantile estimation alone can not motivate one to choose a method rather than an other.

Estimation duration

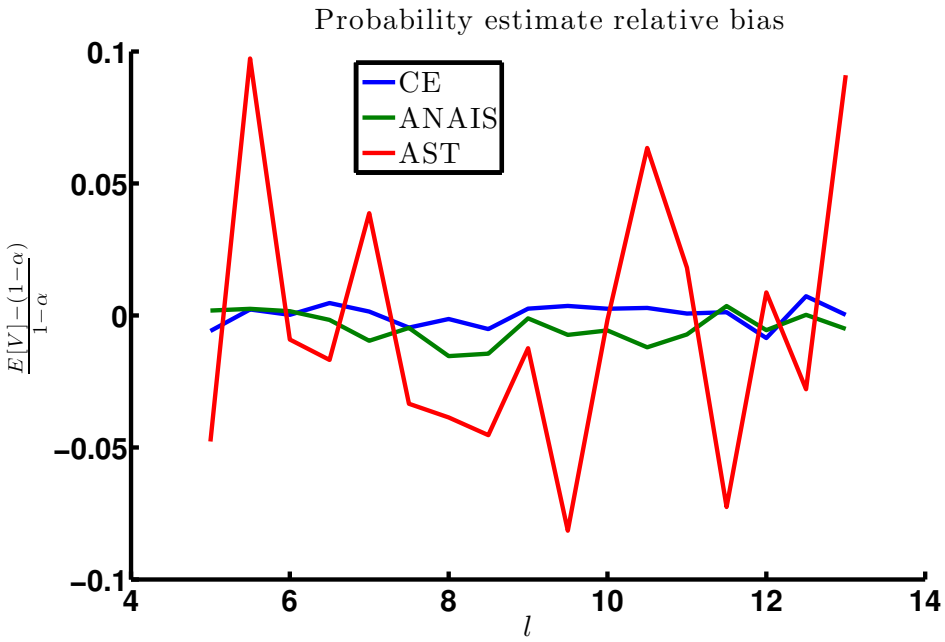
The same comments as when estimating probabilities in Section 6.1.2 applied as Figure 6.11 and Figure 6.12 show.

6.2.3 Outcome: robustness to rarity

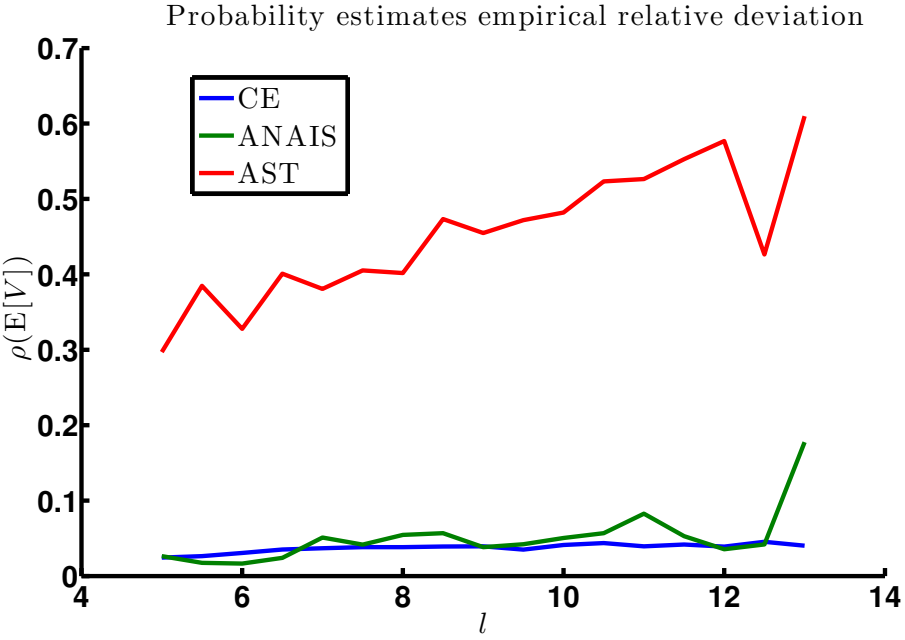
Comments were similar to those of Subsection 6.1.3.

1. CE shows almost no bias, has little variance and calculation time cost thanks to an unrealistic abundance of information.

2. ANAIS estimates both extreme quantiles and rare event probabilities with little bias and variance. Maybe the way the `KDE toolbox` is coded can account for ANAIS taking ten

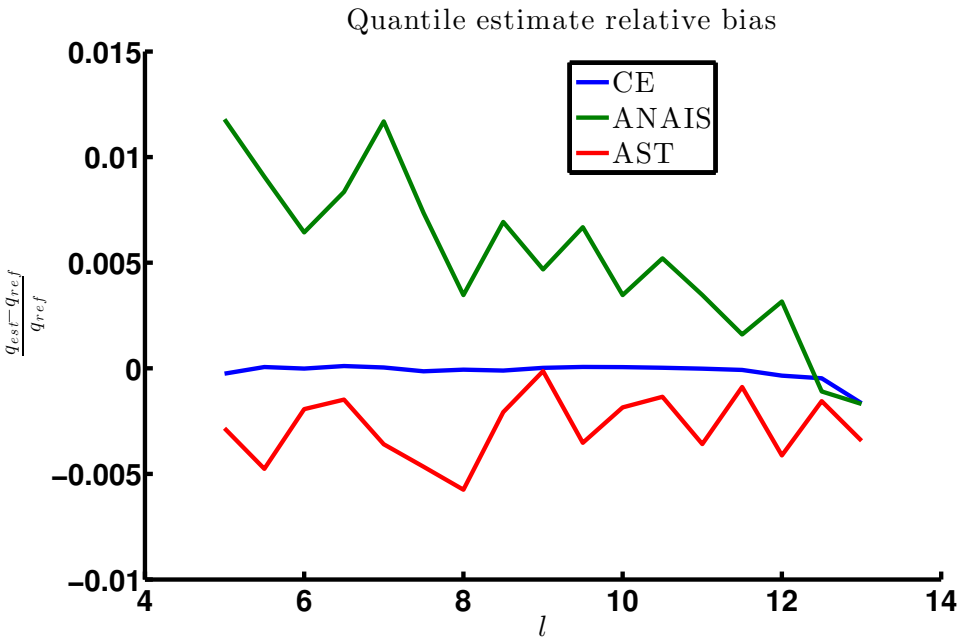


(a) Probability estimates relative bias.

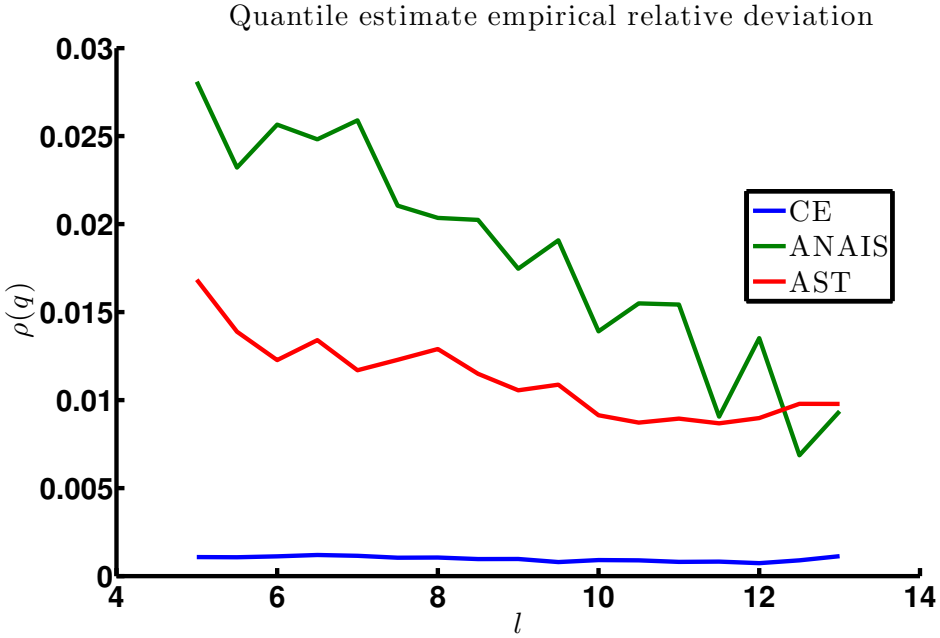


(b) Probability estimates empirical relative deviation.

Figure 6.9: CE, ANAIS and AST estimate $\mathbb{E}[V_l]$ as *per* Section 6.2. The relative bias probability reference is $\alpha_l = 1 - 10^{-\frac{l+9}{2}}$.

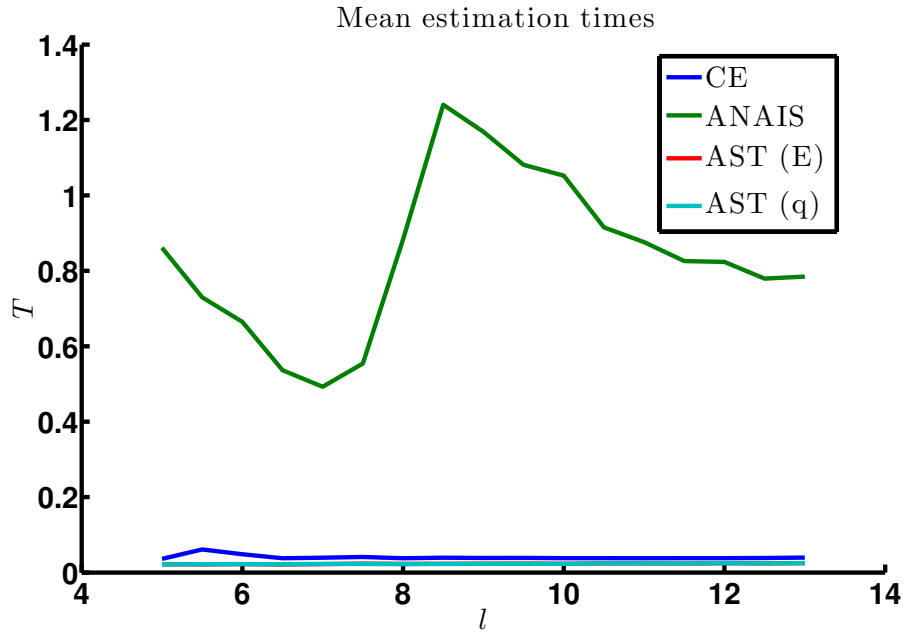


(a) Quantile estimates relative bias.

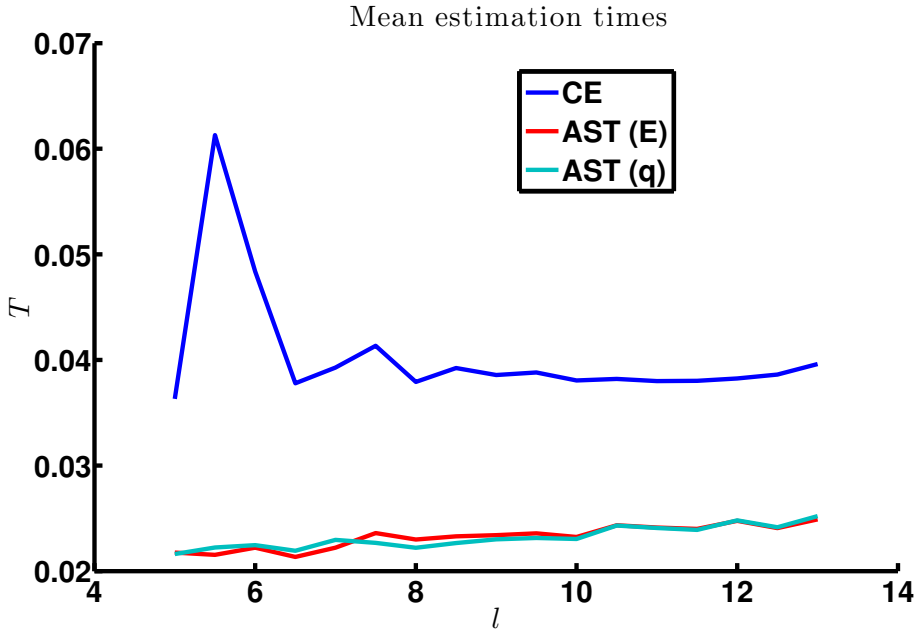


(b) Quantile estimates empirical relative deviation.

Figure 6.10: CE, ANAIS and AST estimate α_l as *per* Section 6.2. The relative bias probability reference is $\alpha_l = 1 - 10^{-\frac{l+9}{2}}$.



(a) ANAIS takes way more time than other methods.



(b) AST is faster than CE, be it for quantile or probability estimation, and seems less time sensitive. (Zoom of above plot, after removing ANAIS data.)

Figure 6.11: CE, ANAIS and AST estimation times as the rarity increases. l corresponds to a probability $\alpha_l = 1 - 10^{-\frac{l+9}{2}}$.

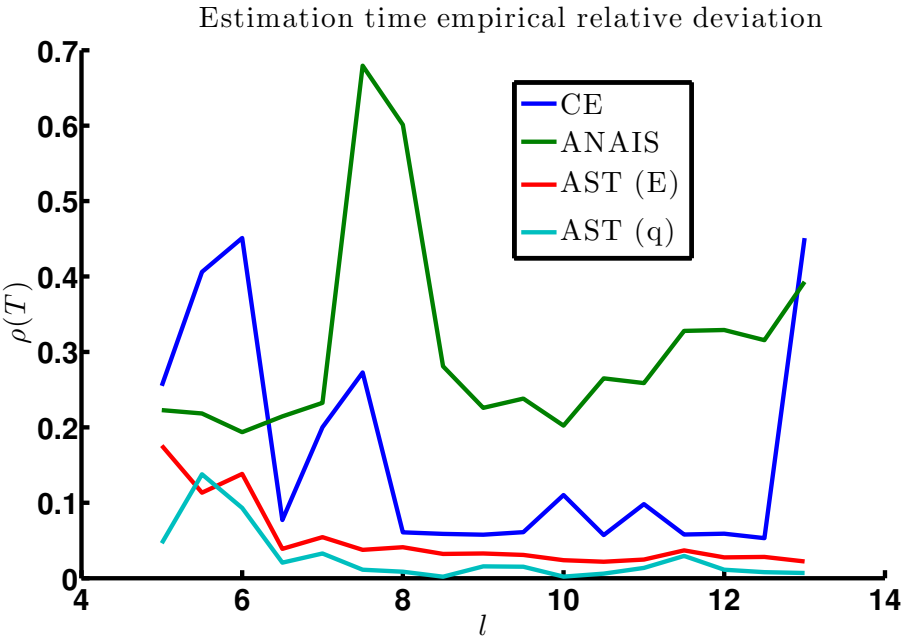


Figure 6.12: CE, ANAIS and AST estimation time empirical relative deviation as the rarity increases. l corresponds to a probability $\alpha_l = 1 - 10^{-\frac{l+9}{2}}$.

times longer than the other methods, but one should remember the involved densities are essentially complex.

3. AST is fast, shows little bias, has a little quantile estimation variance but high and increasing probability estimation variance, even though it consumed a greater budget.

However the techniques could seemingly be ranked in this more contrasted fashion.

6.3 Conclusion

We fixed a simulation budget and tried ANAIS, CE and AST capacity to accurately estimate Rayleigh distribution probabilities and quantiles first when the input dimension increased and second when the event rarity increased. According to our experiments, CE seemed the best choice because it always delivered the most accurate estimations. Then comes ANAIS which was victim of the dimensionality curse but was otherwise accurate. Last comes AST. Even though AST appeared as insensitive to dimension, its probability estimation grew a lot with rarity and never was less then a third.

In retrospect, one explanation may be that the CE was more simple and suited better the circumstances . Besides, this study has two main limitations.

Tunings were not optimal. Indeed, optimal tunings are unknown but tunings were not adapted at all even though the setting changed a lot. In the absence of tuning rational we used crude rules. Maybe the performances would have been different had the algorithms been better tuned.

The Euclidean norm used to transform the Gaussian distributions into Rayleigh distribution

does not yield the many possible feature a real life transfer function can. Though it is not universal, it was nevertheless a reasonable way to exhibit the algorithms behaviours.

Theory only can put a final end to such experiments. But in the absence of such results to enlighten this assessment, ONERA gave us an opportunity to practice in a case very similar to the ones of its interest: the estimation of the probability of collision of two satellites as presented in Chapter 7.

Chapter 7

The Iridium Cosmos case

We decided to estimate the probability of collision between satellites Iridium and Cosmos for two reasons. First, we need a rare event probability estimation practice more realistic than an Rayleigh distribution in order to better discover CE, ANAIS and AST. The second objective was estimating the probabilities of such event accurately. Actually, they are of major importance in the satellite safety protocols the ONERA develops.

7.1 The Galactic void is not void anymore

Over the last century, mankind has extended its realm to space. Though it is limited to the immediate vicinity of the Earth, this newly explored domain already bears the marks of human activities: space junks. Those were ignored until the last decade when the orbits of practical human interest became conspicuously polluted. Space debris *are* a threat to human space activity [37, 46].

7.1.1 The threat of the increasing space pollution

Dealing with space debris rises a number of questions. What to do with active satellites turning inactive? How to clean at least the frequently used orbits? How to design future satellite so that they pollute as less as possible when they go out of order? How to ensure the safety of active satellites [26]? How to measure the risk they are exposed to? We focused on this last question[1].

The safest practice with respect to satellites is avoiding collision [87]. Avoidance maneuvers are efficient but costly and reduce the possible usage time. However, there is no point in saving fuel if it ends up spread in the void after a collision. Satellite safety responsible teams have to design a trade-off between fuel saving and collision avoiding. Avoidance maneuvers are decided according to monitoring process based, among other parameters, on the estimated collision probability.

[1] This chapter is somewhat between a follow–up and a correction of [68]

7.1.2 Iridium 33 & Cosmos 2251 unexpected rendez-vous

The February 10^{th} 2009, though the probability of their collision was reported insignificant [47], active commercial satellite Iridium and out of order Russian satellite Cosmos did collid. The impact produced number of smaller debris. Most of them can destroy any artificial space object, whether in use or not, they might encounter.

The Iridium satellite program is worth at least 200 M$, so one could at first think the loss was worth that amount. However, the smaller debris are the real international concern as well. Actually, they can hinder future space mission and even cause more dangerous debris themselves. Besides, they are harder to spot and numerous. To avoid the generation of such secondary debris, preventing the decomposition of bigger debris is a priority.

To prevent such ordeal to occur again, the whole satellite monitoring process is under review. This includes in particular the estimation of the probability of collision between a given satellite and a debris.

7.1.3 Further than hypothesis based numerical integration

Crude Monte Carlo (CMC) would be a way if it could cope with very small probabilities, say 10^{-6}, within the available simulation budget and time as stated in Subsection 1.2.5. Actually, in a real context, the simulation budget is limited and even scarce when compared to rare event probabilities. According to our knowledge, the current methodology in the NASA [21] to estimate the collision probability between two orbiting objects is integrating a Rician probability density function (pdf) over a circular sub domain of the collision plan.

Actually, the Gaussian uncertainty with respect to the real location and speed of the satellites when last observed *via* Earth based radars is deemed to remain Gaussian as the satellites go on their tracks, although the dynamics is nothing but linear. Said tracks are assumed, under specified hypothesis, to be straight lines in the encounter region so as to conveniently define a collision plan. The error ellipsoids are combined then projected on the collision plan and eventually numerically integrated.

However easily implemented and fast to calculate this method is, the two hypothesis it is based on are major drawbacks.

1. Uncertainty linear propagation.

2. Straight line tracks in the vicinity of the time of closest approach.

The previously introduced rare event techniques were expected to help avoid such hypothesis when estimating the probability. So, CE, ANAIS and AST were used in this car both as a way to test them on a real life case and to find an appropriate tool with respect to this specific issue.

7.2 Spacecraft encounter: a random minimal distance problem

We first formalized the question “what was the probability a commercial Iridium communications satellite and a defunct Russian satellite Cosmos collided?”.

In order to introduce the predicament an active spacecraft managing team can find itself in, we described the geometrical issue at hand and then the main source of randomness. To keep things simple, orbital mechanics being not our topic, we used Kepler mechanics. One could use more advanced models such as SGP4 [63] if wanted. The discussed probability estimation methodologies are independent of the method.

7.2.1 A deterministic geometry framework

We considered two satellites orbiting around the Earth in a Galilean frame of reference with our planet as origin and equipped with the Euclidean distance. This three-body problem was considered a double two-body problem: each satellite interacted through gravity only with the Earth and not with the other satellite. Besides, the Earth and the satellites were assumed to be homogeneous spheres with radii d_E, d_1 and d_2. The collision distance was therefore $d_c = d_1 + d_2$.

At time t, the satellites were represented by their states $\vec{s}_1(t)$ and $\vec{s}_2(t)$ i.e. their positions $\vec{r}_1(t)$ and $\vec{r}_2(t)$ and their speeds $\vec{v}_1(t)$ and $\vec{v}_2(t)$ such that $\vec{s}_i = (\vec{r}_i, \vec{v}_i)$.

In our setting, the speeds evolved according to the same well-known Ordinary Differential Equations defining the two body problem

$$\forall t, \frac{d\vec{v}_i}{dt} = -a\frac{\vec{r}_i}{r_i^3} \tag{7.1}$$

where a is a positive constant given by physics.

This ordinary differential equation (*ode*) is analytically solved in many textbooks [31, 72]. Its solution depends continuously and in a bijective fashion on the initial conditions *i.e.* its value $\vec{s}_i^m$ at t_i^m, the measurement time, through ϕ the ode's resolvant i.e. its solution map.

$$i \in \{1, 2\}, \forall t \in I, \vec{s}_i(t) = \phi(\vec{s}_i^m, t_i^m, t) \tag{7.2}$$

At this point, there was a natural way to clear out the collision issue using

$$\delta = \min_{t \in I}\{\|\vec{r}_2 - \vec{r}_1\|(t)\} \tag{7.3}$$

$t \in I \mapsto \|\vec{r}_2 - \vec{r}_1\|(t)$ experimental convexity, Figure 7.1, makes δ available through numerical optimization, the associated test

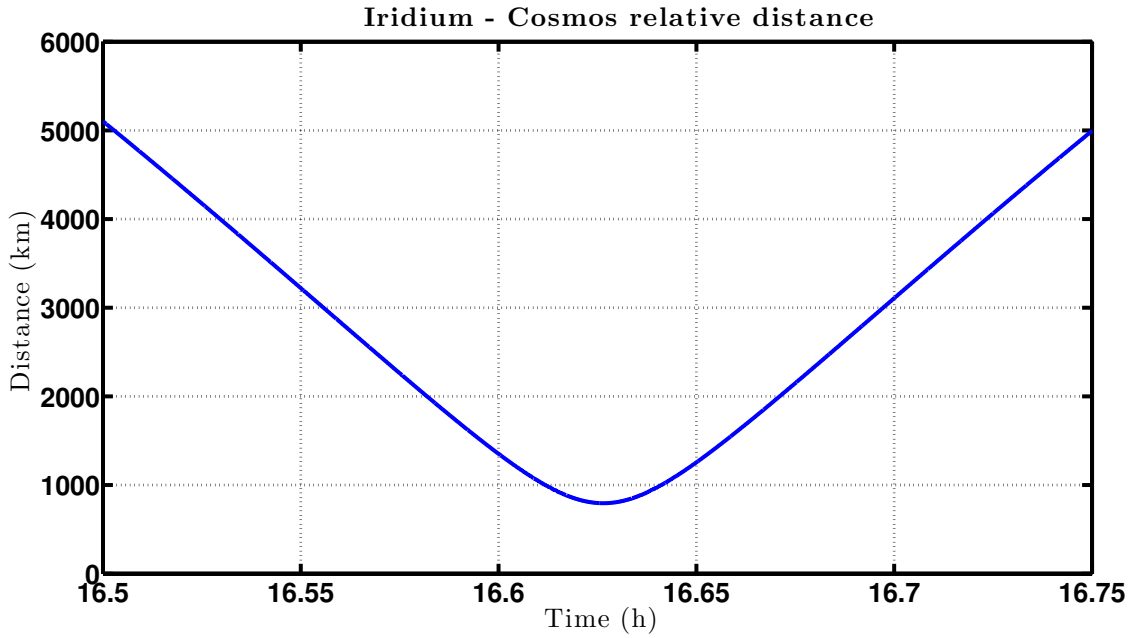
$$\xi(\vec{s}_1^m, t_1^m, \vec{s}_2^m, t_2^m, I) = \begin{cases} 1 & \text{if} \quad \delta \leq d_c \\ 0 & \text{otherwise} \end{cases} \tag{7.4}$$

eventually closing the deal. Things would be all that easy and deterministic, had randomness not barged in.

7.2.2 Random initial conditions lead to uncertainty

Actually, the states are not monitored around-the-clock but merely measured from times to times, by a radar. The Two Line Elements (TLE) provided by NORAD sum up this information and feed the models with the $(\vec{s}_i^m, t_i^m)$ pairs. However, TLEs are inaccurate and their inaccuracy is unknown.

Figure 7.1: Iridium-Cosmos distance on the day before the collision according to TLE.



This uncertainty was our very issue. To cope with this and to reflect the reality, we added independent, identically distributed (*iid*) noises $\vec{\epsilon}_1$ and $\vec{\epsilon}_2$, with density $f_{\vec{\epsilon}}$, to the models' inputs $\vec{s}_i^m$.

$$\vec{E} = \begin{pmatrix} \vec{\epsilon}_1 \\ \vec{\epsilon}_2 \end{pmatrix} \qquad f_{\vec{E}}(\vec{e}) = f_{\vec{\epsilon}}(\vec{\varepsilon}_1) \cdot f_{\vec{\epsilon}}(\vec{\varepsilon}_2) \tag{7.5}$$

$f_{\vec{\epsilon}}$ was chosen Gaussian as described in Table 7.1 based on the field experience of the lab but actually estimating the accuracy of the TLE is a current research topic [28, 55].

The collision issue could not be answered in a cut-and-dried way anymore. It had to be rephrased in a probabilistic fashion itself: what was the probability of collision between the two satellites? *Via* the random counterpart of our deterministic geometrical problem

$$\begin{array}{rcl} i \in \{1,2\}, \forall t \in I, \quad \vec{S}_i & = & \Phi(\vec{s}_i^m + \vec{\epsilon}_i, t_i^m, t) = (\vec{R}_i(t), \vec{V}_i(t)) \\ \Delta & = & \min_{t\in I}\{\|\vec{R}_2 - \vec{R}_1\|(t)\} \\ \Xi & = & \xi(\vec{s}_1^m + \vec{\epsilon}_1, t_1^m, \vec{s}_2^m + \vec{\epsilon}_2, t_2^m, I) \end{array} \tag{7.6}$$

this question was equivalently stated as

$$\mathbb{P}[\{\text{ The satellites collide during } I\}] = \mathbb{P}[\Delta < d_c] = \mathbb{E}[\Xi] \tag{7.7}$$

We then faced a plain expectation estimation problem.

7.2.3 A huge CMC estimation as reference

A huge Monte Carlo estimate was done to serve as a reference. Such budget and time will not be available in a realistic context. It benefited using fast to evaluate Kepler dynamics instead

of SGP4 and an unreasonable more than a week calculation time on four computers.

$$N = 77 \cdot 10^8 \qquad \mathbb{P}\left[\Delta \leq d_c\right] \approx 1.15 \cdot 10^{-6} = \overline{\Xi}_N \tag{7.8}$$

This demonstrated that CMC would be too expensive a choice in practice and that a specific tool was required. Besides, this result served as a reference later on, when testing AST, ANAIS and CE.

7.3 Five collision probability estimations

So as to estimate the unlikely collision probability *i.e.* the expectation of black box test function Ξ with a reasonable calculation cost, a rare event dedicated technique was needed. In order to find the most appropriate rare event technique to this specific case, we tried AST, ANAIS and CE. All numerical settings and parameters are given in Table 7.1. As $d_c = 10$km and the TLE are the penultimate before collision, we estimated the probability that the relative distance went under a safety threshold. The reasoning with a smaller d_c and the actual ultimate TLE would be the very same. All results are given in Table 7.2. In case a non-elliptic trajectory was generated, it was systematically replaced with an elliptic one distributed according appropriately.

We used a small CMC estimation as a way to show the possible improvement. Its mean estimate is close to the reference value. However, as the empirical deviation shows, no estimate was actually accurate. Being multiples of $\frac{1}{N}$, they could not be.

Table 7.1: Estimation parameters and settings in the Iridium-Cosmos collision case

$$\text{TLEs} : 02/09/09 \qquad D = 10^2 Diag(4, 4, 10, 4, 4, 7)$$

$$f_{\vec{\epsilon}} \sim \mathcal{N}(0_6, D^2) \qquad \mathrm{M}\left(x,\, \mathrm{d}y\right) \sim \mathcal{N}\left(\frac{\theta X}{\sqrt{1+\theta^2}}, \frac{\theta^2}{1+\theta^2} D^2\right)$$

$$d_c = 10^4 \qquad N = 10^5 \qquad m = 50 \qquad \iota = 1.05$$

$$n_{AST} = 500 \qquad \beta_{AST} = \frac{4}{5} \qquad \omega = 25 \qquad \theta_0 = 1$$

$$n_{ANAIS} = 1000 \qquad \beta_{ANAIS} = \frac{4}{5} \qquad \kappa = 100 \qquad f_Z^{\dagger,0} = f_{\vec{\epsilon}}$$

$$n_{CE} = 1000 \qquad \beta_{CE} = \frac{4}{5} \qquad \kappa = 100 \qquad f_Z^{\theta} = f_{\vec{\epsilon}}$$

7.3.1 Adaptive Splitting Technique

AST Algorithm 3.2.2 was used and tuned according to Table 7.1. θ was specifically tuned according to Proposition 3.1.4 with the heuristic tuning procedure described in Subsubsection 4.1.3.1. Results showed that AST estimated more accurately than CMC as its estimators

Table 7.2: Iridium-Cosmos collision probability estimations

	Mean estimate	*Erd*	Mean simulation cost	*Erd*
CMC	$1.0 \cdot 10^{-6}$	3.0305	100000	0
AST	$1.0 \cdot 10^{-6}$	0.4294	89375	4.4%
ANAIS	0	0	100000	0
CE	✗	✗	?	?

have a way lesser relative variance as it was *divided by 4*, and for a very similar cost as it consumed on average the same amount of points. The $\frac{2}{5}$ empirical relative deviation was in line with the Chapter 6 experiments.

7.3.2 ANAIS

ANAIS Algorithm 5.1.1 failed to produce one single collision. The curse of dimensionality could be an explanation as the input space dimension is twelve. However, in an attempt to improve ANAIS, two key points of the algorithm, presented in page 77, were reconsidered: the intermediate threshold definition and the auxiliary density definition.

In Algorithm 5.1.1, the threshold sequence (S_k) is not monotonous. Actually, S_{k+1} being greater or lesser than S_k is a random event, as can be seen on Figure 7.2. However, it makes sense to impose that (S_k) is monotonous so as not to step back from the targeted threshold. Therefore, assuming, the target is a great threshold, S_{k+1} should not be allowed to be lesser than any of the previous intermediate thresholds. Likewise, $f_Z^{\dagger,k+1}$ is designed to better estimate the probability of $\{S_{k+1} \leq Y\}$. It needs not using points sampled according to $f_Z^{\dagger,i}, 0 \leq i \leq k$, because these intermediate densities are designed with respect to previous lower thresholds.

These ideas still have to be formulated into an efficient algorithm.

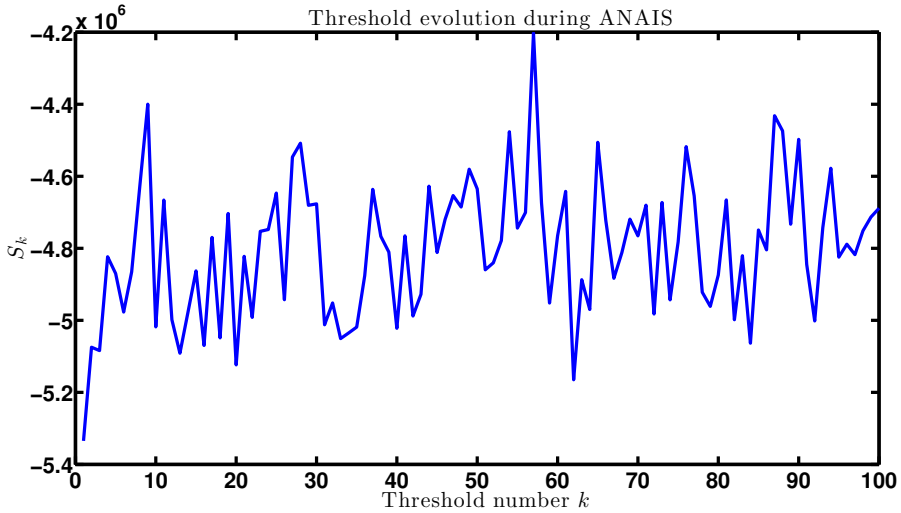


Figure 7.2: Intermediate threshold sequence while estimating the probability of collision *via* ANAIS.

7.3.3 Cross-Entropy

CE Algorithm 2.2.1 failed completely: it crashed. Actually it generated only non-elliptic trajectories. We used AST to generate points distributed according to conditionally to resulting into the trespassing of the security thresholds and made an histogram in each direction of $\mathbb{R}^1 2$. As Figure 7.3 shows, the chosen uni-modal parametric family was not adapted. Choosing a more complex exponential change of measure seemed a reasonable choice as a well chosen polynomial exponential change of measure, as defined in Definition 7.3.1, could theoretically adapt the auxiliary density to the multi-modality.

Definition 7.3.1 (Polynomial Exponential Family)**.** *The density function f_Z^θ is said to be part of the Polynomial Exponential Family (PEF) if the following holds.*

1. $\forall x \in \mathbb{R}^d, p_\theta(x)$ *is a polynomial whose coefficients are in* $\theta \in \mathbb{R}^l$*, for some* l*,*
2. $I(\theta) = -\ln\left(\int_{\mathbb{R}^d} \exp\left(p_\theta(x)\right) \lambda(\mathrm{d}x)\right)$*,*
3. $\forall x \in \mathbb{R}^d, f_Z^\theta(x) = \exp\left(p_\theta(x) + I(\theta)\right)$

The NEF is a subset of the PEF which is part of the class of exponential changes of measure from Definition 2.2.1.

However sampling according to an exponential change of measure is an issue as soon as one steps out the well-known cases *i.e.* Poisson, binomial, negative binomial, normal, and gamma. Besides, there was no way to know beforehand the number of mode in each dimension and set the polynomial degrees accordingly. Refined cluster counting techniques such as [8] are statistical and can therefore only be used once the data has already been generated.

7.3.4 Outcome

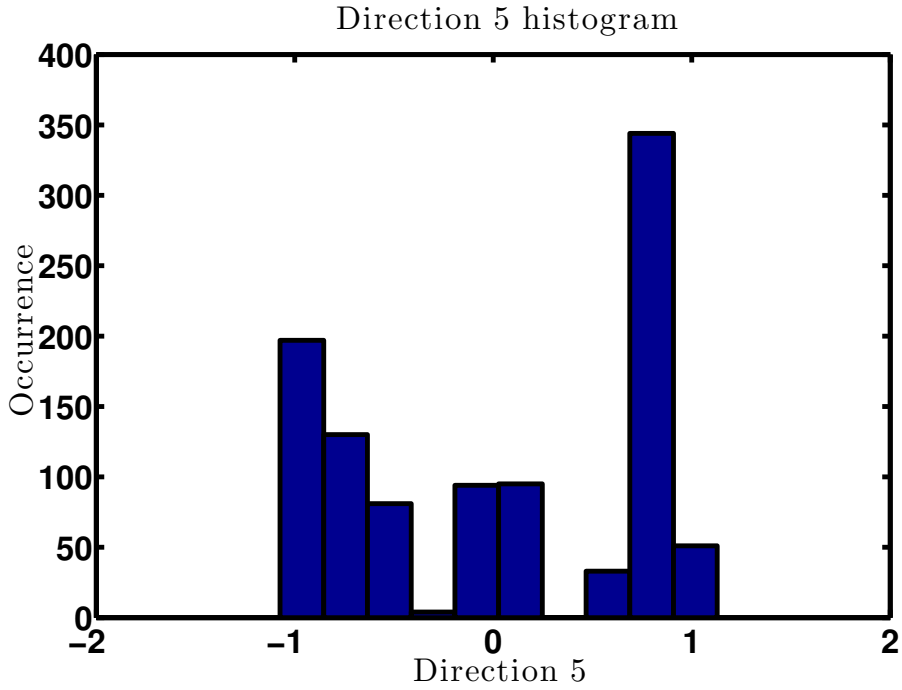
CE crashed because of a poor choice of the parametric family of the auxiliary densities: it was not adapted to multi-modality. In practice, using a non-parametric method before switching to a parametric one after enough information has been accrued to make an educated seemed a good choice. As for ANAIS, it suffered from the high dimension of the system and of tuning that allowed the auxiliary sampling density bandwidth to go to zero. Its enhanced version, performed better but failed nonetheless. Unexpectedly, only AST delivered a useful estimation, despite its crude tuning. In order to improve it, we performed the following experience.

7.4 Experiments on AST sensitivity

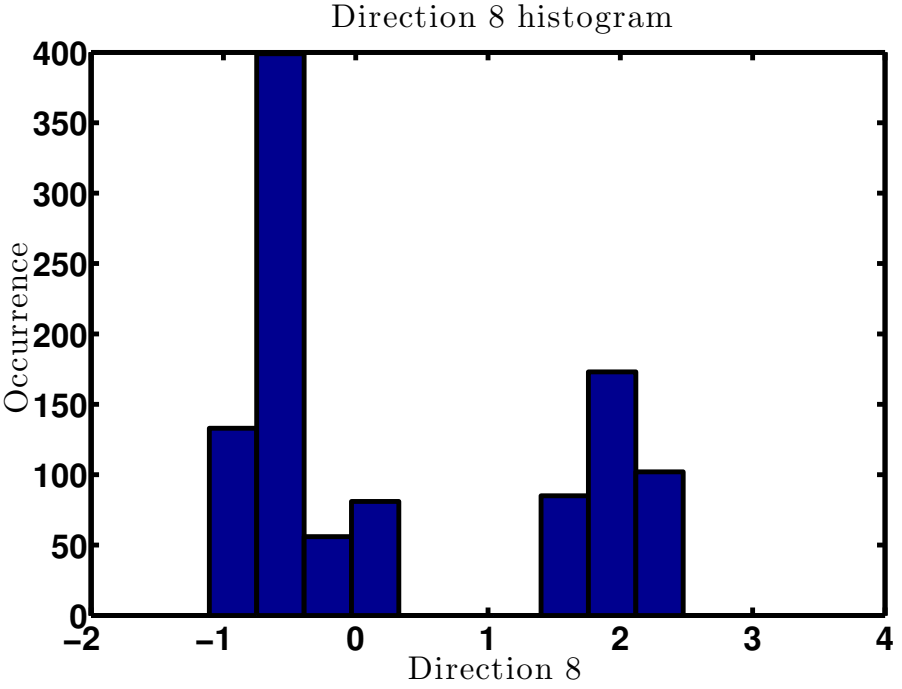
AST performance came with a price: parameter tuning. We knew this much since Chapter 6. However the unexpected good performance of AST renewed our interest in a sound tuning. So, we looked for an empirical realistic tuning rule.

To test AST's sensitivity[2] to its variance variation coefficient and number of iterations,

[2] AST estimator dependence w.r.t. to number of cloud particles n has already been partly discussed in Section 3.3 when it came to its asymptotic behavior in the ideal case where every particle set is *iid*.



(a) Direction 5: Iridium



(b) Direction 8: Cosmos

Figure 7.3: Histogram along direction 5 and 8 of the input space of 1029 samples distributed according to $f_{\vec{\epsilon}}$ conditionally to resulting into the security thresholds .

their values were changed to see their impacts on the estimator's behavior *equibus paribus*.

$$\iota : \text{variation coefficient of } \theta \qquad \beta = \frac{4}{5} : \text{quantile level} \tag{7.9}$$

$$\omega : \text{number of kernel applications} \qquad n = 2000 : \text{number of particles} \tag{7.10}$$

$$T : \text{simulation duration in seconds} \qquad \Gamma : \text{simulation cost} \tag{7.11}$$

$$\kappa_{\max} = 300 : \text{maximal number of thresholds} \qquad m = 100 : \text{number of estimations} \tag{7.12}$$

$$\mathrm{d}_{\mathrm{col}} = 10\mathrm{km} : \text{collision distance} \qquad \widetilde{\Xi}_n : \text{estimated value} \tag{7.13}$$

7.4.1 Settings.

Here is how we interpreted the numerical experiment presented in Table 7.3.

- Column-wise: number of iterations ω influence
  - Column $\iota = 1$ showed that if the transition kernel variance θ does not evolve, regardless of how many times the transition kernel is applied, AST fails to provide an estimation. A possible explanation is that a static θ can not adapt itself so as to efficiently explore the input space.
  - Columns $\iota = 1.05$ and $\iota = 1.10$ showed that the more the transition kernel is applied, the closer to $\overline{\Xi}_N$ the estimation and the lower the variance but the costlier the estimation. Intermediate estimations are then more consistent but come at a higher cost.
- Row-wise: variance variation coefficient ι influence
  - All three rows suggest that a gradual variation of θ leads to a smaller variance.

Table 7.3: AST sensitivity experiment w.r.t. number of kernel application ω and kernel parameter variation coefficient ι as *per* algorithm 3.2.2.

Parameters $n = 2000$ and $\alpha = \frac{4}{5}$ were fixed while ω and ι varied.

ω \ ι	1			1.05			1.1		
5	869.2	2992	$5.25 \cdot 10^{-26}$	205.0	657.5	0.40	180.5	640.1	0.6
	3.3	0	98.9	5.1	1.8	25.1	3.5	1.8	23.9
10	1784.8	5982	$9.45 \cdot 10^{-26}$	376.8	1235.6	0.96	340.2	1232	1.01
	5.4	0	79.6	5.8	1.6	19.9	5.9	1.5	20.7
20	3470.2	11962	$1.8 \cdot 10^{-25}$	748.3	2449.6	1.06	787.2	2450.8	1.07
	3.46	0	56.5	3.8	1.3	15.1	3.9	1.4	18.0

Legend: $\overline{T}_m$ $\quad \overline{\Gamma}_m \cdot 10^{-3}$ $\quad \overline{\widetilde{\Xi}}_{nm} \cdot 10^{6}$

$\rho(T, m) \cdot 10^2$ $\quad \rho(\Gamma, m) \cdot 10^2$ $\quad \rho\left(\widetilde{\Xi}_n, m\right) \cdot 10^2$

7.4.2 A rule of the thumb for tuning

The best trade-off seemed to have θ close to but different from 1 and as many kernel application as possible and, as could be expected, high n *i.e.* as high a simulation budget as possible. This was not much but this was as far empirical work could get us: there is no going around theoretical study when dealing with AST.

7.5 Conclusion

As a final way to test CE, ANAIS and AST ability to estimate rare event probability without any prior about the transfer function, we estimated the probability that satellites Iridium and Cosmos got dangerously close. This case study was an attempt at not relying on on restrictive hypothesis to do the estimation.

CE crashed because the chosen parametric density family was not adapted to the presence more than one connected component of interest in the input space. Actually, the number of connected component can not be known before performing the estimation. Besides, sampling to an auxiliary density that is an exponential change of measure and is adapted to multiple connected components is an issue. So we decided not to use CE unless the number of connected component and a way to sample according to an adapted exponential change of measure is known. These conditions were not met in the last part of our work.

ANAIS failed to produce a single collision and was therefore partly rethought.

AST delivered the most accurate estimation, though crudely tuned. Experimentally, a shaking parameter close to one but not equal to one and many transition kernel applications seemed a good rule of the thumb. This empirical intuition remained to be backed up with theoretical results.

Conclusion

We built some experience of the Cross-Entropy (CE), Non parametric Adaptive Importance Sampling (NAIS) and Adaptive Splitting Technique (AST) estimating a 10^{-5} probability and $(1-10^{-5})$-quantile twice.

We first used a $\mathcal{R}_2$ *i.e.* a dimension 2 Rayleigh random variable as this random variable is completely known theoretically. It came out that though all estimators eventually delivered accurately and reasonably fast, they all come with their own drawbacks. CE's optimisation step is cumbersome as soon as the the auxiliary density is outside the Natural Exponential Family such as when the subset of interest is multi-modal. NAIS can not work until the initial distribution does generate rare events. AST requires a defter tuning that induces a random and hard to control cost, and that may result in a higher variance.

The second test involved a $\mathcal{R}_{20}$. This way, the estimator behaviours w.r.t. a high dimension input was observed. NAIS fails in high dimension and should not be used in such cases. CE and AST quantile estimations were not affected by the dimension increase. CE and AST estimated the probability without bias but with higher variance than with $\mathcal{R}_2$. CE and AST appeared as stable with respect to high dimensions.

The Non parametric Adaptive Importance Sampling (NAIS) algorithm can be not user friendly because it requires an initial auxiliary density that straight away generates rare events. Such a density being precisely what the user is looking for, NAIS is of very limited practical interest without important *a priori* information. To provide NAIS with a kickstand, we proposed the Adaptive NAIS (ANAIS) algorithm which does allow the user to choose the original density as initial auxiliary density: there is no need for *a priori* information. To this purpose, ANAIS takes a leaf out of CE and AST books and builds its auxiliary density *via* an increasing sequence of empirical quantile based thresholds. This circumvents the initial density choice issue of NAIS. Besides, experiments show that ANAIS outperforms NAIS, especially in high dimensions. This is why NAIS was discarded and only ANAIS was used from then on, even though ANAIS requires theoretical deepening.

ONERA gave us an opportunity to practice *via* the estimation of the probability of collision of two satellites. It was an opportunity to improve the satellite monitoring procedure and be able to estimate such probabilities without hypothesis.

CE crashed because the chosen uni-modal parametric density family was not adapted. We considered building a exponential change of measure adapted to the presence of more than one connected component of interest but sampling according to a such law requires a specific theoretical development that is yet to be done.

ANAIS did not produce any collision and returned a zero probability. Ways of improvement have been spotted but they still remain to be translated into an efficient algorithm.

AST estimated the aimed probability with $\frac{2}{5}$ empirical relative deviation, which is very similar

to the precision it yielded in simpler cases. Though AST requires study with respect to its tuning, it seemed the more mature choice.

This case study resulted in two major decisions. The fist one was not to use CE without a reasonable about of information about the system. Actually, this technique requires an unaffordable amount of beforehand information about the transfer function to choose the auxiliary density family. As we considered black box systems, CE was simply not used from then on. The second was to point out AST as a rare event probability estimation technique of great interest for our black box case, when tuning is mastered. As for ANAIS, it seemed to need more maturing.

We then moved on to another case study in order to practice more realistic extreme quantile estimation. We had to determine a spacecraft propeller fallout safety zone estimation in Part III.

Part III

From safety distance to safety area *via* Extreme Minimum Volume Sets

Introduction

In Part II, we had grown some practical experience of CE, ANAIS and AST. This practice had culminated in the successful estimation of the probability of collision between two satellites. However, the ONERA is not interested in probability estimation only but in quantile estimation as well.

Actually, ONERA, as a partner of major aerospace French projects, often has to design and assess safety. Safety is to be understood on a case by case basis. However, the reference measure is the quantile. As safety requirements increase, the quantile level becomes extreme and therefore requires dedicated attention. To evaluate the practical interest of rare event specific tools, the ONERA proposed us an extreme quantile estimation case.

The case was presented as typical safety distance estimation problem. It involved a spacecraft booster in free fall. The booster is expected to hit the ground on a risk free location but actually falls in a random location arounds its target. The spread is due to random exterior events, such as the atmospheric conditions, and because the actual dropping conditions, such as the angular orientation, are not exactly those that were designed.

We were expected to propose an adapted extreme quantile estimation tool. However, there was more than met the eye.

Chapter 8

Spacecraft booster fallout safety perimeter estimation

After propelling its load, a spacecraft booster separates from it and should fall onto the ground without hitting anything valuable. In this case, the α-level safety zone is commonly defined in the spacecraft industry via the α-quantile of the random distance to the desired landing position on the ground. A point is deemed α-safe if it is more than q_α away from the landing target. This is a common formal representation but in the specific case that was brought to us the maximum safety level was high, $\alpha = 1 - 10^{-6}$, just as the simulation cost. This toy mission came as a good extreme quantile estimation practice ground[1].

8.1 Problem formalisation and random input generation

We first modeled the whole problem in Subsection 8.1.1 and then briefly chose an adapted sampling strategy.

8.1.1 Landing point modeling

The booster landing point on the ground depends on six variables: the angular orientation, location, speed, weight and slope of the booster at dropping time plus the wind strength. Said six parameters, though uncertain, were bounded. We knew nothing more as the whole system had been designed by another department. This is why we modeled the system as follows.

$$\phi : \left| \begin{array}{ccc} [0,1]^6 & \to & \mathbb{R}^2 \\ x & \mapsto & y \end{array} \right. \qquad \delta = \|y\| = \vartheta(x) \tag{8.1}$$

In practice, the mapping was translated so that the target landing position coincided with 0_2.

[1] This work has been published in [65]

8.1.2 Formal problem and simulation budget

Given a set of safety levels and a simulation budget, we had to estimate all associated quantiles. The system simulator was more realistic than the experiments done in Part II.

$$A = \left\{1 - 10^{-1}, 1 - 10^{-2}, 1 - 10^{-3}, 1 - 10^{-4}, 1 - 10^{-5}, 1 - 10^{-6}\right\} \quad N = 5 \cdot 10^4 \quad m = 50 \tag{8.2}$$

Estimating $\Delta = \|Y\|$ quantiles whose level are in A seemed an easy task for AST Algorithm 3.2.4 and ANAIS Algorithm 5.1.1.

8.2 Quantile indeed?

8.2.1 A naive CMC quantile estimation

In order to better compare CMC and ANAIS results we first performed the quantile estimation with CMC. Figure 8.1 showed, as expected, that CMC failed to distinguished the quantile of most extreme levels. Empirical relative deviations were small because CMC is a consistent estimator, but the estimation was a failure. We then moved on to ANAIS estimation.

8.2.2 A successful AST quantile estimation

In order to benefit from the previous work, we used a Gaussian vector to simulate $\mathcal{U}([0,1]^6)$ using Rayleigh's law cumulative distribution function's inverse.

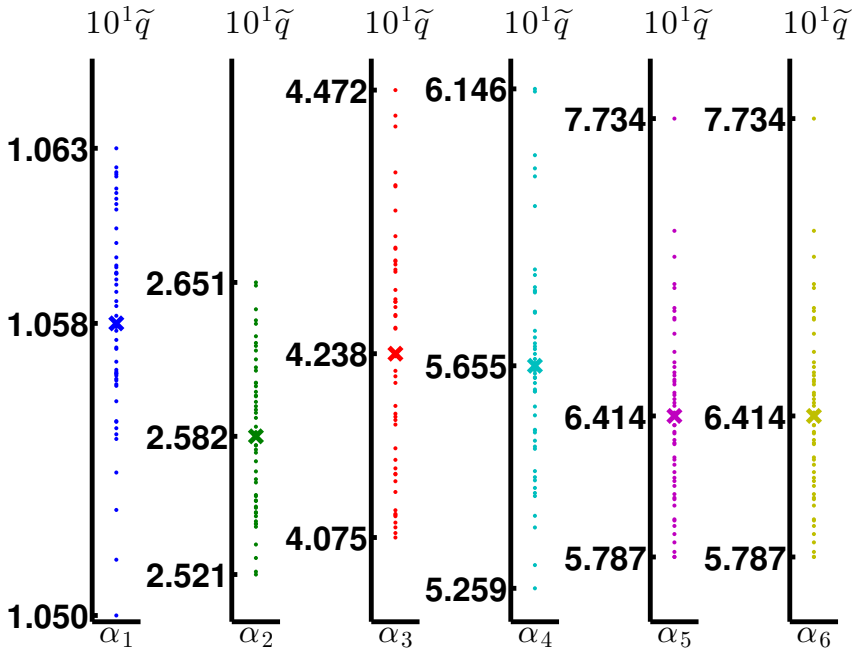
$$\begin{aligned}
H &\sim \|\mathcal{N}(0_2, I_2)\| && (8.3)\\
&\sim f_H(h)dh && (8.4)\\
f_H(h) &= \mathbf{1}_{\mathbb{R}^+}(h) h \exp^{-\frac{h^2}{2}} && (8.5)\\
F_H(h) &= \int_{-\infty}^{h} f_H(r)dr && (8.6)\\
&= 1 - \exp^{-\frac{\max(h,0)^2}{2}} && (8.7)\\
0 \leq a < b \leq 1 &\Rightarrow \mathbb{P}[a \leq F_H(H) \leq b] = b - a && (8.8)
\end{aligned}$$

The six $\mathcal{U}([0,1]^6)$ components were independently, identically distributed as $F_H(\|Z\|)$ with $Z \sim \mathcal{N}(0_2, I_2)$. This allowed the use of the same framework as before. Actually, we had no explicit Markovian kernel reversible with respect to the uniform law.

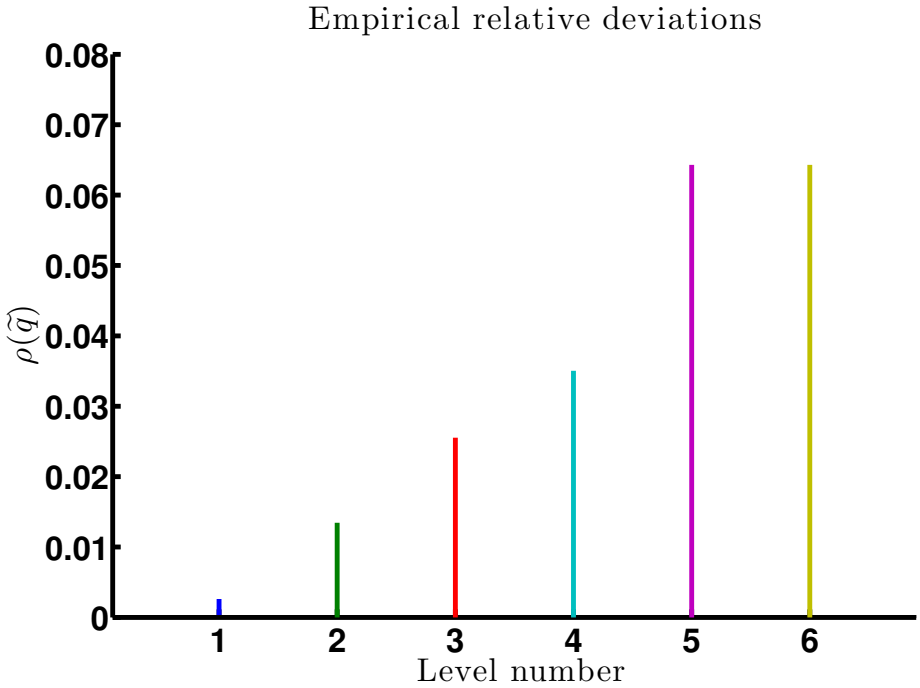
The algorithm was tuned according to Proposition 4.1.1 and set as follows.

$$\begin{array}{cccccccccc}
\kappa_{max} & = & 20 & \widetilde{\beta} & = & \frac{3}{5} & \iota & = & 1.1 & \\
n & = & 125 & \omega & = & 25 & \theta^0 & = & 1 & \mathrm{M}(x, \mathrm{d}y) \sim \mathcal{N}\left(\frac{\theta X}{\sqrt{1+\theta^2}}, \frac{\theta^2}{1+\theta^2}\right)
\end{array} \tag{8.9}$$

and Figure 8.2 results were produced. Though κ_{max} was always reached, all quantiles, including those of highest levels, had been distinctly estimated with little empirical relative deviation. We then made an ANAIS estimation.

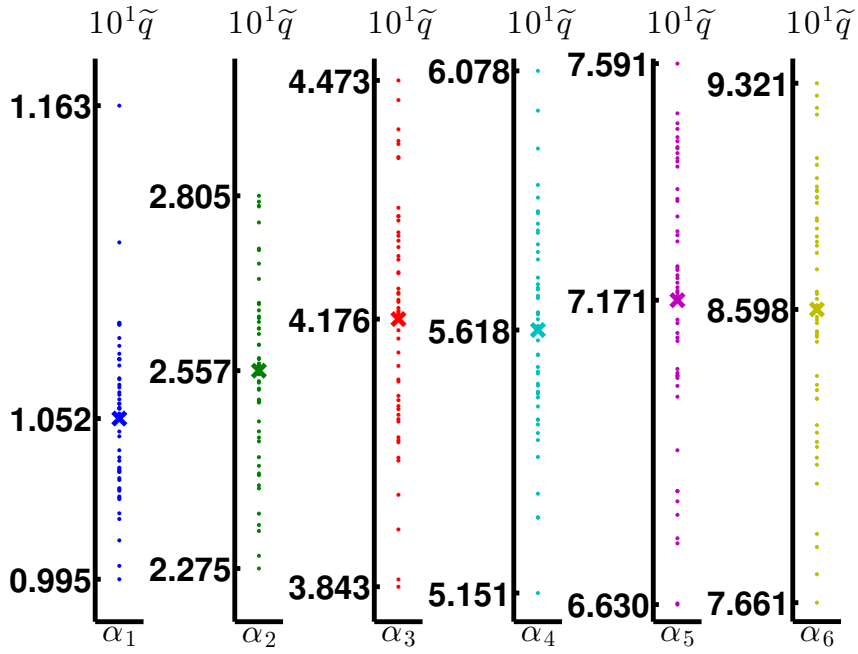


(a) For each level α_i, 50 CMC quantile estimates are represented as $\cdot$ and their mean as $\times$.

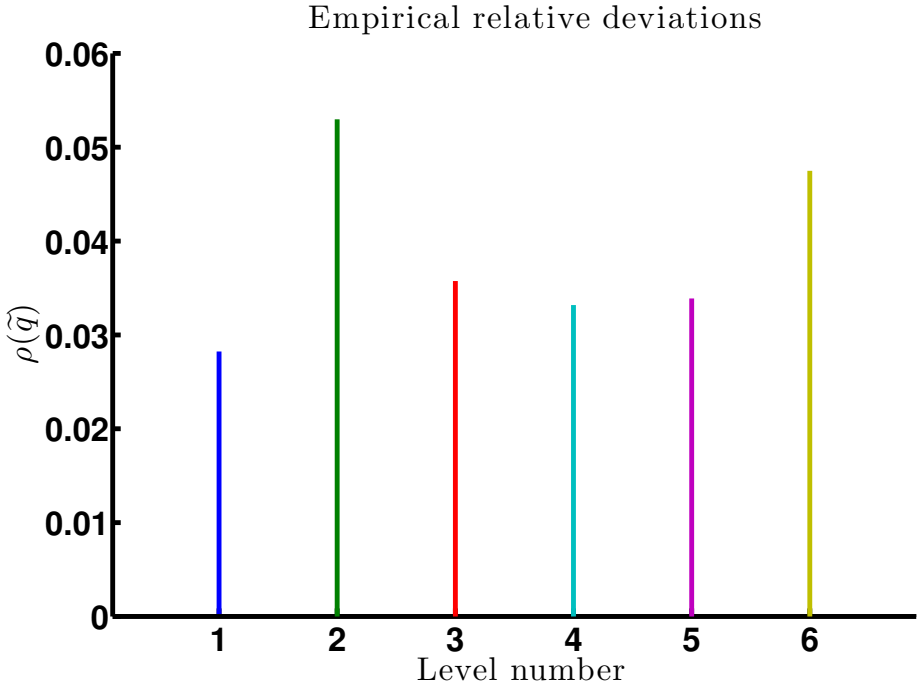


(b) The quantile estimator shows little empirical relative deviation for all levels.

Figure 8.1: Booster fallout safety distance estimation *via* CMC as described in Subsection 8.2.1. A is sorted in increasing order so that α_i is its i^{th} value.



(a) For each level α_i, 50 AST quantile estimates are represented as $\cdot$ and their mean as $\times$.



(b) The quantile estimator shows little empirical relative deviation for all levels.

Figure 8.2: Booster fallout safety distance estimation *via* AST as described in Subsection 8.2.2. A is sorted in increasing order so that α_i is its i^{th} value.

8.2.3 A successful ANAIS quantile estimation

To parameter Algorithm 5.1.1, we chose Epanechnikov kernels because the sampling space is bounded. Besides, we used the rejection method so that all actually used samples are in $[0,1]^6$. Luckily enough normalizing the estimator by the mean likelihood $\overline{\mathrm{w}}$ freed us from needing the auxiliary density normalizing constant explicitly. As for the bandwidth, it was set isotropically *via* `AMISE`. ANAIS was set to estimate the most extreme level quantile and others were estimated along with it.

$$n = 1000 \qquad \kappa = 50 \qquad \widehat{\beta_\dagger} = 0.90 \qquad h \text{ via } \texttt{AMISE} \tag{8.10}$$

$$\forall\,(x,z) \in \mathbb{R}^6 \times \mathbb{R}^6, \quad \underline{K}_6\,(x,z,h) = \frac{\mathbf{1}}{\left\{\left([0,1]^6\right)^2\right\}}\,(x,z) \prod_{i=1}^{6} \left(1 - \left|\frac{x_i - z_i}{h_i}\right|\right) \mathbf{1}_{[0,1]}\left(\left|\frac{x_i - z_i}{h_i}\right|\right) \tag{8.11}$$

All results are shown on Figure 8.3. All the estimated quantiles were different and, greater and rare distances had been generated. The empirical relative deviations were much greater than those of CMC, especially for the quantiles of lowest levels, yet only one empirical relative deviation was over a fifth. Besides, the empirical relative deviations of the three most extreme level quantile estimators had the same order of magnitude of the CMC estimators. This is because ANAIS is designed to estimate one quantile at a time, the one of most extreme level. Intermediate densities are designed to estimate intermediate quantiles, but they are build out of a lesser budget than the last one which uses all the simulation budget and therefore induces smaller variance. As a matter of fact, the other quantiles came as by-products and happened to be consistently estimated. In practice however, one is looking for maximal security and mostly cares for the safest places.

ANAIS estimated extreme quantiles better than CMC with the same budget. That was good news. Then, we looked at the actual landing positions.

8.2.4 Was quantile the answer?

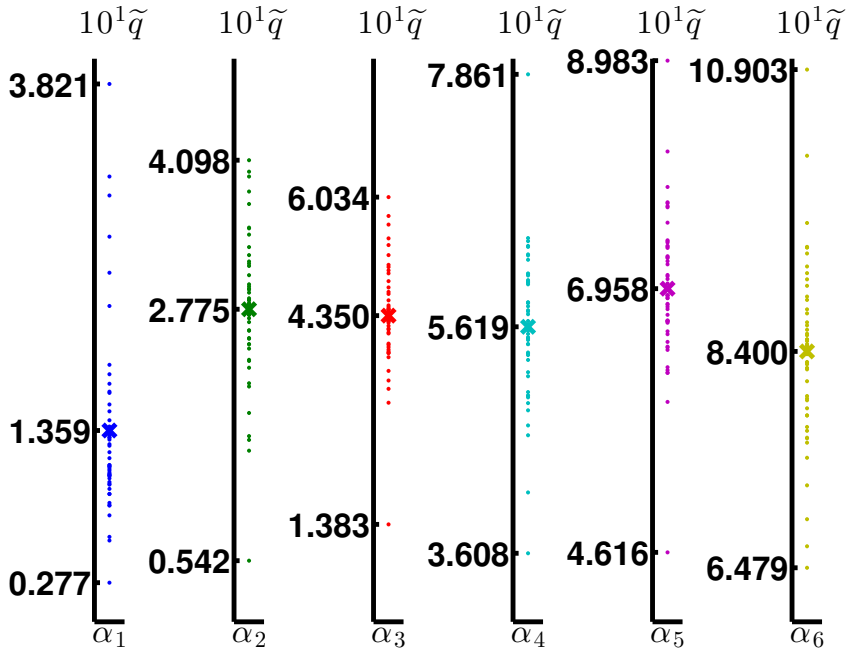
Actually, we represented landing points and safety perimeters generated through CMC and ANAIS on Figure 8.4. We already knew ANAIS had generated rare landing positions. What caught our eye was that the spatial distribution of landing position around the origin was blatantly not isotropic.

Quantiles cannot represent a spatial distribution. We had discarded a wealth of information summarizing the *spatial* distribution through circles with the *scalar* quantiles as *radii*. We did estimate extreme quantiles efficiently, but the very concept of quantile did not stood as the appropriate tool with respect to our purpose.

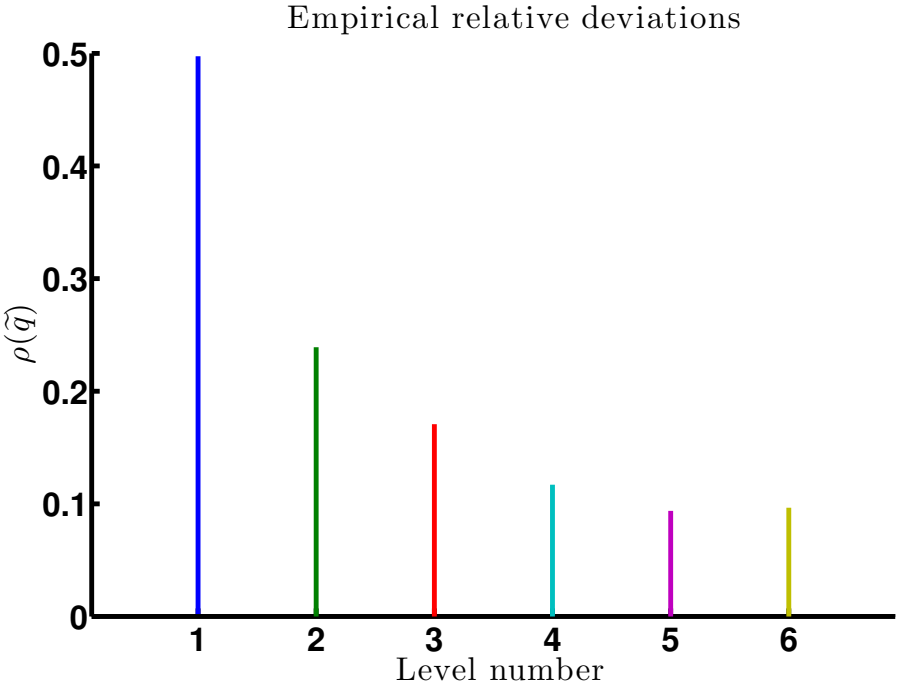
How to deterministically summarize the distribution of a $\mathbb{R}^d$ valued random variable? Fortunately enough, this question was new to neither statistics nor probability.

8.3 Three approaches to spatial distribution

Before interrogating the distribution of a random variable Y on $\mathbb{R}^d$, $d > 1$ *via* deterministic quantities, one has to know which information is to be retrieved and then build the appropriate

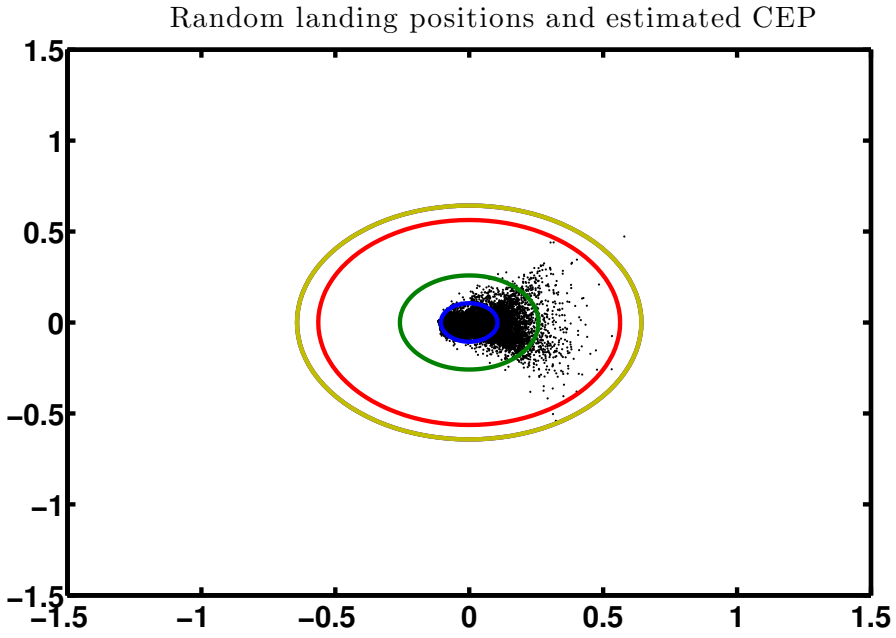


(a) For each level α_i, 50 ANAIS quantile estimates are represented as $\cdot$ and their mean as $\times$.

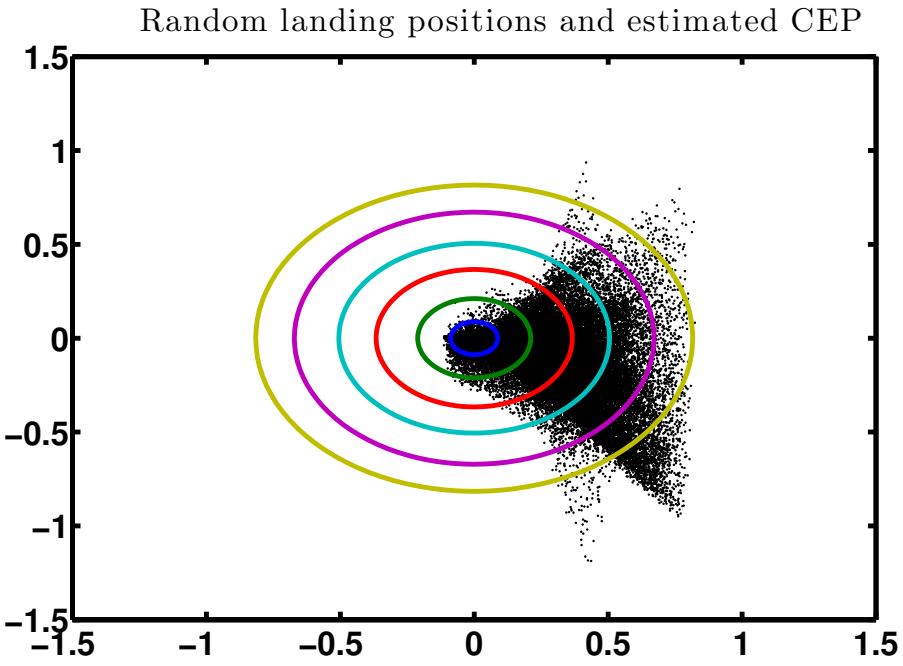


(b) The quantile estimator shows smaller and smaller empirical relative deviation as the level increases until it becomes as small as CMC's for the most extreme level.

Figure 8.3: Booster fallout safety distance estimation *via* ANAIS as described in Subsection 8.2.3. A is sorted in increasing order so that α_i is its i^{th} value.



(a) A set of landing positions with safety perimeters *via* CMC.



(b) A set of landing positions with safety perimeters *via* ANAIS.

Figure 8.4: Landing positions and safety perimeters – Circular Error Probable– estimated *via* CMC and ANAIS with $N = 5 \cdot 10^4$ simulations as described in Section 8.2.

tool. Dimension one quantiles and statistics are of very common use and are efficient as long as $d = 1$ *i.e.* Y is on the line. Using these tools in greater dimensions is theoretically questionable. However, in some industries their counterparts are not known and understood enough so needs can be articulated through their more appropriate means.

The current tool in the aerospace industry is the Circular Error Probable (CEP) which is exactly what has been done in Section 8.1. The scalar score criterion ϑ can be chosen more astutely than a plain Euclidean distance but CEP essentially assumes the random variable spreads isotropically around the user specified center c and estimates $\vartheta(Y - c)$ quantiles. In practice however, there is not always an obvious choice of center and a meaningful scalar criterion is often hard to design. One has to go behind dimension one quantiles. With this respect, statistics and probability come with different solutions: multivariate quantiles – Subsection 8.3.1 – and Minimum Volume Set respectively – Subsection 8.3.2 –.

8.3.1 Multivariate quantiles

Statisticians have been addressing the question of the extension of the quantile definition to greater dimensions for decades from the seminal Mahalanobis distance quantile [56]. However, an unique fundamental mathematical object on which the whole theory could be based still has to be found. As said object is only known through a set of properties it is expected to respect [84], there are many *ad hoc* candidates [39, 49, 83] . We will briefly sketch the state of the art in Subsubsection 8.3.1.1, introduce the properties *a minima* expected from any candidate multivariate quantile in Subsubsection 8.3.1.2 and eventually, in Subsubsection 8.3.1.3, we point to our uncompleted attempt to multivariate quantile that is further detailed in Appendix A.

8.3.1.1 The DORQ paradigm

According to [84], multivariate quantile candidates span four main quantities: Depth, Outlyingness, Ranking and Quantile (DORQ). Depth and Outlyingness are scalar measures of how close to and respectively remote from the distribution core, from its multidimensional median, a point is. A given point's Ranking directionally translates this idea within the unit ball: the origin represents the median and the direction from the median to the point in some sense and the magnitude is the Outlyingness. The multivariate Quantile associates a $\mathbb{R}^d$ point to a Ranking and can be induced by the Depth. Besides, Quantile and Ranking on the one hand and Outlyingness and Depth on the other hand are (inversely) equivalent. There are hence four entrances to the DORQ paradigm. The only entrance requirements are invariance and equivariance.

8.3.1.2 Invariance and equivariance

A very pragmatic idea is that the information retrieved from a given data set should not depend on how the data is represented and ideally should not change whether one considers the original sample set or its image through a bijection. Invariance formalizes this idea when talking about Outlyingness and Depth while Equivariance does so as far as Ranking and Quantile are concerned. However, bijection equivariance is a very strong requirement and in practice equivariance to translation, rotation and rescaling is enough so that no usual data

representation fouls the retrieved information. We developed multivariate quantile proposal as well.

8.3.1.3 Isoquantile curves

When designing the safety zone, we suggested isoquantile curves as a way to go beyond the Circular Error Probable (CEP) definition of the fallout zone. The α-level c-centered isoquantile curve of a $\mathbb{R}^d$ random variable Y basically links the α-quantiles of Y conditionnaly to being on $c + \mathbb{R}^+\vec{u}$, assuming this makes sense, when $\vec{u}$ describes the unit sphere. Isoquantile curves describe the *spatial* spread of the random landing position around the intended one. However, this work is far from completion, for lack of equivariance analysis for instance. Its current embryonic stage is presented in Appendix A nonetheless. Besides, we addressed the fallout zone definition from the other side of data with the probability density driven Minimum Volume Set definition.

8.3.2 Minimum Volume Sets

Probability theory comes with a tool to spatially summarise spatial distributions just as statistics: Minimum Volume Sets (MVS) . At first sight, MVS abandons the very concept of quantile to focus on high probability low volume sets [82].

Definition 8.3.1 (Minimum Volume Set class)**.** *Let $Y \sim \eta$ be a random variable on $\mathbb{R}^d$, λ the Lebesgue measure and $\mathcal{C}$ a set of $\mathbb{R}^d$ subsets that are measurable for both η and λ. Given $\alpha \in [0, 1]$, the α-level minimum volume set class $\mathcal{M}_\alpha^{\mathcal{C}}$ of Y with respect to the Lebesgue measure over $\mathcal{C}$, is the set of $\mathcal{C}$ elements whose volume is the smallest among those whose probability is at least α.*

$$\mathcal{C}(\alpha) = \{\mathbf{B} \in \mathcal{C} \,|\, \eta(\mathbf{B}) \geq \alpha\} \tag{8.12}$$

$$\mathcal{M}_\alpha^{\mathcal{C}} = \underset{\mathbf{B} \in \mathcal{C}(\alpha)}{\arg\min} \{\lambda(\mathbf{B})\} \tag{8.13}$$

If η is absolutely continuous with respect to λ, Y has a probability density function f_Y w.r.t. λ. Then a sensible intuition is that a $\mathbb{R}^d$ subset that concentrates high density points – *i.e.* dangerous points in our case – in a low volume is a $\mathcal{M}_\alpha^{\mathcal{C}}$ element. Such domains would be defined spatially and have a practical interpretation: any given point of theirs would be a likely Y value. In our case, Y is the random landing position of the spacecraft booster. Such a high density subset would indicate the places *not* to be for the booster is very likely to hit the ground (and you) there.

However, a MVS is a $\mathbb{R}^d$ subset that solves a constrained integral valued optimisation problem. This is not something easy to handle. Could it be rephrased in a more convenient fashion – Subsubsection 8.3.2.1–? How to actually estimate it in practice – Subsubsection 8.3.2.2– ?

8.3.2.1 A Y density level as a MVS

In our setting, $Y = \phi(X)$ and the existence of f_Y is a matter of Geometric Measure Theory [36, 64]. To avoid technicalities, we will assume f_Y exists. Necessarily the input space dimension is

greater or equal to the output space dimension. This is usually the case in complex systems. Furthermore, the class of sets of interest will not be the Borel algebra on $\mathbb{R}^d$ but the following.

$$\mathcal{C} = \left\{ \mathbf{B} \subset \mathbb{R}^d \middle| \mathbf{B} = \bigcup_{i=1}^{n} I_i, n < \infty, \text{ where } I_i \text{ is a } d\text{-dimensional interval} \right\} \cup \emptyset \tag{8.14}$$

This class is wide enough for practical purposes. As for the "high density point subset", it is defined as a f_Y level set.

Definition 8.3.2 (Density level set)**.** *Let f_Y be the density of Y w.r.t. the Lebesgue measure.*

$$\forall t \in [0, \sup f_Y], \qquad \mathbf{B}_t = \left\{ y \in \mathbb{R}^d \,|\, f_Y(y) \geq t \right\} \qquad \alpha_t = \eta(\mathbf{B}_t) \qquad u_t = \lambda(\mathbf{B}_t) \tag{8.15}$$

The density level set $\mathbf{B}_t$ is the set of all points whose density is greater than or equal to t.

Such sets can be measured by both λ and $\eta = f_Y \lambda$. [75] formalises this intuition and gives conditions under which such a subset belongs to $\mathcal{M}_\alpha^{\mathcal{C}}$. Its main result can now be stated. The proof can be found therein.

Theorem 8.3.1. *Let $\mathcal{C}$ be defined according to equation (8.14). Furthermore, choose any $t \in [0, \sup f_Y]$.*

- $\left[\lambda\left(f_Y^{-1}(t) \right) = 0 \right] \Rightarrow \left[\mathbf{B}_t \in \mathcal{M}_{\alpha_t}^{\mathcal{C}} \right]$.
- $\left[\lambda\left(f_Y^{-1}(t) \right) > 0 \right] \Rightarrow \left[\forall \alpha \in \left] \lim_{h \to 0} \alpha_{t+h}, \alpha_t \right[, \nexists t' \in [0, \sup f_Y] \text{ such that } \mathbf{B}_{t'} \in \mathcal{M}_{\alpha_t}^{\mathcal{C}} \right]$.

Not all minimum volume sets are density level sets but the later definitely make sense with respect to our safety zone definition objective. This density based approach ties the MVS of interest, which are $\mathbb{R}^d$ subsets, with the scalar mapping f_Y but is not operational. Actually, f_Y is usually unknown and objectives are more often probabilities than density thresholds. How can Theorem 8.3.1 be more practice oriented?

8.3.2.2 MVS density level set estimation conditions

To make MVS density level set estimation tractable for a given $\alpha \in [0, 1]$, [76] suggests to rephrase their definition using the fact that "by monotonicity of the Lebesgue measure, $\lambda(\mathbf{B}_t)$ is a decreasing function of t and minimizing $\lambda(\mathbf{B}_t)$ is equivalent to maximizing t."

$$[\mathbf{B}_t \in \arg\min \{ \lambda(\mathbf{B}_l) \,|\, \eta(\mathbf{B}_l) \geq \alpha \}] \Leftrightarrow [t \in \arg\max \{ l \in [0, \sup f_Y] \,|\, \eta(\mathbf{B}_l) \geq \alpha \}] \tag{8.16}$$

Then, as $\eta(\mathbf{B}_t) = \mathbb{P}[f_Y(Y) \geq t]$, [76] authors make a decisive step toward *terra cognita*:

$$\arg\max \{ t \in [0, \sup f_Y] \,|\, \mathbb{P}[f_Y(Y) \geq t] \geq \alpha \} \text{ is the } (1-\alpha)\text{-quantile set of } f_Y(Y). \tag{8.17}$$

The good news is MVS density level set estimation can this way be boiled down into estimating a quantile of $f_Y(Y)$. The bad news is f_Y is unknown.

[14] considers the plug-in estimator idea to replace f_Y with its kernel estimation $\hat{f}_Y$ and proves it to lead to a consistent approximation of the sought density level set. If one can evaluate the integral, that is... [76] then comes handy again. In their after-specified result and under the same hypothesis [14] used, their authors show that approximating the integral optimisation with a $\hat{f}_Y(Y)$ quantile CMC empirical search, as [42] proposed, is consistent.

Theorem 8.3.2 (Plug-in CMC estimator)**.** *Let* $(Y_1, \cdots, Y_N)$ *be* iid *according to* f_Y *on* $\mathbb{R}^d$*,* $\hat{f}_Y$ *be* f_Y *kernel estimator and* $\mathcal{B}_t$ *a* $\hat{f}_Y$ *level set.*

$$\forall y \in \mathbb{R}^d, \hat{f}_Y(y) = \frac{1}{h_N^d N} \sum_{i=1}^{N} K\left(\frac{y - Y_i}{h_N}\right) \qquad \forall t \in \left[0, \sup \hat{f}_Y\right], \mathcal{B}_t = \left\{y \in \mathbb{R}^d \,\middle|\, \hat{f}_Y(y) \geq t\right\} \tag{8.18}$$

$$\forall \alpha \in [0,1], q_\alpha \text{ is the } \alpha\text{-quantile of } \hat{f}_Y(Y) \qquad \overline{q}_\alpha \text{ is the CMC } \alpha\text{-quantile of } \hat{f}_Y(Y) \tag{8.19}$$

If

1. *The kernel function* K *is such that*
   (a) $\forall y \in \mathbb{R}^d, K(y) = K(\|y\|)$.
   (b) K *is continuously differentiable and monotone non increasing, and has compact support.*
2. *The density function* f_Y *is such that*
   (a) f_Y *is twice continuously differentiable.*
   (b) $f_Y(y) \to 0$ *as* $\|y\| \to \infty$.
   (c) $\forall t \in]0,1[, \inf_{f_Y^{-1}(t)} \|\nabla f_Y\| > 0$*, where* ∇f_Y *is the gradient of* f_Y.
3. *The bandwidth* h_N *is such that*
   (a) $N h_N^{d+4} (\ln(N))^2 \underset{N\to\infty}{\longrightarrow} 0$.
   (b) $\frac{N h_N^{d+2}}{\ln(N)} \underset{N\to\infty}{\longrightarrow} \infty$.

then

1. $\forall t \in]0,1[, \lambda\left(f_Y^{-1}([t-\epsilon, t-\epsilon])\right) \to 0$ *as* $\epsilon \to 0$.
2. $\forall \alpha \in]0,1[, \mathbb{P}\left[Y \in \mathcal{B}_{\overline{q}_{(1-\alpha)}}\right] \to \alpha$ *in probability.*
3. $\forall \alpha \in]0,1[, \lambda\left(\left(\mathbf{B}_{q_{(1-\alpha)}} \cap \mathcal{B}^c_{\overline{q}_{(1-\alpha)}}\right) \cup \left(\mathbf{B}^c_{q_{(1-\alpha)}} \cap \mathcal{B}_{\overline{q}_{(1-\alpha)}}\right)\right) \to 0$ *in probability.*

Within this framework

$$\left[h_N = \frac{1}{N^a}\right] \Rightarrow \left[\frac{d+3}{(d+2)(d+4)} < a < \frac{2d+3}{2(d+2)^2}\right] \tag{8.20}$$

and in particular

$$h_N = \frac{1}{N^a} \text{ with } a = \frac{1}{2}\left(\frac{d+3}{(d+2)(d+4)} + \frac{2d+3}{2(d+2)^2}\right) \tag{8.21}$$

is a lawful bandwidth choice.

This theorem seemed just what we needed to estimate the safety zone with respect to the booster fallout. It was hence our second try at estimating them, as detailed in Section 8.4. However, it would not be our last.

8.4 MVS estimation *via* CMC

As quantile is the not the good tool to summarize a spatial distribution, we decided to switch to Minimum Volume Density Level Set as detailed in Subsection 8.3.2. As Theorem 8.3.2 can be translated into a CMC estimation algorithm, we did our second attempt at estimating the booster fallout safety zones. However, we were dealing with extreme levels of safety. How could the CMC based algorithm able to handle them ?

We used the plug-in estimator algorithm according to Theorem 8.3.2 – subsection 8.4.1– and assessed whether it could be used straight away to estimate the MVS, including those of extreme levels – subsection 8.4.2.1–.

8.4.1 CMC plug-in MVS algorithm

The algorithm is straight forward. From a Y sample set, one builds a f_Y kernel estimation $\hat{f}_Y$, evaluates it over a well-chosen grid to find out how grid points divide up in between the empirical $\hat{f}_Y(Y)$ quantiles.

Algorithm 8.4.1 (CMC plug-in MVS algorithm)**.** *To estimate the Minimum Volume Density Level Set of $\mathbb{R}^d$ valued random variable $Y = \phi(X)$ which probability is α procede as follows.*

1. *Sample $(X_1, \cdots, X_N)$ according to f_X in a* iid *fashion.*
2. *Evaluate $\forall i \in \{1, \cdots, N\}$,$Y_i = \phi(X_i)$.*
3. *Construct f_Y kernel estimation $\hat{f}_Y$ as per equations (8.18) and (8.21):*

$$a = \frac{1}{2}\left(\frac{d+3}{(d+2)(d+4)} + \frac{2d+3}{2(d+2)^2}\right) \qquad h_N = \frac{1}{N^a} \tag{8.22}$$

$$\forall y \in \mathbb{R}^d, \qquad \hat{f}_Y(y) = \frac{1}{h_N^d N}\sum_{i=1}^{N} K\left(\frac{y - Y_i}{h_N}\right) \tag{8.23}$$

4. *Evaluate $\forall i \in \{1, \cdots, N\}$,$V_i = \hat{f}_Y(Y_i)$.*
5. *Estimate the $(1-\alpha)$-quantile of $f_Y(Y)$ with $V_{(\lfloor(1-\alpha)N\rfloor+1)}$.*
6. *Choose a $\mathbb{R}^d$ grid $\{y_1, \cdots, y_u\}$.*
7. *Evaluate $\forall i \in \{1, \cdots, u\}$,$v_i = \hat{f}_Y(y_i)$.*
8. *Approximate the α-Minimum Volume Density Level Set with $\left\{y_i \middle| v_i \geq V_{(\lfloor(1-\alpha)N\rfloor+1)}\right\}$.*

The trickiest part of this algorithm is checking the hypothesis of the underlying theorem.

8.4.2 Safety area CMC estimation

Regularity can be expected from f_Y because the booster falls according to classical physics laws. And for the same reason, f_Y can reasonably be assumed to have bounded support. We therefore could use Algorithm 8.4.1 in line with Theorem 8.3.2 to estimate the minimum volume sets whose safety levels are in A as intended in Subsection 8.1.2.

$$A = \left\{1 - 10^{-1}, 1 - 10^{-2}, 1 - 10^{-3}, 1 - 10^{-4}, 1 - 10^{-5}, 1 - 10^{-6}\right\} \tag{8.24}$$

8.4.2.1 Kernel density estimator settings

We used Epanechnikov kernels to build a density kernel approximation of f_Y. They are bounded but they are not *continuously* differentiable. We used them in the absence of another convenient choice. We chose the bandwidth according to Equation 8.21.

$$\forall (v,y) \in \mathbb{R}^2 \times \mathbb{R}^2, \qquad K_2(v,y,h) = \left(\frac{3}{4}\right)^2 \prod_{i=1}^{2} \left(1 - \left|\frac{v_i - y_i}{h_i}\right|\right) \mathbf{1}_{[0,1]}\left(\left|\frac{v_i - y_i}{h_i}\right|\right) \tag{8.25}$$

Then we estimated $f_Y(Y)$ quantiles.

8.4.2.2 $f_Y(Y)$ quantile estimation

Each safety level α_l is associated with the $(1-\alpha_l)$-quantile of $f_Y(Y)$. The quantile of highest level was consistently estimated. However, for the safety level of interest were extreme, so were the quantile levels. As it could be feared, CMC could not distinguish them as shown on Figure 8.5. The common value that the five quantiles of lowest levels had was the smallest a kernel center could have. At this point, we already knew the associated MVS would overlap. We wished to see them nonetheless.

8.4.2.3 Grid choice

We used the same deterministic rectangular grid $\mathbb{G}$ to represent the Minimum Volume Sets.

$$\left\{y \in \mathbb{R}^2 \,|\, f_Y(y)\right\} \supset \left[-\frac{12}{100}, \frac{3}{10}\right] \times \left[-\frac{3}{10}, \frac{3}{10}\right], \qquad \mathtt{y}_0 = \frac{1}{100}\begin{pmatrix} -12 \\ -30 \end{pmatrix} \tag{8.26}$$

$$\Delta_1 = \frac{1}{1000}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \Delta_2 = \frac{1}{1000}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{8.27}$$

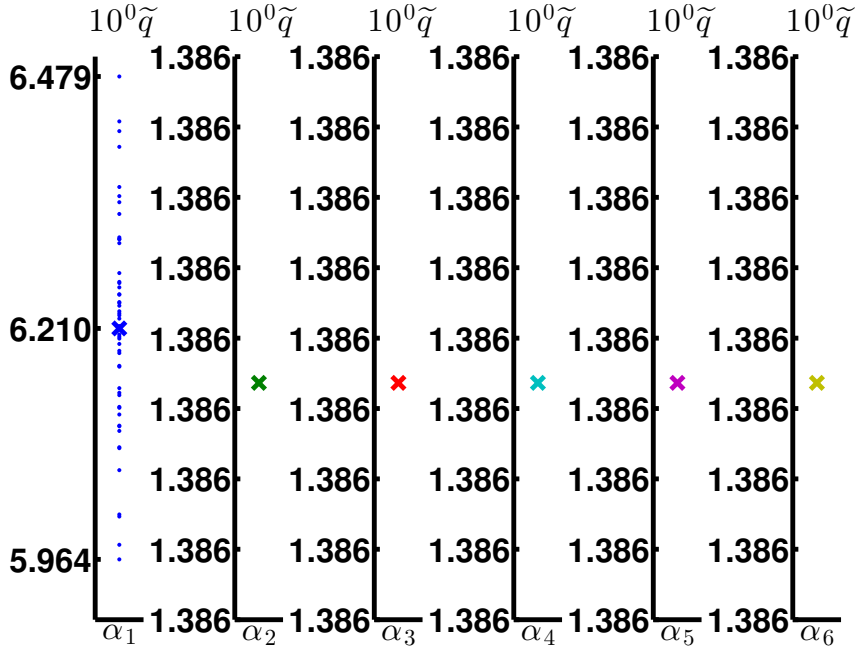
$$i \in \{1, \cdots, 600\} \qquad j \in \{1, \cdots, 420\} \tag{8.28}$$

$$\mathtt{y}_{ij} = \mathtt{y}_0 + i\Delta_1 + j\Delta_2 \qquad \mathbb{G} = \left\{\mathtt{y}_{ij}\right\} \tag{8.29}$$

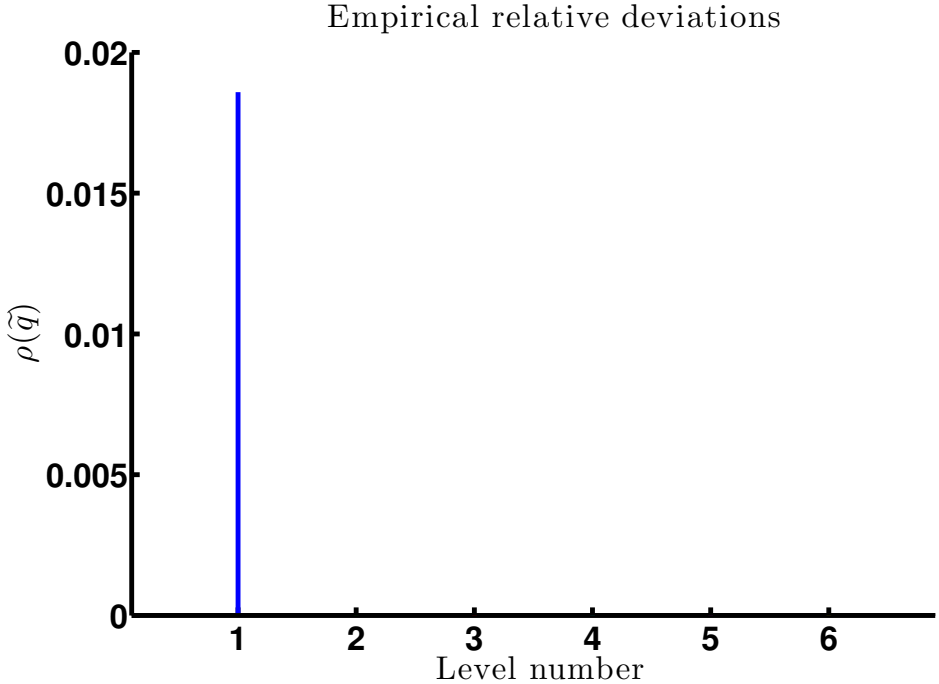
$\mathbb{G}$ was actually a *subset* of f_Y support chosen for it encompassed the bulk of the distribution and still not exceeded our calculation power. Ideally, one should use Algorithm 8.4.1 to estimate the support and take a slightly wider grid.

8.4.2.4 MVS as safety zones

We calculated f_Y approximation over $\mathbb{G}$ and used the quantiles to build the Minimum Volume Sets. Our first MVS estimation results can be seen Subfigure 8.6a. The first observation was that MVS moulded to the actual shape of the samples as expected. Unfortunately, the second observation is that the highest level MVS overlapped, as expected as well. Building low level MVS – Subfigure 8.6b – showed the overlapping was due to the levels being extreme and not to the algorithm.

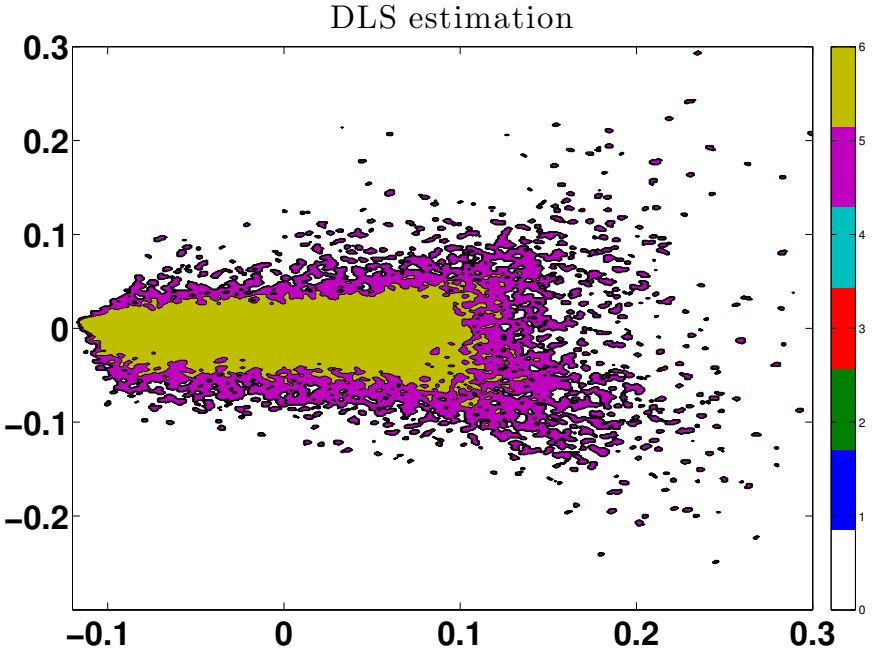


(a) Estimated quantiles.

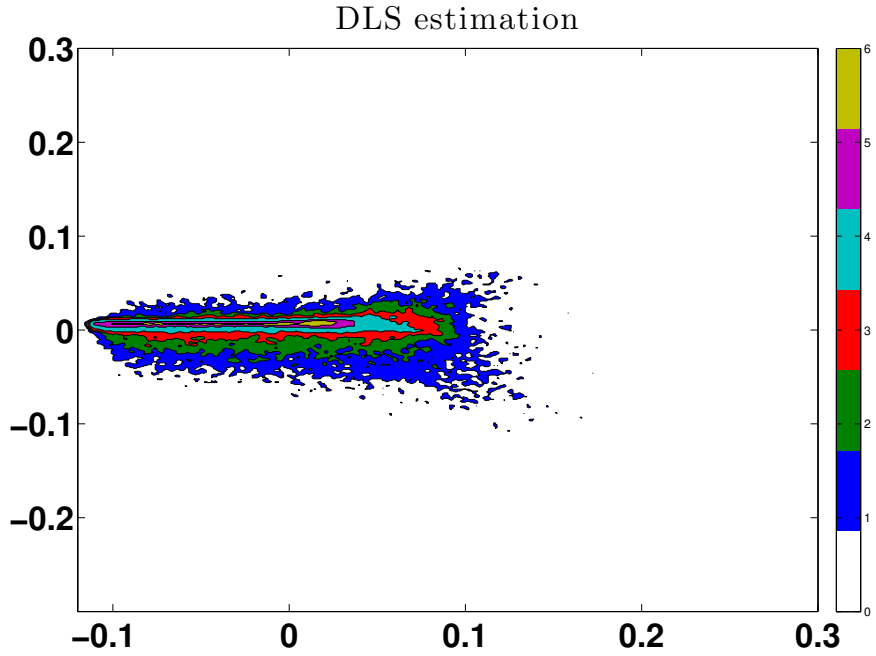


(b) Estimator empirical relative deviations.

Figure 8.5: 50 $f_Y(Y)$ extreme quantile estimations *via* CMC as described in Subsection 8.4.2. Levels are in $1 - A = \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$.



(a) An extreme level MVS estimation *via* CMC. Levels are in $A = \{1-10^{-1}, 1-10^{-2}, 1-10^{-3}, 1-10^{-4}, 1-10^{-5}, 1-10^{-6}\}$.



(b) A low level MVS estimation *via* CMC. Levels are in $B = \{\frac{2}{10}, \frac{4}{10}, \frac{6}{10}, \frac{7}{10}, \frac{8}{10}, \frac{9}{10}\}$.

Figure 8.6: Two safety area (MVS/DLS) estimations *via* CMC as described in Subsection 8.4.2.

8.4.2.5 Outcome

Minimum Volume Set is a good tool to define the safety areas because it moulds to the actual distribution and therefore make practical sense. Unfortunately, CMC is not able to deal with the level of safety we required. We had to look for an extreme level minimum volume set estimator.

8.5 Conclusion

We had been asked to estimate the safety distance way from the targeted landing position of a fall space shuttle booster. The α safety distance was defined as the α-quantile of the random distance from the target landing position to the actual landing position. We estimated it, even for extreme safety levels, using ANAIS and AST. However, we noticed the landing distribution was not isotropic and looked better tool than quantile to summarize the spatial distribution of a random variable.

To define the α safety area, we selected from literature the Minimum Volume Set (MVS). It is the smallest surface, with respect to Lebesgue measure, in with the booster falls with α probability. To be safe with probability α, one has to be *outside* it. This made practical sense. We estimated extreme level MVS with the plug-in CMC estimator. Though the MVS geometry was adapted to the actual landing points distribution, said estimator was unable to distinguish extreme level MVS just CMC cannot distinguish extreme level quantiles.

When comparing rare event dedicated techniques in Part II, we had grown some experience of extreme quantile estimation. Maybe there was a way to capitalize on them to estimate extreme Minimum Volume Sets. This attempt is described in Chapter 9.

Chapter 9

Safety area estimation *via* Extreme Minimum Volume Set

A MVS estimation requires a quantile estimation. An extreme MVS estimation requires an extreme quantile estimation. And the later can be better done through ANAIS or AST than CMC. Besides designing associated extreme MVS estimation algorithms[1] in Section 9.1 and Section 9.2 , we decided to compare their results in Section 9.3.

9.1 ANAIS extreme MVS estimation

We modified the CMC MVS plug-in estimator Algorithm 8.4.1 so that it benefited from ANAIS in Subsection 9.1.1. Then, we estimated the MVS with it in Subsection 9.1.2.

9.1.1 ANAIS plug-in MVS algorithm

The main idea of the ANAIS Algorithm 5.1.1 is building up an Importance Sampling change of measure $f_Z^\dagger$ to estimate $Y = \phi(X)$ extreme quantiles as explained in Chapter 5. This change of measure happens in the *input* space. However, MVS estimation requires $V = f_Y(Y)$ quantiles: things happen in the *output* space. To estimate extreme MVS, we suggest to make an Importance Sampling kernel based density estimation $\hat{f}_V^\dagger$ of V pdf in the output space and to use it, in f_V stead, in order to estimate V extreme quantiles. Inopportunely, f_Y is unknown in practice. It will be approximated by its iteratively updated kernel IS approximation $\hat{f}_Y^\dagger$. We stop updating $f_Z^\dagger$, $\hat{f}_Y^\dagger$ and $\hat{f}_V^\dagger$ when ANAIS proposes a threshold that is above the IS estimation of the V quantile for the algorithm would otherwise spend its estimation power on a useless quantity. Then, we use the remaining simulation budget, if any, to put a final touch to the density estimators updating them one last time to better estimate the desired quantile. Eventually, the MVS is estimated *via* a grid as done in CMC Algorithm 8.4.1. The following algorithm structures these ideas.

Algorithm 9.1.1 (ANAIS plug-in MVS algorithm)**.** *To estimate the α probability Minimum Volume Density Level Set $\mathbf{S}_\alpha$ of the random variable $Y = \phi(X)$, where $Y \in \mathbb{R}^d$ and $X \in \mathbb{R}^p$, with simulation budget N, proceed as follows.*

[1] The work presented here goes further than [66] and [70], and will be submitted soon.

1. *Set :*

$$a = \frac{1}{2}\left(\frac{d+3}{(d+2)(d+4)} + \frac{2d+3}{2(d+2)^2}\right) \quad k = 0 \quad f_Z^{\dagger,k} = f_X \quad \widehat{q}_{\dagger_k} = 0 \quad t_k = \infty. \tag{9.1}$$

2. *While* $\widehat{q}_{\dagger_k} < t_k$ *do:*

   (a) *Sample* $\left(Z_1^k, \cdots, Z_n^k\right)$ *according to* $f_Z^{\dagger,k}$ *in a* iid *fashion and evaluate*

$$\forall i \in \{1, \cdots, n\}, \qquad Y_i^k = \phi\left(Z_i^k\right) \qquad \mathrm{w}_i^k = \frac{f_X}{f_Z^{\dagger,k}}\left(Z_i^k\right). \tag{9.2}$$

   (b) *Approximate* f_Y *with:*

$$h_Y^k = \frac{1}{(n(k+1))^a} \qquad \forall y \in \mathbb{R}^d, \qquad \hat{f}_Y^{\dagger,k}(y) = \frac{\sum_{j=0}^k \sum_{i=1}^n \mathrm{w}_i^j K_Y\left(\frac{y - Y_i^k}{h_Y^k}\right)}{\left(h_Y^k\right)^d \sum_{j=0}^k \sum_{i=1}^n \mathrm{w}_i^j}. \tag{9.3}$$

   (c) *Approximate* $f_Y(Y)$ via $\hat{f}_Y^{\dagger,k}$*:*

$$\forall i \in \{1, \cdots, n\}, \qquad \forall j \in \{1, \cdots, k\}, \qquad V_i^j = \hat{f}_Y^{\dagger,k}\left(Y_i^j\right). \tag{9.4}$$

   (d) *Estimate the* $(1-\alpha)$*-quantile of* $f_Y(Y)$ *with estimator* $\widehat{F}_{3V}$ *from Definition 2.3.1.*

$$\widehat{q}_{\dagger_\kappa} = \inf\left\{ v \in \mathbb{R}^+ \,\middle|\, 1 - \alpha \leq \frac{\sum_{j=0}^k \sum_{i=1}^n \mathbf{1}_{\left\{v \geq V_i^j\right\}} \mathrm{w}_i^j}{\sum_{j=0}^k \sum_{i=1}^n \mathrm{w}_i^j} \right\}. \tag{9.5}$$

   (e) *Propose a new threshold:* $t_k = \max\left(\widehat{q}_{\dagger_k}, V^k_{(\lfloor(1-\alpha)N\rfloor+1)}\right)$.

   (f) *Build the new auxiliary sampling density:*

$$\text{Fix } h_Z^{k+1} > 0 \quad \forall z \in \mathbb{R}^p, \quad f_Z^{\dagger,k+1}(z) = \frac{\sum_{j=0}^k \sum_{i=1}^n \mathbf{1}_{\left\{t_k \geq V_i^j\right\}} \mathrm{w}_i^j K_Z\left(\frac{z - Z_i^k}{h_Z^{k+1}}\right)}{\left(h_Z^{k+1}\right)^d \sum_{j=0}^k \sum_{i=1}^n \mathbf{1}_{\left\{t_k \geq V_i^j\right\}} \mathrm{w}_i^j}. \tag{9.6}$$

   (g) *Set* $k = k + 1$.

3. *Set* $\kappa = k$ *and* $\aleph = N - n\kappa$

4. *Sample* $(Z_1^\kappa, \cdots, Z_\aleph^\kappa)$ *according to* $f_Z^{\dagger,\kappa}$ *in a* iid *fashion and evaluate.*

$$\forall l \in \{1, \cdots, \aleph\}, \qquad Y_l^\kappa = \phi\left(Z_i^\kappa\right) \qquad \mathrm{w}_l^\kappa = \frac{f_X}{f_Z^{\dagger,\kappa}}\left(Z_l^\kappa\right) \tag{9.7}$$

5. *Put the final touch to* f_Y *estimator.*

$$h_Y^\kappa = \frac{1}{N^a} \quad \forall y \in \mathbb{R}^d, \quad \hat{f}_Y^{\dagger,\kappa}(y) = \frac{\sum_{j=0}^k \sum_{i=1}^n \mathrm{w}_i^j K_Y\left(\frac{y - Y_i^k}{h_Y^\kappa}\right) + \sum_{l=1}^\aleph \mathrm{w}_l^\kappa K_Y\left(\frac{y - Y_l^\kappa}{h_Y^\kappa}\right)}{\left(h_Y^k\right)^d \left(\sum_{j=0}^k \sum_{i=1}^n \mathrm{w}_i^j + \sum_{l=1}^\aleph \mathrm{w}_l^\kappa\right)}. \tag{9.8}$$

6. *Update* $f_Y(Y)$ *approximation for the last time* via $\hat{f}_Y^{\dagger,\kappa}$.

$$\begin{array}{rcl} \forall i & \in & \{1,\cdots,n\} \\ \forall j & \in & \{1,\cdots,k\} \\ \forall l & \in & \{1,\cdots,\aleph\} \end{array} , \begin{array}{rcl} V_i^j & = & \hat{f}_Y^{\dagger,\kappa}\left(Y_i^j\right) \\ V_l^\kappa & = & \hat{f}_Y^{\dagger,\kappa}\left(Y_l^\kappa\right) \end{array} \tag{9.9}$$

7. *Eventually, estimate the* $(1-\alpha)$*-quantile of* $f_Y(Y)$ *with estimator* $\widehat{F}_{3V}$ *from Definition 2.3.1.*

$$\widehat{q_\dagger}_\kappa = \inf\left\{ v \in \mathbb{R}^+ \middle| 1-\alpha \leq \frac{\sum_{j=0}^{k}\sum_{i=1}^{n} \mathbf{1}_{\{v \geq V_i^j\}} \mathrm{w}_i^j + \sum_{l=1}^{\aleph} \mathbf{1}_{\{v \geq V_l^\kappa\}} \mathrm{w}_l^\kappa}{\sum_{j=0}^{k}\sum_{i=1}^{n} \mathrm{w}_i^j + \sum_{l=1}^{\aleph} \mathrm{w}_l^\kappa} \right\}. \tag{9.10}$$

8. *Choose a* $\mathbb{R}^d$ *grid* $\mathbb{G} = \{y_1,\cdots,y_u\}$.

9. *Evaluate* $\forall i \in \{1,\cdots,u\}, v_i = \hat{f}_Y^{\dagger,\kappa}(y_i)$.

10. *Approximate the* α*-Minimum Volume Density Level Set* $\mathbf{S}_\alpha$ *with* $\mathcal{S}_\alpha = \left\{ y_i \middle| v_i \geq \widehat{q_\dagger}_\kappa \right\}$.

This algorithm does deserve some theoretical deepening and a ANAIS counterpart of the handy CMC Theorem 8.3.2 would be very much appreciated. Nonetheless, we boldly moved on to numerical experiments.

9.1.2 MVS estimation *via* ANAIS

In the absence of theoretical results, we decided to be conservative and based our tuning on previous experiments. The main difference was that instead of f_X and f_Y, we used Importance Sampling kernel based counter-parts *i.e.* weights. The safety targets were the same as in Subsection 8.1.2, with respect to the same falling space shuttle booster.

$$A = \left\{1-10^{-1}, 1-10^{-2}, 1-10^{-3}, 1-10^{-4}, 1-10^{-5}, 1-10^{-6}\right\} \tag{9.11}$$

9.1.2.1 Setting the kernel densities

f_Z was tuned just as when estimating CEP *via* ANAIS in Subsection 8.2.3 and simulated *via* the rejection method. f_Y was tuned just as when estimating MVD *via* CMC in Subsection 8.4.2.

$$n = 1000 \qquad \kappa = 50 \qquad \widehat{\beta_\dagger} = 0.90 \qquad h_Z \textit{ via } \texttt{AMISE} \qquad h_Y \textit{ via Equation 8.21} \tag{9.12}$$

$$\forall (x,z) \in \mathbb{R}^6 \times \mathbb{R}^6, \quad \underline{K}_6^Z(x,z,h_Z) = \mathbf{1}_{\left([0,1]^6\right)^2}(x,z) \prod_{i=1}^{6} \left(1 - \left|\frac{x_i - z_i}{h_{Z,i}}\right|\right) \mathbf{1}_{[0,1]}\left(\left|\frac{x_i - z_i}{h_{Z,i}}\right|\right) \tag{9.13}$$

$$\forall (v,y) \in \mathbb{R}^2 \times \mathbb{R}^2, \quad K_2^Y(v,y,h_Y) = \left(\frac{3}{4}\right)^2 \prod_{i=1}^{2} \left(1 - \left|\frac{v_i - y_i}{h_Y}\right|\right) \mathbf{1}_{[0,1]}\left(\left|\frac{v_i - y_i}{h_Y}\right|\right) \tag{9.14}$$

We could then estimate the quantiles of $f_Y(Y)$.

9.1.2.2 $f_Y(Y)$ quantile estimates

Contrarily to Subsubsection 8.4.2.2 and as Figure 9.1 showed, the estimator had distinguished all the $f_Y(Y)$ quantiles. As in Subsection 8.2.3, the most extreme level quantile had the lowest variance because it was the one we targeted in the first place, the other ones being by-products. We could expect a good news from the MVS estimation.

9.1.2.3 MVS as safety zones

We used the same deterministic rectangular grid $\mathbb{G}$ as Subsubsection 8.4.2.3 to represent the MVS. A ANAIS estimation can be seen on Figure 9.2. All MVS were estimated and distinguished unlike when using CMC as can be seen on Figure 8.6. We had reached our objective! We then tried to do the same adapting AST to extreme MVS estimation.

9.2 AST extreme MVS estimation

We followed the same road map as when adapting ANAIS to extreme MVS estimation in Section 9.1: first, in Subsection 9.2.1, we adapted Algorithm 3.2.4 and then, in Subsection 9.2.2, used the algorithm to estimate the extreme MVS.

9.2.1 AST plug-in MVS algorithm

The AST plug-in estimator is based on the very same idea as the ANAIS one. The only difference is that intend of using ANAIS to build up the auxiliary sampling density, we use AST. The rest of the algorithm is virtually unchanged. Only weights require a specific comment before stating the algorithm itself.

9.2.1.1 How to weight discarded points?

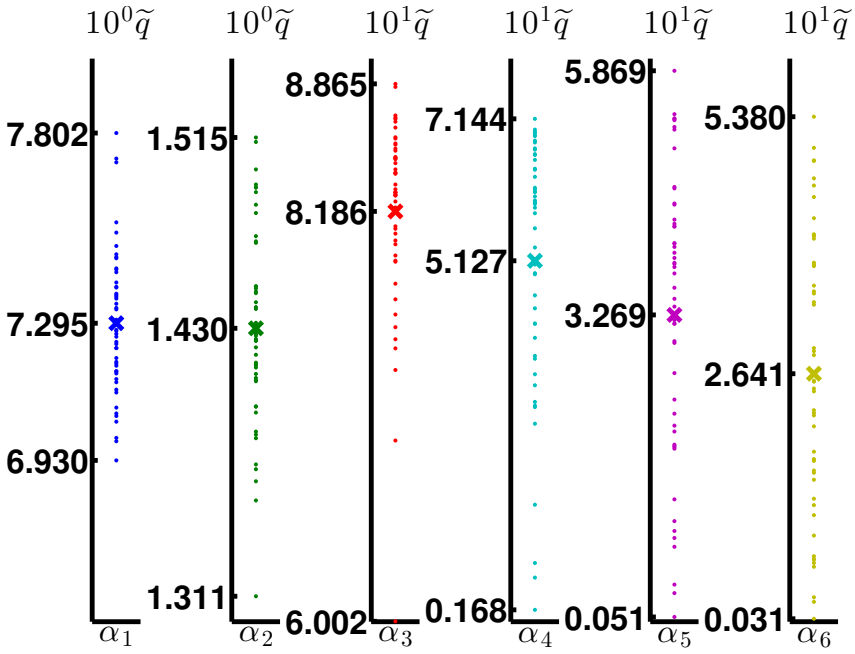
During AST quantile estimator Algorithm 3.2.4, Algorithm 3.1.1 procedure is iterated several times in order to produce *iid* points distributed according to a conditional distribution. Only the points that are actually distributed according to the conditional distribution of interest are kept and all user are discarded. Besides, only the final generation of them proceed to the next step because they are deemed *iid*. This Algorithm 3.1.1 procedure is used in the coming algorithm as well but more points are kept as the procedure is iterated. Actually, in the *input* space, only the last generation is kept so as in line with Algorithm 3.2.4. But, in the *output* space, all points that are distributed according to the conditional distribution are kept, whatever their correlation, in order to build the importance sampling estimator $\hat{f}_Y$ of the f_Y.

Selected points are distributed according to the conditional distribution

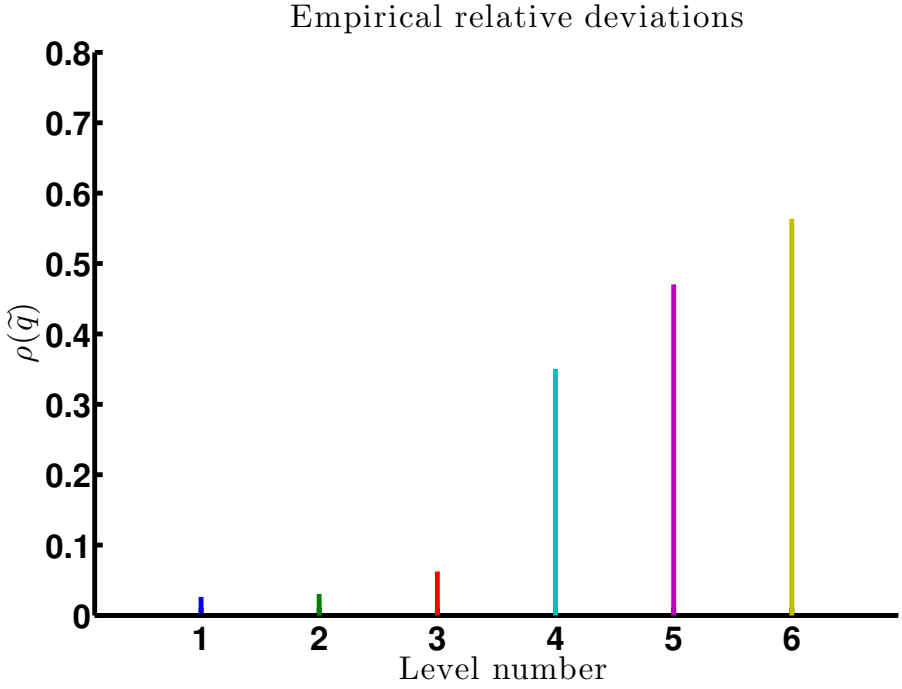
$$\forall z \in \mathbb{X}, f_{X|\mathbf{A}}(z) = \frac{\mathbf{1}_{\mathbf{A}}(z \in \mathbf{A}) f_X(z)}{\int_{\mathbf{A}} f_X(v)\, \mathrm{d}v} \tag{9.15}$$

and have a weight equal to $\mathrm{w} = \int_{\mathbf{A}} f_X$ so as to balance the conditioning.

$$Z \sim f_{X|\mathbf{A}} \Rightarrow \forall \mathbf{A}' \in \mathcal{X}, \mathbb{E}\left[\mathrm{w} \underset{\mathbf{A}}{\mathbf{1}}(Z \in \mathbf{A}')\right] = \int_{\mathbf{A}' \cap \mathbf{A}} f_X(x)\, \mathrm{d}x \tag{9.16}$$



(a) Estimated quantiles.



(b) Estimator empirical relative deviations.

Figure 9.1: 50 $f_Y(Y)$ extreme quantile estimations *via* ANAIS as described in Subsection 9.1.2. Levels are in $1 - A = \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$.

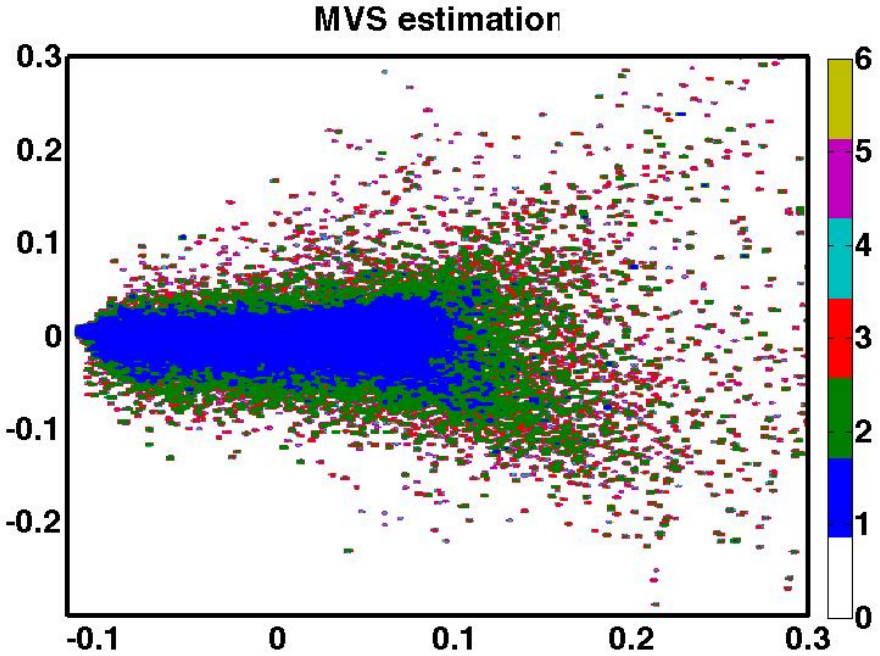


Figure 9.2: Safety area (MVS) estimations *via* ANAIS as described in Subsection 9.1.2. Levels are in $A = \{1-10^{-1}, 1-10^{-2}, 1-10^{-3}, 1-10^{-4}, 1-10^{-5}, 1-10^{-6}\}$. See Figure 8.6 page 127 for comparison with CMC.

We wanted to use discarded points to build $\hat{f}_Y$ as well. But we could not find the normalizing constant.

$$Z \sim \int_{\mathbf{A}} f_{X|\mathbf{A}}(v)\, \mathrm{M}(v, \mathrm{d}z)\, \mathrm{d}v \qquad \int_{\mathbb{X}} \left\{ \int_{\mathbf{A}} f_{X|\mathbf{A}}(v)\, \mathrm{M}(v, \mathrm{d}z)\, \mathrm{d}v \right\} \mathrm{d}z = ? \tag{9.17}$$

Without it, it was impossible to weight discarded point and they were therefore discarded when building $\hat{f}_Y$. Had we had this weight, we would have used them. To nonetheless mark their future use in the AST plug-in estimator Algorithm 9.2.1, they appear at item 5c but with a zero weight.

9.2.1.2 The algorithm

The plug-in AST estimator algorithm Algorithm 9.2.1 comes with the same requirements as the AST extreme quantile estimator Algorithm 3.2.4 plus the f_Y approximation constraints. For the later, in the absence of an adapted theory, we stuck to Theorem 8.3.2.

Algorithm 9.2.1 (AST plug-in MVS algorithm)**.** *To estimate the α probability Minimum Volume Density Level Set $\mathbf{S}_\alpha$ of the random variable $Y = \phi(X)$, where $Y \in \mathbb{R}^d$ and $X \in \mathbb{R}^p$, with simulation budget N, proceed as follows.*

1. *Set* $k = 0$ *and:*

$$a = \frac{1}{2}\left(\frac{d+3}{(d+2)(d+4)} + \frac{2d+3}{2(d+2)^2}\right) \qquad \begin{array}{c}\kappa_{max} \in \mathbb{N}^* \\ n \in \mathbb{N}^*\end{array} \qquad \begin{array}{c}\omega \in \mathbb{N}^* \\ \widetilde{\beta} \in]\alpha, 1[\end{array} \qquad \begin{array}{c}\widetilde{q}^k = -\infty \\ \mathrm{M} \text{ is } f_X\text{-reversible}\end{array} \tag{9.18}$$

2. *Generate the starting sample set.*

$$\forall i \in \{1, \cdots, n\} \qquad X_i^k \overset{\text{iid}}{\sim} f_X \qquad Y_i^k = \phi\left(X_i^k\right) \tag{9.19}$$

3. *Build the initial approximation of* f_Y*.*

$$h_Y^k = \frac{1}{(n(k+1))^a} \qquad \forall y \in \mathbb{R}^d, \qquad \hat{f}_Y^k(y) = \frac{\sum_{i=1}^n K_Y\left(\frac{y - Y_i^k}{h_Y^k}\right)}{\left(h_Y^k\right)^d n}. \tag{9.20}$$

4. *Build the first intermediate threshold as* $V = \hat{f}_Y^k(Y)$ *empirical of level*[2] $\widetilde{\beta}$*.*

$$\forall i \in \{1, \cdots, n\} \qquad V_i^k = \hat{f}_Y^k\left(Y_i^k\right) \qquad S_k = V^k_{(\lfloor n\widetilde{\beta} \rfloor)} \tag{9.21}$$

5. *While* $\widetilde{q}^k < S_k$ *and* $k \leq \kappa_{max}$*, do as follows.*

   (a) *Replace* $\left\{X_i^k \middle| V_i^k > S_k\right\}$ *points with points uniformly selected with replacement within* $\left\{X_i^k \middle| V_i^k \leq S_k\right\}$ *and call the sample set* $\left(Z_1^{k,0}, \cdots, Z_n^{k,0}\right)$*.*

   (b) *Set* $j = 0$ *and do as follows until* $j = \omega$*:*

      i. *Generate the proposal set* $\left(Z'^{k,j}_1, \cdots, Z'^{k,j}_n\right)$ *such that* $Z'^{k,j}_i \sim \mathrm{M}\left(Z_i^{k,j}, \mathrm{d}z'^{k,j}_i\right)$*.*

      ii. *Form* $\left(Z_1^{k,j+1}, \cdots, Z_n^{k,j+1}\right)$ *such that* $Z_i^{k,j+1} = \begin{cases} Z'^{k,j}_i & \text{if } \hat{f}_Y^k\left(Z'^{k,j}_i\right) \leq S_k \\ Z_i^{k,j} & \text{otherwise} \end{cases}$*.*

      iii. *Increment* j *by one:* $j = j + 1$*.*

   (c) *Weight all points*[3] *and use them to build a new* f_Y *approximation.*

$$\forall (i,j) \in \begin{pmatrix} 1 \\ \vdots \\ n \end{pmatrix} \times \begin{pmatrix} 1 \\ \vdots \\ \omega \end{pmatrix} \qquad \mathrm{w}_i^{k,j} = \begin{cases} \widetilde{\beta}^k & \text{if } \hat{f}_Y^k\left(Z'^{k,j}_i\right) \leq S_k \\ 0 & \text{otherwise} \end{cases} \tag{9.22}$$

$$\frac{1}{h_Y^k} = \left(\sum_{l=0,i=1,j=1}^{k,n,\omega} \mathbf{1}_{\left\{\hat{f}_Y^k\left(Z'^{k,j}_i\right) \leq S_k\right\}}\right)^a \qquad \mathrm{W}^k = \sum_{l=0,i=1,j=1}^{k,n,\omega} \mathrm{w}_i^{l,j} \tag{9.23}$$

$$\forall y \in \mathbb{R}^d, \qquad \hat{f}_Y^{k+1}(y) = \frac{\sum_{l=0,i=1,j=1}^{k,n,\omega} \mathrm{w}_i^{k,j} K_Y\left(\frac{y - Z'^{k,j}_i}{h_Y^{k+1}}\right)}{\left(h_Y^{k+1}\right)^d \mathrm{W}^k} \tag{9.24}$$

[2] For clarity sake and without loss of generality, we assume $\left\lfloor n\widetilde{\beta} \right\rfloor \geq 1$.

[3] Read Subsection 9.2.1 for a discussion about weighting.

(d) Use the last generation as if iid *like* f_X *conditionally to* $\{V \leq S_k\}$.

$$\forall i \in \{1, \cdots, n\} \qquad X_i^{k+1} = Z_i^{k,\omega} \qquad Y_i^{k+1} = \phi\left(X_i^{k+1}\right) \tag{9.25}$$

$$V_i^{k+1} = \hat{f}_Y^{k+1}\left(Y_i^{k+1}\right) \qquad S_{k+1} = V_{(\lfloor n\widetilde{\beta} \rfloor)}^{k+1} \tag{9.26}$$

(e) Increment k *by one:* $k = k + 1$.

(f) Estimate $(1-\alpha)$*-quantile of* $f_Y(Y)$ *with estimator* $\widehat{F}_{3V}$ *from Definition 2.3.1.*

$$\widetilde{q}^k = \inf\left\{ r \in \mathbb{R} \,\middle|\, \sum_{l=0,i=1,j=1}^{k,n,\omega} \mathrm{w}_i^{l,j} \underset{\left\{\hat{f}_Y^k\left(Z^{l,j}_i\right) \leq r\right\}}{\mathbf{1}} \geq (1-\alpha)\,\mathrm{W}^k \right\} \tag{9.27}$$

6. *Set* $\kappa = k$ *and choose a* $\mathbb{R}^d$ *grid* $\mathbb{G} = \{y_1, \cdots, y_u\}$.

7. *Evaluate* $\forall i \in \{1, \cdots, u\}, v_i = \hat{f}_Y^{\kappa}(y_i)$.

8. *Approximate the* α*-Minimum Volume Density Level Set* $\mathbf{S}_\alpha$ *with* $\mathcal{S}_\alpha = \{y_i \,|v_i \geq \widetilde{q}^{\kappa}\}$.

Though this algorithm does deserver some theoretical deepening, we moved directly to numerical simulation.

9.2.2 MVS estimation *via* AST

Again, we chose our settings according to previous experiments. The AST part of the algorithm was tuned according to Subsection 8.2.2 but with $\kappa_{max} = 21$, and the f_Y estimation part on Subsubsection 9.1.2.1. The levels of interests were the same.

9.2.2.1 $f_Y(Y)$ quantile estimates

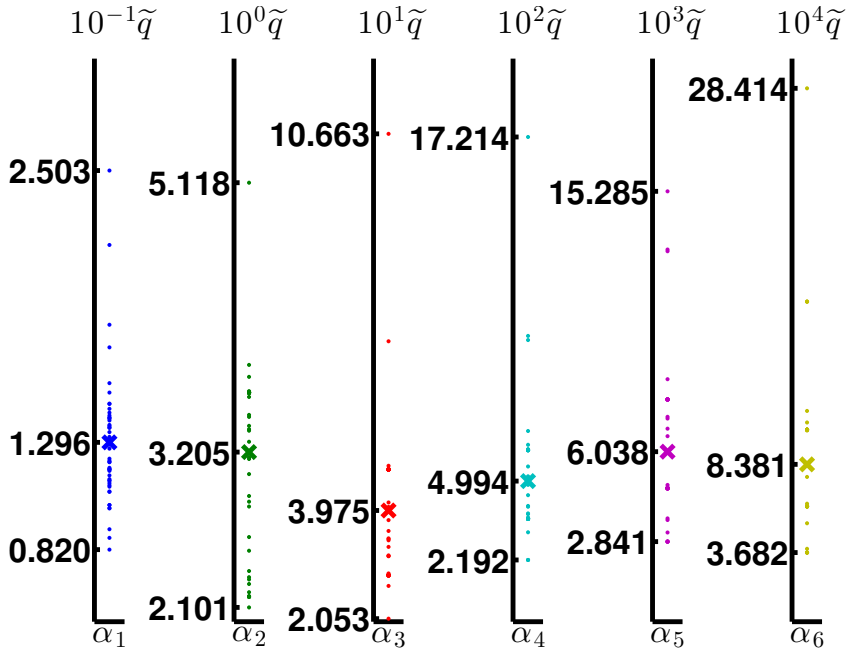
Figure 9.3 shows that all $f_Y(Y)$ quantiles were distinctly estimated but with a significant empirical relative variance because occasional but massive leaps aside the mean are possible. That was in line with Part II experiments and better than what CMC had given. Again, we could expect a good news from the MVS estimation.

9.2.2.2 MVS as safety zones

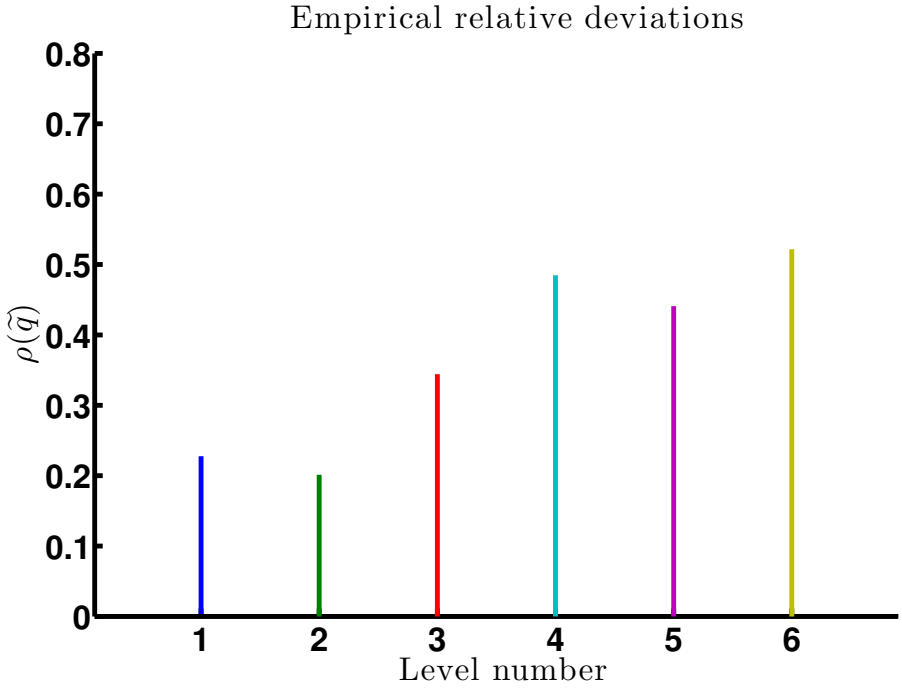
Again, we used the same deterministic rectangular grid $\mathbb{G}$ as Subsubsection 8.4.2.3 to represent the MVS and an AST estimation can be seen on Figure 9.4. All MVS were estimated and distinguished unlike when using CMC as can be seen on Figure 8.6. We had reach our objective but in an obviously different manner than with ANAIS as can be seen on Figure 8.6. So we decided to compare all three MVS estimators: CMC, ANAIS and AST.

9.3 MVS estimator comparison

The first comment was that the CMC estimations – Figure 8.6 – were dichromatic while those of ANAIS – Figure 9.2 – and AST – Figure 9.4 –were showed the whole color spectrum. This



(a) Estimated quantiles.



(b) Estimator empirical relative deviations.

Figure 9.3: 50 $f_Y(Y)$ extreme quantile estimations *via* AST as described in Subsection 9.2.2. Levels are in $1 - A = \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$.

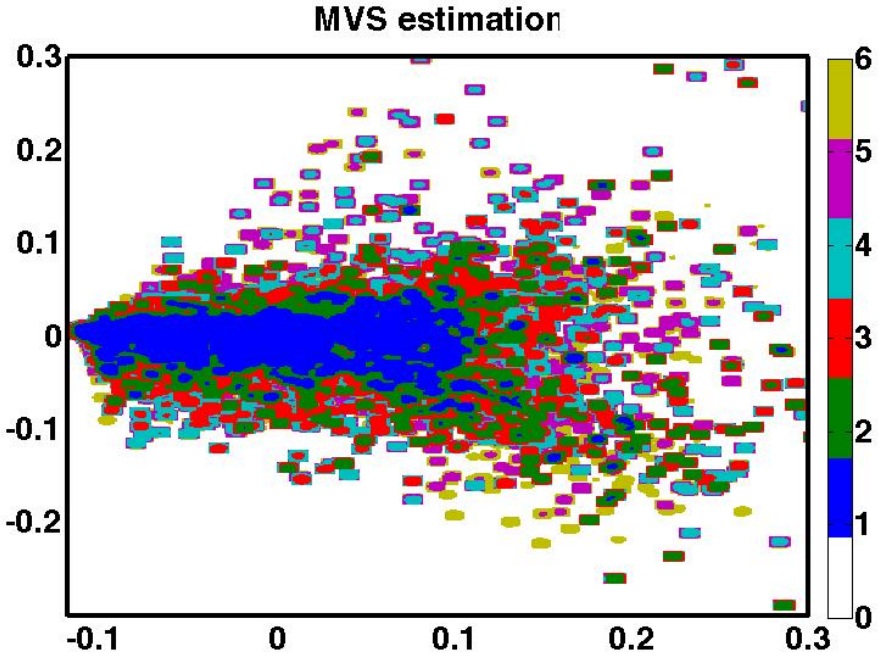


Figure 9.4: Safety area (MVS) estimations *via* AST as described in Subsection 9.2.2. Levels are in $A = \{1-10^{-1}, 1-10^{-2}, 1-10^{-3}, 1-10^{-4}, 1-10^{-5}, 1-10^{-6}\}$. See Figure 8.6 page 127 for comparison with CMC.

visually translated the incapacity of the CMC plug-in MVS estimator to distinguish extreme $f_Y(Y)$ quantiles while ANAIS and AST could, as expected. However, there was more to be said. We compared CMC , ANAIS and AST with respect to two other aspects: their geometries – Subsection 9.3.1– and their variances – Subsection 9.3.2–.

9.3.1 Geometry

Though all MVS had a pointillist look, the size of the dots was much larger in the AST case than in the others. This is because less points are used to build the f_Y approximation with this technique: typically only 21440 out the 50000 generated points. This translated into larger bandwidth and was due to the weighting issue discussed in Subsubsection 9.2.1.1.
As for the MVS shape, it was schematically a collection of co-axial symmetrical cones such that the bigger angular spread, the higher the level. In the cones of most extreme levels, landings were rare so the approximations of f_Y yielded pikes that could not be smoothened. This shows more in the AST because of larger size of its dots. and accounts for the $f_Y(Y)$ quantile estimation dramatic differences between AST and both CMC and ANAIS as well.
We then moved on measuring the consistency of the estimators.

9.3.2 Variance

Defining the MVS estimator variance was very important to measure the reliability of the provided results. We based our definition on the grid approximating the surface of interest

and used it straight away on our practical purpose.

9.3.2.1 Definition

Given $\mathbb{G}$ node, and a safety level, could we tell for sure if it was in the associated MVS?

$$\forall \mathtt{y} \in \mathbb{G}, \qquad \forall \alpha \in A, \qquad \mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right] =? \tag{9.28}$$

Ideally the probability would be either 1 or 0 *i.e.* "yes" or "no". As a matter of fact, the hypothesis testing theory from statistics stood as a good way to decide in which MVS a given node was. But we went for something more visual. We used the variance of the underlying binomial law to quantify our uncertainty.

$$\underset{\mathtt{y}\in\mathtt{S}_\alpha}{\mathbf{1}} \sim \mathcal{B}\left(\mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right]\right), \qquad \mathbb{V}\left[\underset{\mathtt{y}\in\mathtt{S}_\alpha}{\mathbf{1}}\right] = \mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right]\left(1 - \mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right]\right) \tag{9.29}$$

This variance was zero when we could tell for sure whether $\mathtt{y}$ was in $\mathtt{S}_\alpha$ – $\mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right]$ is 1 or 0 –, grew with our cluelessness – $\left|\frac{1}{2} - \mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right]\right| \in \left]0, \frac{1}{2}\right[$ – until it was maximal as we were clueless *i.e.* $\mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right] = \frac{1}{2}$. We just had to use the m simulated $\mathtt{S}_\alpha$ to estimate $\mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right]$.

$$\mathbb{P}\left[\mathtt{y} \in \mathtt{S}_\alpha\right] \approx \frac{1}{m}\sum_{i=1}^{m} \underset{\mathtt{y}\in\mathtt{S}_\alpha^i}{\mathbf{1}} = \mathtt{p} \qquad \mathbb{V}\left[\mathtt{S}_\alpha, \mathbb{G}\right] = \left\{ \begin{pmatrix} \mathtt{y} \\ \mathtt{p}\left(1-\mathtt{p}\right) \end{pmatrix} \in \mathbb{R}^3 \,|\, \mathtt{y} \in \mathbb{G} \right\} \tag{9.30}$$

This is how we measured and represented the estimator variance.

9.3.2.2 Variance representation and comparison

Two ANAIS MVS variances can be seen on Figure 9.6. This highlights the MVS border as the part of the "red zone". Ideally, a perfect estimator variance would be all white. Until such an estimator exists, one can use this tool to visualize the uncertainty and use statistics to decide to which MVS a given node belongs to.

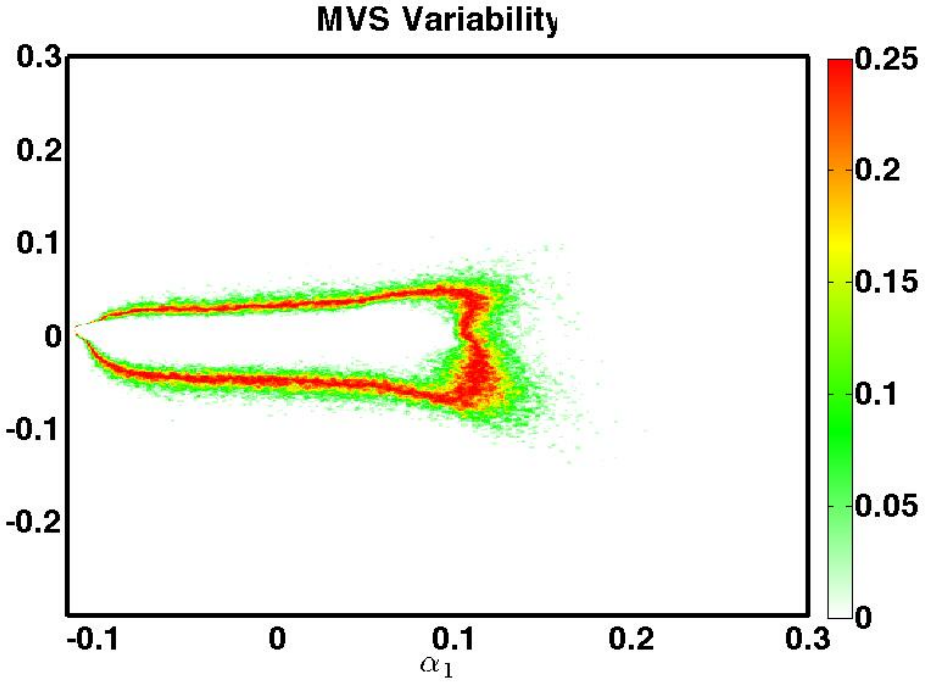
For both $\alpha = 1 - 10^{-1}$ and $\alpha = 1 - 10^{-6}$, variances ranked, in increasing order, CMC, ANAIS and AST. However, CMC underestimated $f_Y\left(Y\right)$ quantile of the highest levels. So the ANAIS plug-in estimator seemed the most reliable.

9.3.3 Outcome

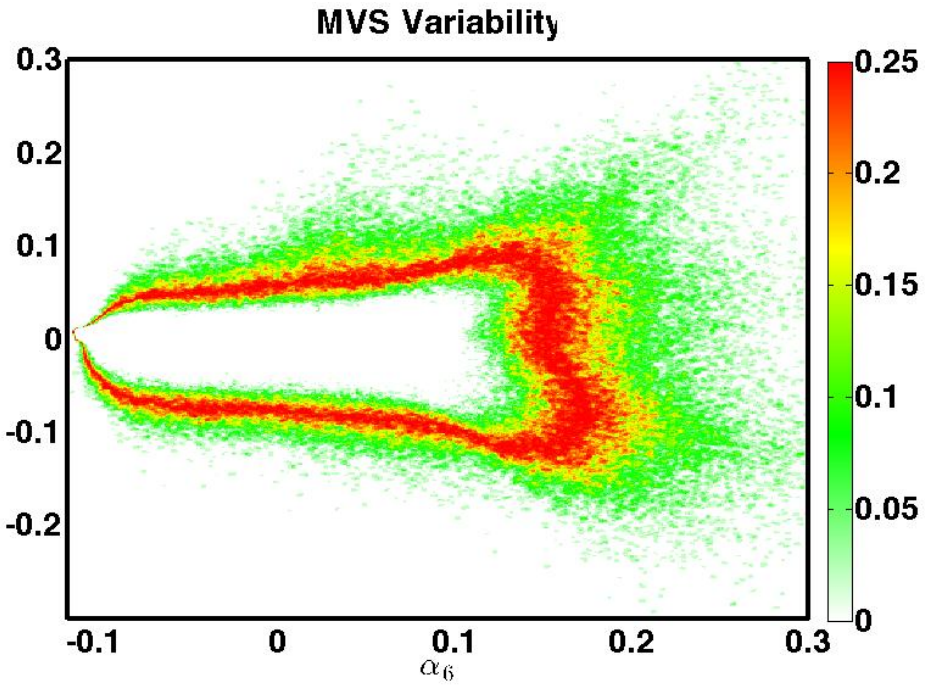
To replace CMC when estimating extreme level MVS we had to choice between ANAIS and AST. We chose ANAIS because it had a higher spatial resolution and smaller variance. Had we been able to weight all AST points, the outcome might have been different.

9.4 Conclusion

In Chapter 8, we had decided to use the Minimum Volume Sets CMC estimator to estimate the safety areas with respect to a spacecraft booster fallout. Given the simulation budget, CMC could not distinguished the most extreme safety level areas. We therefore adapted in this

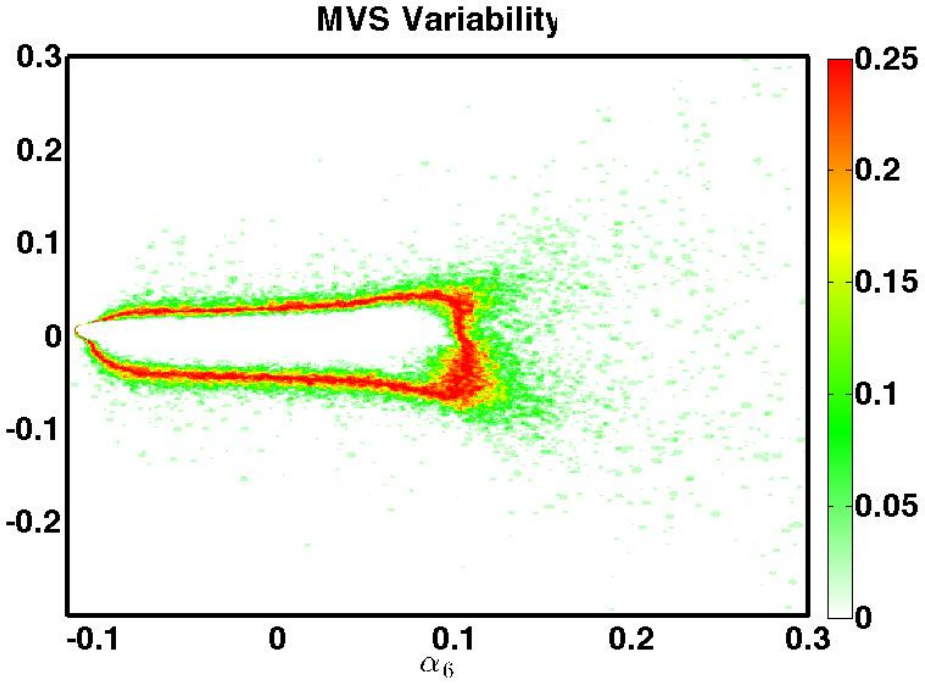


(a) $\alpha = 1 - 10^{-1}$

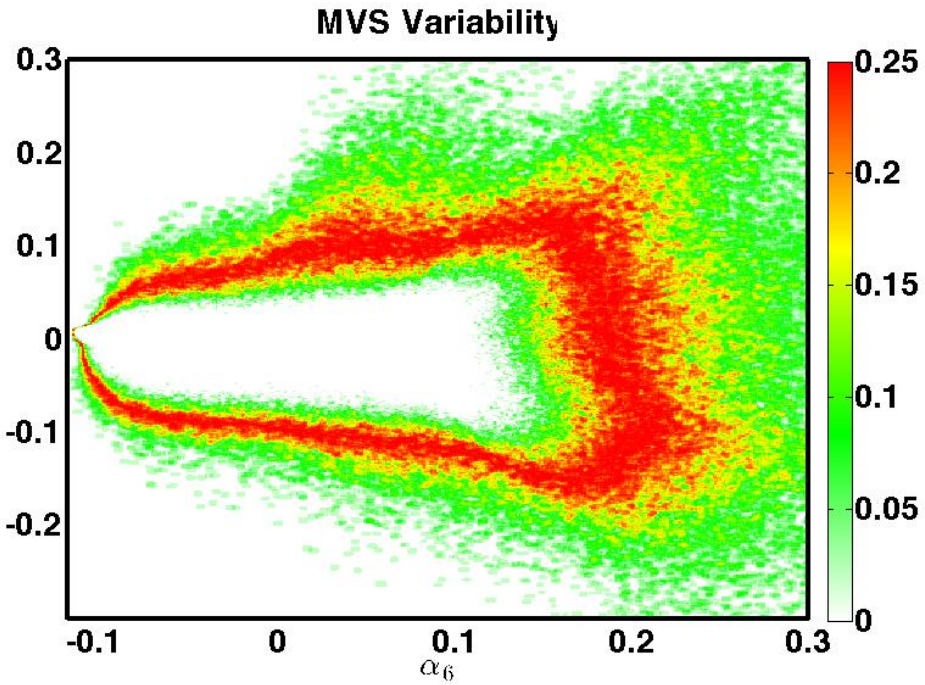


(b) $\alpha = 1 - 10^{-6}$

Figure 9.5: Safety area (MVS) CMC estimator variance.

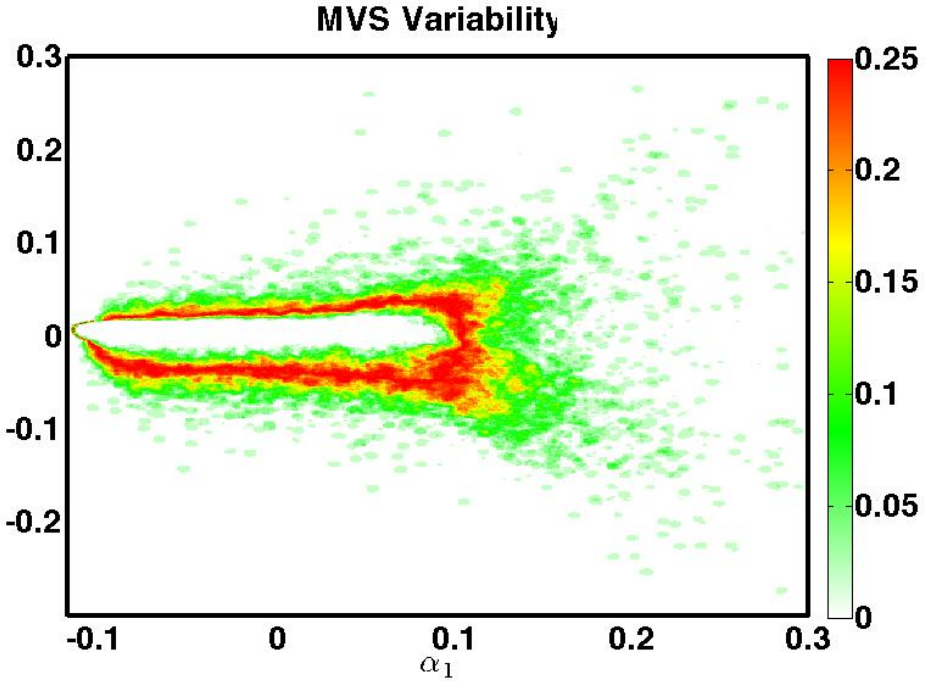


(a) $\alpha = 1 - 10^{-1}$

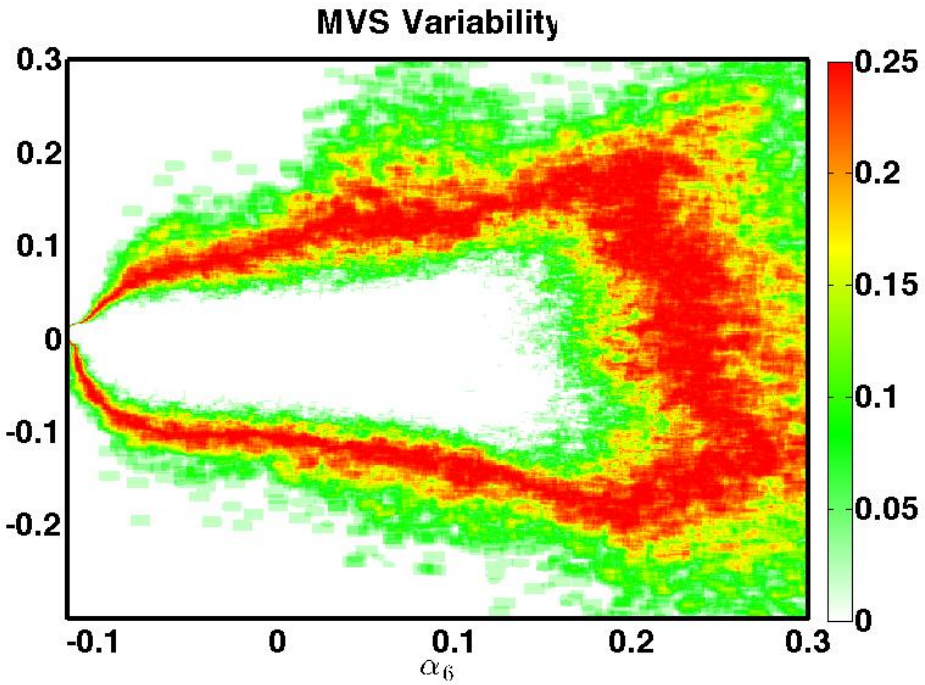


(b) $\alpha = 1 - 10^{-6}$

Figure 9.6: Safety area (MVS) ANAIS estimator variance.



(a) $\alpha = 1 - 10^{-1}$



(b) $\alpha = 1 - 10^{-6}$

Figure 9.7: Safety area (MVS) AST estimator variance.

chapter, ANAIS and AST into Minimum Volume Sets plug-in estimators that did distinguished them, though they needed theoretical deepening. To represent said estimator's consistency, we defined a three dimensional variance which roughly highlighted the MVS border. However, to decide in which safety area a given point is, combining the estimator with a statistical tool seemed a good idea. Unfortunately, we did not have the time to try it out.
AST was hindered by our incapability to weight part of the generated points whereas we could use them all with ANAIS. Therefore, until new theoretical results, we reckon ANAIS plug-in estimator is the best choice to estimate extreme level minimum volume sets because it has higher spatial resolution and smaller variance.

Conclusion

In Part III we engaged the extreme quantile estimation problem provided by the ONERA. It stemmed from the safety requirements with respect to the free fall of a spacecraft booster.

Chapter 8 experiments showed that CMC could not distinctly estimate the extreme quantiles with the available budget. Yet, as expected, AST and ANAIS could in a seemingly equivalent fashion. However, quantiles translated into circular safety zones that did not match the geometry of the actual landing distribution of the booster. So we looked through the literature for a better suited tool. Though the research of multidimensional extension to quantiles is an active statistics topics, we chose a probability tool. The Minimum Volume Set appeared as a reasonable choice because it moulds to the spatial distribution of the random variable of interest and allows to define the safety zones as density level sets that make practical sense. However, CMC is unable to estimate extreme level sets accurately because they are based on an extreme quantile.

Therefore, in Chapter 9, we adapted AST and ANAIS into plug-in minimum volume set estimators and compared them *via* both their geometry and their variances. Because the AST version is hindered by a theoretical question we were unable to solve, we advocate the use of the ANAIS estimator. However both require important theoretical deepening. Actually, both AST and ANAIS plain probability and quantile estimators are yet to be mastered theoretically. Besides, the weighted kernel density estimation underlying the MVS estimation is roughly adapted to the CMC MVS estimator and not to the ANAIS and AST versions.

In conclusion, we fulfilled the expectation of ONERA and proposed what we think is a better suited strategy in the safety area context. However, the ideas we came up with will be fully operational only when important theoretical questions will have been answered.

Conclusion: from practice to theory

In order to equip the ONERA with user-friendly tools able to handle rare events generated by a static random variable *via* a black box mapping, we had the following strategy. First, we swept the literature. Then we tested selected techniques with toy cases and eventually applied them to two cases of interest for the lab.

In Part I, we decided on a probability approach , because the lab produces the data itself, and focused on CE, NAIS and AST. The two first are Important Sampling techniques and the last one is based on splitting.

After trying the three methods on two crash tests in order to get practical experience in Part II, we proposed a NAIS variation, called ANAIS, designed to be of more convenient use. As ANAIS allows one to initiate the algorithm with the natural sampling density whereas NAIS requires one to straightaway know a auxiliary density which generates rare events, we decided to use ANAIS in NAIS stead. However, it was unable to estimate the probability of collision between satellites Iridium and Cosmos.

As for CE, though it performed best on the crash tests, it could not be carried out on this real test. The fact that the chosen parametric density family can not fit more than one connected component of interest in the input space can account for it. Choosing the auxiliary density in a adapted subset of the Exponential Family was considered, but sampling according to such distribution came as an issue. Besides, choosing this adapted subset can not be done before some information about the black box system is gathered. CE was therefore deemed inappropriate to this aim.

Unlike ANAIS and CE, AST performance did not deteriorate in the collision case with respect to the crash tests. AST estimations were accurate but yielded a two fifth empirical relative deviation. It came out as the best choice to estimate rare event probabilities though its tuning requires theoretical study. Furthermore, the systems of interest for the ONERA are mostly of high dimension and AST performance did not deteriorate in high dimensions. We therefore considered that AST would be a good candidate with respect to the lab objective with a rational and operational tuning rule. Besides, the lab often deals with random processes instead of random variables. Actually, systems are often dynamical and evolve as the time goes, facing a random event at each instant. The literature about this dynamic case is abundant as far as the Splitting Technique is concerned [18, 52, 54]. Focusing on the static case seemed *a posteriori* a good investment as it stood as a good practicing ground before the dynamical case. As a matter of facts, a PhD dealing with the dynamical case started as a sequel of this work.

The original objective of Part III was confronting AST and ANAIS to a realistic extreme quantile estimation problem. In this instance, the quantile was the distance from the forecasted landing position of a spacecraft booster that ensures safety with a given very close to one

probability. Both techniques estimated the quantile accurately and consistently. There is no clear cut answer about which estimator is the most suitable for the black box case. Actually, AST and ANAIS are still hindered by theoretical questions. A practical argument is that because of AST greater robustness to high dimensions, it should b used in such cases and ANAIS in low dimension.
Going back to the safety distance estimation problem, from a practical perspective, the induced circular safety zone was inappropriate because it did not match the actual landing surface geometry. We then selected a better suited tool from the spatial distribution literature, the Minimum Volume Set. The MVS does mold to the distribution geometry but the CMC plug-in estimator of MVS is not able to estimate MVS of extremely close to one probabilities with a limited budget. To this end, ANAIS and AST were adapted into plug-in estimators of extreme level MVS. An unsolved theoretical question hindered the AST version and resulted in its being able to only value half its simulation budget. The AST version did outperform the CMC estimator nonetheless. The ANAIS version yielded the best results but it needs theoretical scrutiny nonetheless.

Actually, the studies, though fruitful, were many times carried out until the border of their theoretical mathematical core and raised many theoretical questions. They can be organized as follows.

In the CE case, it would be interesting to be able to sample according to a polynomial exponential density once the number of modes in each dimension is known.

As for ANAIS, some ways of improvement are making the threshold sequence monotonous and finding a rational choice of density kernels and bandwidth.

As far as AST is concerned, tuning is a question as well. Besides, as the MH algorithm is an issue of its owns, having more explicit pairs of probability measure and reversible Markovian kernels would be interesting.

With respect to MVS estimation, the main question is weighted kernel density estimation.

From a practical point of view and it applies to all algorithms, knowing haw to share the simulation budget would be convenient.

Besides, it would be interesting to see how automatized tuning procedures such as the one presented in [58] can be of help.

This thesis allowed the ONERA to start readying for rare events by pointing out a theoretical tool adapted to handling the black box with random input case it faces regularly: AST. Its interest has been demonstrated through the Iridium-Comos case study and its dynamic form is now being tested as well. Besides, we led the ONERA to introduce the MVS in its random spatial dispersion quantification tools. The CMC MVS plug-in estimator being ill-suited for extreme level MVS estimation two dedicated algorithms have been constructed: ANAIS-MVS and AST-MVS.

This work was mostly geared toward applications because of ONERA's quest of operational tools, probability being not its core science. Nonetheless, all these tools require theoretical deepening.

Part IV

Appendix

Appendix A

Isoquantile curves

Various reliability or hedging problems boil down to quantile estimation. However, real life systems are usually multidimensional. In this case, a multidimensional counterpart of quantiles is required: this is the very purpose of Isoquantiles. We define the Isoquantile and propose a conewise approximation. A Crude Monte Carlo isoquantile estimation algorithm is provided with basic variance results. However, increasing safety standards created a need for extreme quantile estimation Crude Monte Carlo cannot fulfill. The Splitting Technique is therefore introduced with an isoquantile estimation algorithm to cope with extreme isoquantile estimation. Finally, we present some numerical results in a the bidimensional Gaussian case and a launcher impact safety zone estimation context as an exemple of the kind of results to expect in practice.

A.1 Introduction

When facing uncertainty, one might want a way to sum up what is known about it through some *deterministic* figures as this helps proportion a system to its uncertain environment. The quantile is used to represent the spread of the randomness at hand *on the line*. This problem is very crucial when hedging in finance, pricing an insurance or proving an industrial process' reliability.

However most systems are multidimensional and require their spread to be reprensented multidimensionally. We aim at catching a random system's spreading around a target through an envelope enclosing it with a given probability, going further than the benchmark one-dimensional answer, the Circular Error Probable (CEP) which disregarded completely the angular dependency. The scope of potential application includes, to say the least, finance, security design and obviously guidance. The definition of quantiles is therefore extended to higher dimensions through this paper's novelty, the Isoquantile Surfaces (IS), as defined and approximated through cones in section A.2.

An Isoquantile estimation algorithm via Crude Monte Carlo (CMC) is then given with a conewise confidence interval. However, to cope with CMC's unability to deal with extreme quantiles and therefore extreme isoquantiles, a rare event technique is required.

The Extreme Value Theory, as discribed in [27] for instance, has a statistical approach of the density's tail based on a given sample set. But, instead of working with a classically generated sample set to infer the extreme values'density, we decided to generate the sample

set so as to get directly the quantiles we are seeking. The Importance Sampling technique, as introduced in [10], suggests to sample according to another well-chosen probability law and then use weights to estimate the desired probability. Choosing the new probability law being an issue as it requires a lot of ex ante information -though this is tackled in [94] for exemple- and this choice being specific to one probability estimation, while we are looking for many quantiles, we looked for another way. The Splitting Technique, in its adaptive form, was the answer: it needs less a priori information and allows to estimate many isoquantiles at once. This technique is presented in section A.4, as well as how to estimate isoquantiles with it.

The bi-dimensional Gaussian toy case is then presented in section A.5 as a first chance to see the announced results and eventually an application to a launcher impact safety zone estimation problem is done in section A.6. It allows a comparison between CEP and IS and comes as a display case of the technique's estimating power.

A.2 Isoquantile Surfaces

Given an $\mathbb{R}$-valued random variable, a quantile is the capping threshold associated with a given probability. However, in practice most random variable are $\mathbb{R}^d$-valued with $d > 1$ and one wonders about *the spreading around a specified center*. Think about a golf player. Not only is he interested in how far from the hole his ball lands, but also in whether it does so too much on the right or in the front: he is looking for *a landing domaine*. This appears as the well-known target-to-impact spread characterisation problem engineers and scientists face on a daily basis.

Up to now, the benchmark answer was Circular Error Probable (CEP) : one would compute the quantiles of the random distance from the landing point to the aim and draw circles around it. This disregarded completely the angle dependency, wasting this wealth of yet available information. Isoquantiles benefit from it atop allowing extreme quantile estimation.

In the following exposÈ, we work in a very general context as a real life system's state space can be of any dimension. Besides, the target stands as origin as it is the focus of our attention and all coming results are target dependent. We will therefore equvalently use the cartesian coordinates and the hyperspherical coordinates, both originating for the target: a $\mathbb{R}^d$ point can be represented by both x and (r, θ), where $r \geq 0$ and $\theta \in \mathbb{A}^d = [0, \pi]^{d-2} \times [-\pi, \pi]$. This notations remind polar coordinates in order to keep the feeling the golf player's landing surface.

A.2.1 Definition of the Isoquantile

Let $X \sim \mu_X$ be a random variable on $\mathbb{R}^d$ such that

$$\begin{aligned} \mu_X(dx) &= f_X(x)dx && \text{(A.1)} \\ \mu_X(drd\theta) &= r^{d-1} f_X(r,\theta) g_d(\theta) drd\theta && \text{(A.2)} \\ \text{with } g_d(\theta) &= \sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2)\cdots\sin(\theta_{n-2}) && \text{(A.3)} \end{aligned}$$

$\vec{u}(\theta) = \vec{u}$ will denote the direction vector.

Given $\theta \in \mathbb{A}^d$, we will denote X^θ the $\mathbb{R}^+$-valued random variable defined as follows and name it X's behaviour in direction θ.

$$\mu_X^\theta(dr) = \frac{\mathbf{1}_{\mathbb{R}^+}(r) r^{d-1} f_X(r,\theta) g_d(\theta) dr}{\int_0^\infty \rho^{d-1} g_d(\theta) f_X(\rho,\theta) d\rho} \tag{A.4}$$

$$X^\theta \sim \mu_X^\theta(dr) \tag{A.5}$$

$\forall \alpha \in [0,1]$,X^θ's α-quantile will then be denoted x_α^θ

$$x_\alpha^\theta = \inf\{r \in \mathbb{R} | \int_{-\infty}^r \mu_X^\theta(du) \geq \alpha\} \tag{A.6}$$

$$= \sup\{r \in \mathbb{R} | \int_{-\infty}^r \mu_X^\theta(du) \leq \alpha\} \tag{A.7}$$

it captures X's distribution on $\mathbb{R}^+\vec{u}(\theta)$.

X's α level isoquantile or α-isoquantile is eventually defined by its polar equation as

$$\partial S_\alpha^X : \left| \begin{array}{ccc} \mathbb{A}^d & \to & \mathbb{R}^+ \\ \theta & \mapsto & x_\alpha^\theta \end{array} \right. \tag{A.8}$$

so that its interior S_α^X receives X's drawn with α probability as can be seen thanks to the following calculus.

$$\mu_X(S_\alpha^X) = \int_{\mathbb{A}^d} \{\int_0^{x_\alpha^\theta} \mathbf{1}_{\mathbb{R}^+}(r) r^{d-1} f_X(r,\theta) g_d(\theta) dr\} d\theta \tag{A.9}$$

$$= \alpha \int_{\mathbb{A}^d} \{\int_0^{\infty} \mathbf{1}_{\mathbb{R}^+}(r) r^{d-1} f_X(r,\theta) g_d(\theta) dr\} d\theta \tag{A.10}$$

$$= \alpha \tag{A.11}$$

S_α^X is said to be X's α-isoquantile domain. Its shape gives intel on how the system misses its target and can therefore be used to both correct it and protect the space at risk.

Though S_α^X is what we are actually looking for, we will focus on estimating ∂S_α^X.

A.2.2 Approximation through cones

As x_α^θ can not be estimated for all values of θ, a $\mathbb{A}^d$ partition is chosen and associated random variables are defined.

$$\mathbb{A}^d = \bigcup_{i=1}^q I_i \tag{A.12}$$

$$\mu_X^{I_i} = \frac{\mathbf{1}_{\mathbb{R}^+ \times I_i}(r,\theta) r^{d-1} g_d(\theta) f_X(r,\theta) dr d\theta}{\int_{I_i} \int_0^\infty \rho^{d-1} g_d(\eta) f_X(\rho,\eta) d\rho d\eta} \tag{A.13}$$

$$X^{I_i} \sim \mu_X^{I_i} \tag{A.14}$$

$$\mu_{\|X\|}^{I_i} = \frac{\mathbf{1}_{\mathbb{R}^+}(r) r^{d-1} \int_{I_i} g_d(\eta) f_X(r,\eta) d\eta dr}{\int_{I_i} \int_0^\infty \rho^{d-1} g_d(\eta) f_X(\rho,\eta) d\rho d\eta} \tag{A.15}$$

$$\|X^{I_i}\| \sim \mu_{\|X\|}^{I_i} \tag{A.16}$$

For a given $\alpha \in]0,1]$, one can then build ∂S_α^X's approximation $\widehat{\partial S_\alpha^X}$ thanks to the $\|x^{I_i}\|_\alpha$s i.e. the $\|X^{I_i}\|$s'α-quantiles.

$$\widehat{\partial S_\alpha^X} \quad : \quad \left| \begin{array}{ccc} \mathbb{A}^d & \to & \mathbb{R}^+ \\ \theta & \mapsto & \|x^{I_i}\|_\alpha \text{ if } \theta \in I_i \end{array} \right. \tag{A.17}$$

Besides, $\widehat{S_\alpha^X}$ encompasses $100\alpha\%$ of X's realisations too.

$$\begin{aligned} \mu_X(\widehat{S_\alpha^X}) &= \int_{\bigcup_{i=1}^q I_i} \{\int_0^{\|x^{I_i}\|_\alpha} \mathbf{1}_{\mathbb{R}^+}(r) r^{d-1} f_X(r,\theta) g_d(\theta) dr\} d\theta && \text{(A.18)} \\ &= \alpha \int_{\bigcup_{i=1}^q I_i} \{\int_0^{\infty} \mathbf{1}_{\mathbb{R}^+}(r) r^{d-1} f_X(r,\theta) g_d(\theta) dr\} d\theta && \text{(A.19)} \\ &= \alpha \int_{\mathbb{A}^d} \{\int_0^{\infty} \mathbf{1}_{\mathbb{R}^+}(r) r^{d-1} f_X(r,\theta) g_d(\theta) dr\} d\theta && \text{(A.20)} \\ &= \alpha && \text{(A.21)} \end{aligned}$$

$\widehat{\partial S_\alpha^X}$ will be a good approximation of ∂S_α^X if $\forall \theta \in \mathbb{A}^d$ such that $\theta \in [\theta - \epsilon, \theta + \epsilon] = I_\theta$, the quantiles of $\|X^{I_\theta}\|$ converge toward those of X^θ as $\max\{\epsilon_i\}$ decreases toward zero, where $\epsilon \in [0,\pi]^{d-1}$ and $[\theta - \epsilon, \theta + \epsilon]$ is to be understood componentwise.

As can be seen in [92], Lemma 21.2 p.305, if the cumulative distribution of $\|X^{I_\theta}\|$ converges towards X^θ's as the cone around the direction θ narrowes, the quantiles of $\|X^{I_\theta}\|$ converge toward those of X^θ.

As we cannot describe the whole class of density functions for wich this property, from now on denoted $(\mathcal{C})$, holds, we will give sufficient conditions so that the integral of $\mu_{\|X\|}^{I_i}$'s numerator on $[0,\rho]$ converges towards that of $\mu_X^{I_i}$'s numerator, for all ρ in $\overline{\mathbb{R}}^+$.

For instance, $(\mathcal{C})$ holds under the following -harsh- hypothesis:

$$\begin{aligned} \forall \theta \in \mathbb{A}^d, \forall M > 0, \exists \varepsilon \geq 0, \forall \eta \in [\theta - \varepsilon, \theta + \varepsilon], \forall r > 0, \\ |f_X(r,\eta) g_d(\eta) - f_X(r,\theta) g_d(\theta)| \leq M f_X(r,\theta) g_d(\theta) \end{aligned} \tag{A.22}$$

It basically means that in all directions, the density will depend as little as wanted on the angle, at any distance from the center, in a narrow enough cone. If this hold, then, $\forall \rho \in \overline{\mathbb{R}}_+$, one can write

$$\begin{aligned} |\int_0^\rho r^{d-1} \int_{I_\theta} f_X(r,\eta) g_d(\eta) d\eta dr - \int_0^\rho r^{d-1} \int_{I_\theta} f_X(r,\theta) g_d(\theta) d\eta dr| \\ \leq M \int_0^\rho r^{d-1} \int_{I_\theta} f_X(r,\theta) g_d(\theta) d\eta dr \end{aligned} \tag{A.23}$$

Furthermore, denoting $\lambda(I_\theta)$ the Lebesgue measure of I_θ,

$$\begin{aligned} & M \int_0^\rho r^{d-1} \int_{I_\theta} f_X(r,\theta) g_d(\theta) d\eta dr \\ &= M\lambda(I_\theta) \int_0^\rho r^{d-1} f_X(r,\theta) g_d(\theta) dr \\ &\leq 2M\pi^{d-1} \int_0^\rho r^{d-1} f_X(r,\theta) g_d(\theta) dr \\ &\leftarrow 0 \text{ as } M \to 0 \end{aligned} \tag{A.24}$$

This subsets is not empty as it contains all isotropic densities. However, even a very commun density as a non-reduced Gaussian is outside. A milder hypothesis such as

$$\sup_{\eta\in I_\theta} \left| \int_0^\infty \rho^{d-1} f_X(\rho,\eta)d\rho - \int_0^\infty \rho^{d-1} f_X(\rho,\theta)d\rho \right| \to 0 \text{ as } I_\theta \to \{\theta\} \tag{A.25}$$

includes any Gaussian. Evenutally, as the whole Isoquantile technique is -mainly- designed for real life systems, the concern is bounded support density functions. For any bounded support density continuous function, hypothesis A.23 holds: for all such functions ($\mathcal{C}$) holds. This is fairly enough for practical purposes.

Everything then boils down to estimating the $\|x^{I_i}\|_\alpha$.

A.3 Isoquantile Estimation via Crude Monte Carlo

Crude Monte Carlo (CMC) provides an easy way to estimate the $\|x^{I_i}\|_\alpha$. One just has to generate an X sample through CMC and for all i, estimate $\|x^{I_i}\|_\alpha$ as the empirical α-quantile of the norms of the samples falling the I_i basket.

The interval of confidence can besides be expressed in a very convenient fashion both globally through the central limit theorem and conewisely using the result given by [16]. If n X samples are generated, they will share themselves out into the cones according to binomial distributions: N_i, the number points in cone I_i out of the n generated, is a random variable distributed as $\mathcal{B}(n, \mathbb{P}[\Theta \in I_i])$, where the random variable X has hyperspherical coordinates (R, Θ). In a given cone I, if $\|X^I\|_{(1)}, \cdots, \|X^I\|_{(N_I)}$ are the N_I realisations falling in cone I sorted in ascending order and $\|x^I\|_\alpha$ is the sought quantile with $\alpha \in]0,1[$, then, for two integers i and j such that $1 \leq i < j \leq N_I$,

$$\mathbb{P}[\|X^I\|_{(i)} \leq \|x^I\|_\alpha \leq \|X^I\|_{(j)}] = \sum_{l=i}^{j-1} \frac{N_I!}{l!(N_I-l)!}\alpha^l(1-\alpha)^{N_I-l}$$

One can therefore choose the level of confidence he wishes and build up accordingly the smallest confidence interval possible.

However, it is a well-known issue that CMC is not reliable when looking for extremely high or low quantiles as it basically does not generate rare event unless one runs a huge number of simulations. The Splitting Technique (ST) is the rare events dedicated technique we will use to cope with this difficulty.

A.4 Isoquantile estimation via the AST

Assume this is the problem to be solved.

$$(\mathcal{P}_\alpha) \begin{cases} X & \text{is a random variable defined on } (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_X) \\ \mathbb{A}^d & = \bigcup_{i=1}^q I_i \\ \text{for } \alpha \in [0,1] & , \quad \widehat{\partial S_\alpha^X} \text{ is to be found.} \end{cases}$$

The A_is are this time an increasing sequence of surfaces defined by their outlines:

$$\partial A_i \ : \ \left| \begin{array}{ccc} I_i & \to & \mathbb{R}^+ \\ \theta & \mapsto & \sum_{j=1}^q \mathbf{1}_{I_j}(\theta) T_i^{I_j} \end{array} \right. \tag{A.26}$$

where $T_i^{I_j}$ is the ith threshold in the cone defined by I_j.

Given $\alpha \in [0,1]$, the following is set

- $N \in \mathbb{N}^*$ is the number of A_i to be generated
- $n \in \mathbb{N}^*$ is the number of points used to estimate all $T_i^{I_j}$ during one step
- $\forall i, \mathbb{P}[H \in A_{i+1} | H \in A_i] = p$ where $p \in [0,1]$

so that this relationship holds

$$\mathbb{P}[H \in S_\alpha^X] = p^N$$

and the following algorithm is ran.

1. Set $A_0 = \Omega$.
2. Given $i \in \mathbb{N}^*$ and associated A_i, proceed iteratively until $i = N$
   (a) Generate n points distributed according to $\mu_{X|A_i}$ through CMC using the μ_X-reversible kernel K.
   (b) Define, for every cone, $T_{i+1}^{I_j}$ as the empirical p-quantile of the norms of the samples falling the I_j basket.
   (c) Increment i by one.
3. Conclude $\widehat{S_\alpha^X} = A_N$.

All cones are explored at the same time, regardless to the probability that X falls in it. Rarely explored cones should be paid more attention to, so as to ensure the estimation quality.

Though a little more technical than CMC, AST will be shown to provide valuable results, especially when dealing with extreme quantiles. Besides, AST appears as a way to solve $(\mathcal{P}_\alpha)$ and solves $(\mathcal{P}_{\alpha'})$ $\forall \alpha' \in [0, \alpha]$ as well: if $p^{i+1} \leq \alpha' \leq p^i$, then $h_{\alpha'}$ is the $\frac{\alpha'}{p^i}$-quantile of $\mu_{H|A_i}$ and can therefore be estimated too.

A.5 Toy case: the bidimensional Gaussian

We will estimate the spreading of $X \sim \mathcal{N}(0_2, I_2)$ around the origin for different values of α. We know the results to be circles whose radii are given by Rayleigh's law's quantiles. These quantiles can be seen in table A.1.

α	$1-0.5$	$1-10^{-1}$	$1-10^{-2}$	$1-10^{-4}$	$1-10^{-6}$	$1-10^{-8}$
h_α	1.1774	2.1460	3.0349	4.2919	5.2565	6.0697

Table A.1: Rayleigh's law theoretical quantiles for some levels.

The estimations were done in the same conditions for CMC and AST.

- $]-\pi, \pi]$ was divided in sixty pieces of length $2\pi/60$.

- $5 \cdot 10^5$ points were generated 100 times.

The continuous lines are the average estimates and the dashed line represent the standard deviations.

Figure A.1a shows the isoquantiles estimated with CMC. They are circular as expected with low deviations. However, only the first 4 have an accurate radius as the last 2 are clearly underestimated and merged with the fouth isoquantile.

Figure A.1b shows the isoquantiles estimated with AST. These parameters were set:

- $X \sim \mathcal{N}(0, I_2)$.
- $K(X, dy) \sim \mathcal{N}(\frac{X}{\sqrt{1+.2^2}}, \frac{.2^2}{1+.2^2} I_2)$.
- $n = 5 \cdot 10^4$ and $p = 0.75$ and hence $N = 14$.

The isoquantile-surfaces are distinct and have small variance. The radii are accurate up to the last one, whose level is $1 - 10^{-8}$, which is underestimated by less than 10%.

We see that AST can estimate accurately higher level isoquantile curves than CMC with the same number of points and low variance.

A.6 Application

We are now addressing a real case isoquantile estimation for a deterministic black box mapping with random inputs. The reader can think of the system facing uncertainty represented as anyone of the many different cases taken from various domains fitting in this framework: in reality, application of these techniques are numerous.

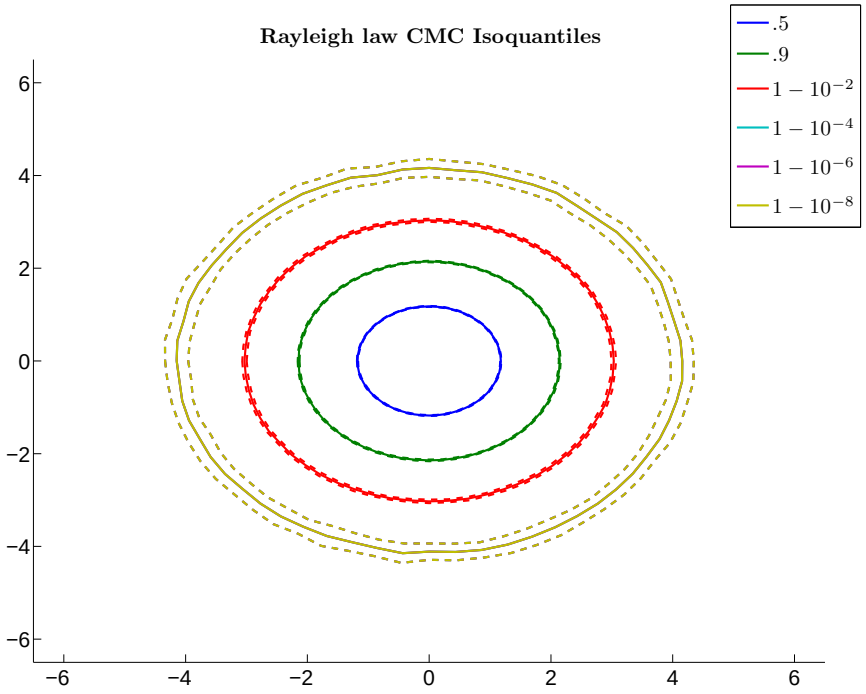
In our situation, it was a launcher impact safety zone estimation problem. After propelling its load, the launcher separates from it and should fall onto the ground without hitting anything valuable. Given the angular orientation, location, speed, weight and slope of the launcher plus the wind strength, we had to represent the spread of the impact location to demonstrate the system's safety.

We will deal with the following real case ($\mathcal{R}$).

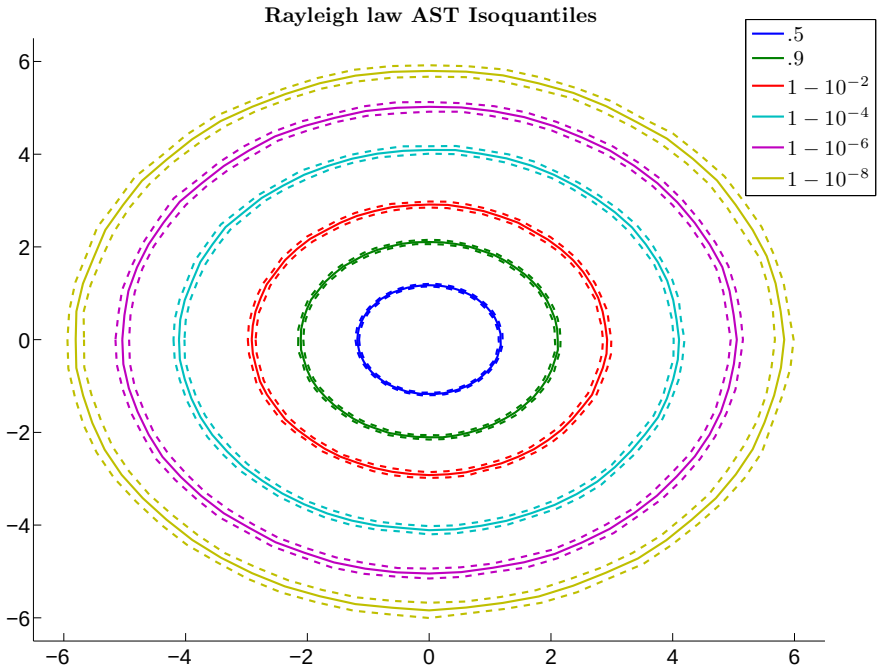
$$\begin{cases} g & \text{is} & \text{a black box mapping from } [0,1]^6 \text{ to } \mathbb{R}^2. \\ X & \sim & \mathcal{U}([0,1]^6) \\ g(X) & \equiv & Y \\ \alpha & \in & \{\frac{1}{2}, \frac{3}{4}, 1-10^{-1}, 1-10^{-2}, 1-10^{-4}, 1-10^{-6}, 1-10^{-8}\} \\ \text{The } S_\alpha^Y\text{s} & \text{are} & \text{to be found.} \end{cases}$$

Only $7 \cdot 10^5$ g evaluations are available to solve ($\mathcal{R}$).

In order to benefit the previous work, we used a Gaussian vector to simulate $\mathcal{U}([0,1]^6)$



(a) Crude Monte Carlo



(b) Adaptive Splitting Technique

Figure A.1: $\mathcal{N}(0_2, I_2)$ law isoquantiles estimations

using Rayleigh's law cumulative distribution function's inverse.

$$\begin{aligned} H &\sim \|\mathcal{N}(0_2, I_2)\| && \text{(A.27)}\\ &\sim f_H(h)dh && \text{(A.28)}\\ f_H(h) &= \mathbf{1}_{\mathbb{R}^+}(h) h \exp^{-\frac{h^2}{2}} && \text{(A.29)}\\ F_H(h) &= \int_{-\infty}^{h} f_H(r)dr && \text{(A.30)}\\ &= 1 - \exp^{-\frac{\max(h,0)^2}{2}} && \text{(A.31)}\\ 0 \leq a < b \leq 1 &\Rightarrow \mathbb{P}[a \leq F_H(H) \leq b] = b - a && \text{(A.32)} \end{aligned}$$

The six $\mathcal{U}([0,1]^6)$ components were independently, identically distributed as $F_H(\|Z\|)$ with $Z \sim \mathcal{N}(0_2, I_2)$. This allowed the use of the same framework as before.

A.6.1 Circular Error Probable Vs Isoquantile Curves

To have a better feeling of the interest of Isoquantiles w.r.t. CEP, we used the same CMC generated samples to draw both the CEPs and Isoquantile curves for the same levels. Besides, the mean impact points envelope was superimposed when useful.

Figure A.2 compares CEP and IS estimations, based on the same one hundred set of $5 \cdot 10^5$ CMC generated samples. The landing surface is clearly not circular and the Isoquantile matches it better than the CEP. This increases the meaningfulness of the estimation as isoquantile surface consider the syteme anisotropy CEP disregarded. Besides IS variance impacts less surface.

As a benefit from IS w.r.t. CEP, the new knowledge IS provides can be of great help. The system we dealt with was a falling launcher. This better idea of its probable impact surface ensures sounder security design. If we consider the system to be a robot going toward a given position and associate the IS with the probabilities of falling in each cone , the designer knows how likely the robot is to miss the target in a given direction and how it does so. This helps him enhance his guidance politic. As a matter of facts, this extra precision is a very convenient way to understand better the randomness at hand and act accordingly in various fields.

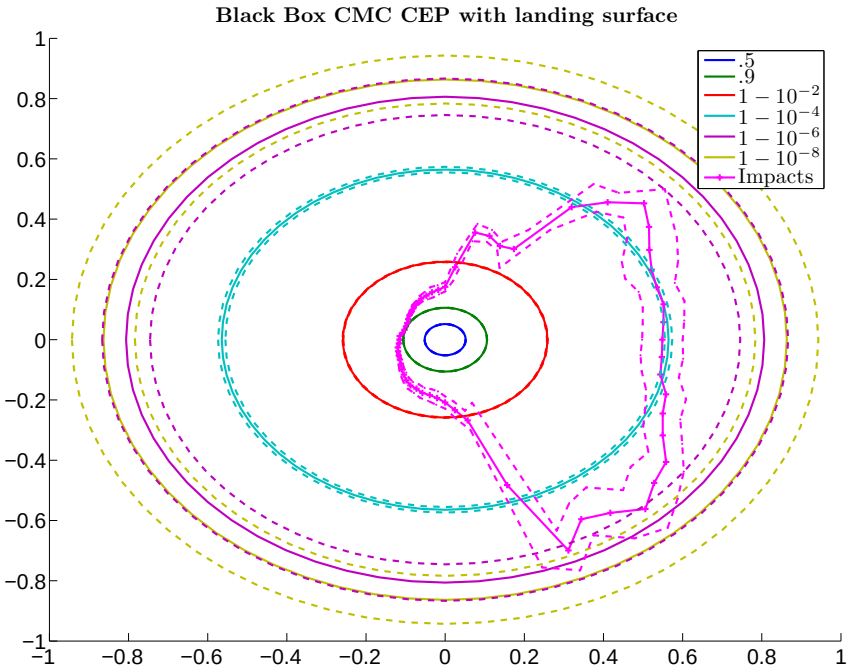
Furthermore, IS can be associated with AST for finer insight than with CMC as explained in the next subsection (A.6.2).

A.6.2 Estimating extreme isoquantiles: CMC Vs AST

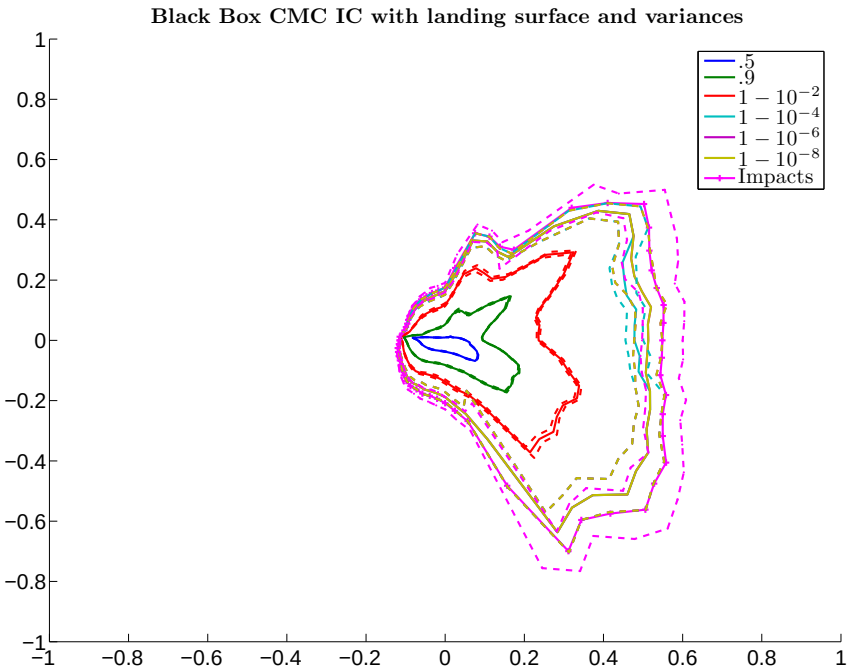
One hundred estimations based on $7 \cdot 10^5$ points each were done using both the CMC and the splitting. The results are presented on figures A.3a and A.3b.

The isoquantiles of level up to $1 - 10^{-4}$ are estimated equivalently w.r.t. both mean and variance. Over this threshold, the CMC's shortcomings appear as it cannot discern the curves. The Splitting, even though it has a greater variance, discerns the last curves and reveals the landing surface to be wider than estimated through CMC.

AST's ability to estimate extreme isoquantile is greater than CMC's. This allows, for instance, better hedging in the finance industry, fitter and less expensive security system design and better guidance in precision robotics.

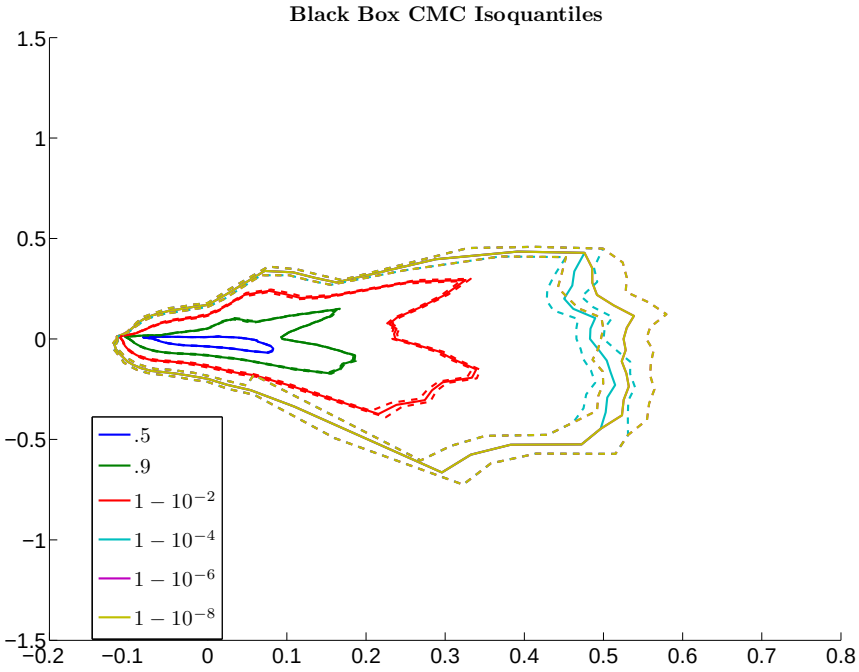


(a) Circular Error Propable

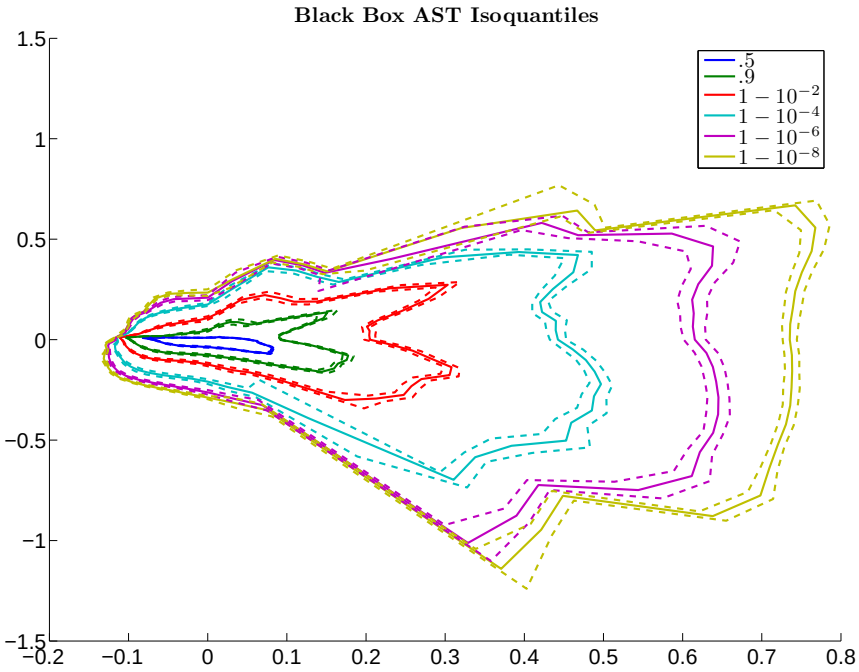


(b) Isoquantile Surface

Figure A.2: Propeller fallout landing surfance with CMC spatial dispersion representation.



(a) Crude Monte Carlo



(b) Adaptive Splitting Technique

Figure A.3: Propeller fallout isoquantile mean and variance

A.7 Conclusion

The quantile definition was extended to the n-dimensional case through the Isoquantile i.e. the border of a centered subset hit by the random variable with a given probability. Isoquantiles offer an angle dependent representation of the random spread around a specified center whereas the benchmark tool, Circular Error Probable assumes isotropy and therefore wastes an information yet available within the data. Isoquantiles provide a finer idea of the spatial randomness at hand and allows a better informed adaptation politic.

Estimating Isoquantile Curves with Adaptive Splitting Technique and Crude Monte-Carlo and comparing the results, we showed AST to be able to estimate isoquantiles of higher level than Crude Monte-Carlo.

In a nutshell, whenever dealing with high level quantiles, we recommend the use of the Adaptive Splitting Technique rather than Crude Monte-Carlo and the use of Isoquantile Curves rather than Circular Error Probable when facing multidimensionality. The AST ensures a better exploration of the probability space, taking rare events into account and IS provides an angle dependant representation of the randomness' spread around a target. This comes very handy in whatever practical purposes that can be seen as a reliability or a impact to target spread characterisation problem and might lead to a very useful density estimation.

Appendix B

Communications

B.1 Publications

- J. Morio and R. Pastel, *Sampling technique for launcher impact safety zone estimation*, Acta Astronautica, Volume 66, Issues 5-6, March-April 2010, pp 736-741.

- J. Morio, R. Pastel and F. Le Gland,*Estimation de probabilités et de quantiles rares pour la caractérisation d'une zone de retombée d'un engin*, Journal de la Société Française de Statistique, (2011), 152, 4, 1-29.

- J. Morio and R. Pastel, *Plug-in estimation of d-dimensional density minimum volume set of a rare event in a complex system*, accepted in Proceedings of the IMechE, Part O, Journal of Risk and Reliability

- J. Morio, R. Pastel and F. Le Gland, *Missile target accuracy estimation with importance splitting*, accepted in Aerospace Science and Technology

B.2 Oral presentations

- Rudy Pastel, Jérôme Morio, François Le Gland, Satellites collision probabilities: switching from numerical integration to the adaptive splitting technique, EUCASS 2011, St Petersburg, Russia, 4-8 July 2011

- Rudy Pastel, Jérôme Morio, François Le Gland, from satellite versus debris collision probabilities to the adaptive splitting technique, RES workshop, 28-29 October 2010, Bordeaux

- R. Pastel, J.Morio, F. Le Gland, Estimating Satellite Versus Debris Collision Probabilities via the Adaptive Splitting Technique, in proceeding of ICCMS 2011

- R. Pastel, J.Morio, F. Le Gland, Representing spatial distribution via the Adaptive Importance Splitting technique and Isoquantile Curves, EMS2010, 17-22 August, 2010 - Piraeus, Greece

B.3 Poster

- R. Pastel, J.Morio, F. Le Gland, Estimating satellite versus debris collision probabilities via the adaptive splitting technique, RESIM 2010, June 21-22, 2010, Cambridge(UK)

List of Tables

List of Figures

Bibliography

[1] Barry C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics (Classics in Applied Mathematics)*. SIAM, 2008.

[2] Yves F. Atchadé, Gareth O. Roberts, and Jeffrey S. Rosenthal. Towards optimal scaling of Metropolis-coupled Markov chain Monte Carlo. *Statistics and Computing*, 20, 2010. Available Online.

[3] Siu-Kui Au and James L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263 – 277, 2001.

[4] J. R. Baxter and Jeffrey S. Rosenthal. Rates of convergence for everywhere-positive Markov chains. *Statistics & Probability Letters*, 22(4):333–338, March 1995. Available Online.

[5] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*.

[6] I. Beichl and F. Sullivan. The importance of Importance Sampling. *Computing in science and engineering*, pages pp.71–73, 1999.

[7] Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49:122–136, 1941.

[8] Gérard Biau, Benoît Cadre, and Bruno Pelletier. A graph-based estimator of the number of clusters. *ESAIM: Probability and Statistics*, 11:272–280, 2007.

[9] P-T. De Boer, D.P. Kroese, S. Mannor, and R.Y. Rubinstein. A Tutorial on the Cross-Entropy Method. *Annals of Operations Research*, pages 19–67, 2004. Page internet du groupe: http://iew3.technion.ac.il/CE/.

[10] P H Borcherds. Importance sampling: an illustrative introduction. *Computing in science and engineering*, pages 405–411, Avril 2000.

[11] Z. I Botev, D.P. Kroese, and T. Taimre. Generalized Cross-Entropy Methods with Applications to Rare-Event simulation and Optimization. *Simulation*, 11:785 – 806, 2007.

[12] James Antonio Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation.* John Wiley, New York, 1990.

[13] James Antonio Bucklew. *Introduction to Rare Event Simulation*. Springer, 2004. Purchase online.

[14] B. Cadre. Kernel estimation of density level sets. *Journal of Multivariate Analysis*, 97:999–1023, 2006.

[15] Claire Cannamela, Josselin Garnier, and Bertrand Iooss. Controlled stratification for quantile estimation. *The Annals of Applied Statistics*, 2(4):1554–1580, December 2008.

[16] Philippe Caperaa and Bernard Van Cutsem. *Méthodes et modèles en statistique non paramétrique*. Les Presses de l'Université de Laval, 1988.

[17] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 2011. Available on line.

[18] Frédéric Cérou and Arnaud Guyader. Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2):417–443, 2007. Available on line.

[19] Frédéric Cérou and Arnaud Guyader. Adaptive particle techniques and rare event estimation. *ESAIM: Proceedings*, 19:65–72, 2007.

[20] I. M. Chakravarti, R. G. Laha, and J. Roy. *Handbook of Methods of Applied Statistics*, volume I. John Wiley and Sons, 1967.

[21] F.Kenneth Chan. *Spacecraft Collision Probability*. The Aerospace Corporation & AIAA, 2008.

[22] Pierre Chauvet. *Aide-mémoire de Géostatistique Linéaire*. Les Presses de l'Ecole des Mines de Paris, 1999. Available on line.

[23] Barry A. Cipra. The Best of the 20th Century: Editors Name Top 10 Algorithms. *SIAM news*, Volume 33:4, 2000. Available Online.

[24] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[25] Mary Kathryn Cowles, Gareth O. Roberts, and Jeffrey S. Rosenthal. Possible biases induced by MCMC convergence diagnostics, 1997. Available Online.

[26] R. Crowther. Orbital debris: a growing threat to space operations. *Philosophical Transactions of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, 361(1802):157, 2003. Available on line.

[27] Laurens de Haan and Ana Ferreira. *Extreme Value Theory: an introduction*. Springer, 2006.

[28] Stéphanie Delavault, Paul Legendre, Romain Garmier, and Bruno Revelin. Improvement of the TLE accuracy model based on a gaussian mixture depending on the propagation duration. In *Astrodynamics Specialist Conference and Exhibit*. AIAA/AAS, 2008.

[29] Jean-François Delmas and Benjamin Jourdain. Does waste-recycling really improve Metropolis-Hastings Monte Carlo algorithm? 2006. Available Online.

[30] Mark Denny. Introduction to Importance Sampling in rare-event simulations. *European Journal of Physics*, pages 403–411, Juillet 2001. mark.denny@baesystems.com.

[31] Luc Duriez. *Eléments de Mécanique Spatiale*. Les documents libres de l'Université de Lille 1, 2005. Available on line in French.

[32] Roger Eckhardt. Stan Ulam, John von Neumann and the Monte Carlo method. *Los Alamos Science*, Special Issue (15):131–141, 1987. Available Online.

[33] Daniel Egloff and Markus Leippold. Quantile estimation with adaptive importance sampling. *Annals of Statistics*, 38:1244–1278, 2010.

[34] Carl-Gustav Esseen. On the Liapunoff limit of error in the theory of probability. *Arkiv for matematik, astronomi och fysik*, A28:1–19, 1942.

[35] N. Etemadi. An elementary proof of the strong law of large numbers. *Probability Theory and Related Fields*, 55:119–122, 1981. Available on line.

[36] Herbert Federer. *Geometric measure theory*, volume Band 153 of *series Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag New York Inc., 1969.

[37] Albert Glassman. The growing threat of space debris. *IIEE-USA Today's Engineer*, 2009. Read on line.

[38] P. Glynn. Importance sampling for Monte Carlo estimation of quantiles. pages 180–185. Publishing House of Saint Petersburg University, 1996.

[39] Marc Hallin, Davy Paindaveine, and Miroslav Šiman. Multivariate quantiles and multiple-output regression quantiles: From l_1 optimization to halfspace depth. *Annals of Statistics*, 38:635–669, 2010.

[40] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. Available on line.

[41] Tim Hesterberg. Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37(2):185–194, 1995.

[42] Rob J Hyndman. Computing and graphing highest density regions. *American Statistician*, 50:120–126, 16 July 1996.

[43] Asaph Hall III. On an experimental determination of π. *The Messenger of Mathematics*, 2:113–114, 1873.

[44] Tomkins R. J. Another proof of Borel's strong law of large numbers. *The American statistician*, 38:208–209, 1984. Available on line.

[45] Adam M. Johansen, Ludger Evers, and Nick Whiteley. *Lecture Notes on Monte Carlo Methods*. Available Online, bristol university edition, 2010.

[46] Nicholas L. Johnson. Orbital debris: the growing threat to space operations. Technical report, 2010. Available on line.

[47] T.S. Kelso. Analysis of the Iridium 33-Cosmos 2251 collision. Technical report, 2009. Available on line.

[48] Martin J. Klein. Principle of detailed balance. *Phys. Rev.*, 97(6):1446–1447, Mar 1955.

[49] L. Kong and I. Mizera. Quantile tomography: using quantiles with multivariate data. *ArXiv e-prints*, May 2008.

[50] Korolev, Victor and Shevtsova, Irina . An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scandinavian Actuarial Journal*, 04 June 2010.

[51] Agnés Lagnoux. Rare event simulation. *PEIS*, 20(1):45–66, January 2006. Available on line.

[52] François Le Gland. *Filtrage bayésien optimal et approximation particulaire*. Cours de l'Ecole Nationale Supérieure de Techniques Avancées, 2008. Available on line in French.

[53] François Le Gland and Nadia Oudjane. *A sequential particle algorithm that keeps the particle system alive*, pages 351–389. Number 337 in Lecture Notes in Control and Information Sciences. Henk Blom and John Lygeros, Springer, Berlin, 2006. Available on line.

[54] Pierre L'Écuyer, François Le Gland, Pascal Lezaud, and Bruno Tuffin. Splitting methods. In Gerardo Rubino and Bruno Tuffin, editors, *Monte Carlo Methods for Rare Event Analysis*, chapter 3, pages 39–61. John Wiley & Sons, Chichester, 2009.

[55] Paul Legendre, Béatrice Deguine, and Bruno Revelin. Two line element accuracy assessment based on a mixture of gaussian laws. In *Astrodynamics Specialist Conference and Exhibit*. AIAA/AAS, 2006.

[56] P.C. Mahalanobis. On the Generalized distance in statistics. *Proceedings of the National Institute of Science of India*, 12:49–55, 1936.

[57] Miguel Piera Martinez, Emmanuel Vazquez, Eric Walter, Gilles Fleury, and Kielbasa Richard. Estimation of extreme values with application to uncertain systems. In *Proceeding of SYSID 2006*, Available Online, 2006.

[58] Julien Marzat. *Diagnostic des systèmes aéronautiques et réglage automatique pour la comparaison de méthodes / Fault diagnosis of aeronautical systems and automatic tuning for method comparison* . PhD thesis, Laboratoire des signaux et systèmes, Novembre 2011.

[59] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculation by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953. Available on line.

[60] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44:335–341, 1949. Available Online.

[61] Nicholas Metropolis. The beginning of the Monte Carlo method. *Los Alamos Science*, Special Issue (15):125–130, 1987. Available Online.

[62] S. P. Meyn and R. L. Tweedie. *Markow chains and stochastic stability*. Springer–Verlag, 1993.

[63] Nicholas Zwiep Miura. *Comparison and design of Simplified General Perturbation Models (SGP4) and code for NASA Johnson Space Center, Orbital debris program office*. PhD thesis, Faculty of California Polytechnic State University, Available Online, May 2009.

[64] F. Morgan. *Geometric measure theory: a beginner's guide*. Academic Press. Academic/Elsevier, 2009.

[65] Jérôme Morio and Rudy Pastel. Sampling technique for launcher impact safety zone estimation. *Acta Astronautica*, 66(5-6):736–741, 2010.

[66] Jérôme Morio and Rudy Pastel. Plug-in estimation of d-dimensional density minimum volume set of a rare event in a complex system. *Journal of Risk and Reliability*, To be published.

[67] Jérôme Morio, Rudy Pastel, and François Le Gland. An overview of importance splitting for rare event simulation. *European Journal of Physics*, 31:1295–1303, 2010.

[68] Jérôme Morio, Rudy Pastel, and François le Gland. Estimating satellite versus debris collision probabilities via the adaptive splitting technique. In *Proceedings of the 3rd International Conference on Computer modeling and simulation, Mumbai,India*, 2011.

[69] Jérôme Morio, Rudy Pastel, and François Le Gland. Estimation de probabilités et de quantiles rares pour la caractérisation d' une zone de retombée d' un engin. *Journal de la Société Française de Statistique*, 152:1–29, 2011.

[70] Jérôme Morio, Rudy Pastel, and François Le Gland. Missile target accuracy estimation with importance splitting. *Aerospace Science and Technology*, To be published.

[71] Carl N. Morris. Natural exponential families with quadric variance functions. *Annals of Statistics*, 10(1):65 – 80, 1982.

[72] C.D. Murray and S.F. Dermott. *Solar system dynamics*. Cambridge University Press, 1999.

[73] Jan C. Neddermeyer. Computationally efficient Nonparametric Importance Sampling. *Journal of the American Statistical Association*, 104(486):788–802, 2009.

[74] Roger B. Nelsen. *An Introduction to Copulas (Springer Series in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[75] Javier Nunez-Garcia, Zoltan Kutalik, Kwang-Hyun Cho, and Olaf Wolkenhauer. Level sets and minimum volume sets of probability density functions. *Approximate Reasoning*, 34, 2003.

[76] Chiwoo Park, Jianhua Z. Huang, and Yu Ding. A computable plug-in estimator of minimum volume sets for novelty detection. *Operations Research*, 58:1469–1480, September 2010.

[77] Richard Rhodes. *The Making of the Atomic Bomb*. Simon & Schuster, 1986.

[78] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical science*, 16(4):351–367, 2001. Available Online.

[79] Jeffrey S. Rosenthal. A review of asymptotic convergence for general state space Markov chains, 1999. Available Online.

[80] R.Y. Rubinstein and D.P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte Carlo Simulation and Machine Learning*. Springer Verlag, 2004.

[81] Liu Jun S. *Monte Carlo Strategies in Scientific Computing*. Lecture Notes in Statistics. 2001.

[82] Clayton Scott and Robert Nowak. Learning Minimum Volume Sets. Technical report, UW-Madison, June 2005.

[83] Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2), 2002. Available on line.

[84] Robert Serfling. Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *Journal of Statistical Planning and Inference*, 22:915–936, 2010.

[85] Claude Shannon. Mathematical theory of communication. *Bell Systems Technical Journal*, 27, 1948.

[86] B.W. Silverman. Density estimation for statistics and data analysis. In *Monographs on Statistics and Applied Probability*. London: Chapman and Hall, 1986.

[87] Nickolay Smirnov. *Space Debris: Hazard Evaluation and Mitigation*. Taylor and Francis, 2002.

[88] Matthew Strand. Metropolis-Hastings Markov Chain Monte Carlo. 2009. Available Online.

[89] Philippe Tassi and Sylvia Legait. *Théorie des probabilités en vue des applications statistiques*. TECHNIP, 1990.

[90] Luke Tierney. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, 22:1701–1728, December 1994. Available on line.

[91] Bruno Tuffin, Pierre L'Ecuyer, and Werner Sandmann. Robustness Properties for Simulations of Highly Reliable Systems, 2006. Available on line.

[92] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, 2006.

[93] Matthew P. Wand, J. S. Marron, and David Ruppert. Transformations in Density Estimation. *Journal of the American Statistical Association*, 86(414), 1991.

[94] Ping Zhang. Nonparametric Importance Sampling. *Journal of the American Statistical Association*, 91(434):1245–1253, September 1996.

Index

ABSTRACT

Rare event dedicated techniques are of great interest for the aerospace industry because of the large amount money that can be lost because of risks associated with minute probabilities. This thesis is focused on the search of probability techniques able to estimate rare event probabilities and extreme quantiles associated with a black box system with static random inputs through two case studies from the industry. The first one is the estimation of the probability of collision between satellites Iridium and Cosmos. The Cross-Entropy (CE), the Non-parametric Adaptive Importance Sampling (NAIS) and an Adaptive Splitting Technique (AST) are compared. Through the comparison, an improved version of NAIS is designed. Whereas NAIS needs to be initiated with a auxiliary random variable which straight away generates rare events, the Adaptive NAIS (ANAIS) allows one to use the original input random as initial auxiliary density and therefore does not require *a priori* knowledge. The second case is the estimation of the safety zone with respect to the fall of a spacecraft booster. Though they can be estimated *via* ANAIS or AST, extreme quantiles are shown to be not adapted to spatial distribution. For that purpose, the Minimum Volume Set (MVS) is chosen from the literature. The Crude Monte Carlo (CMC) plug-in MVS estimator being not adapted to extreme level MVS estimation, both ANAIS and AST are adapted into plug-in extreme MVS estimators. Both the later algorithms outperform the CMC plug-in MVS estimator.
Keywords: rare event probabilities, extreme quantiles, cross-entropy, important sampling, splitting technique, minimum volume sets, density level sets, aerospace, Iridium-Cosmos.

RESUME

Les techniques dédiées aux évènements rares sont d'un grand intérêt pour l'industrie aérospatiale en raison des larges sommes qui peuvent être perdues à cause des risques associés à des probabilités infimes. Cette thèse se concentre la recherche d'outils probabilistes capables d'estimer les probabilités d'évènements rares et les quantiles extrêmes associés à un système boîte noire dont les entrées sont des variables aléatoires. Cette étude est faite au travers de deux cas issus de l'industrie. Le premier est l'estimation de la probabilité de collision entre les satellites Iridium et Cosmos. La Cross-Entropy (CE), le Non-parametric Adaptive Importance Sampling (NAIS) et une Technique de type Adaptive Splitting (AST) sont comparés. Au cours de la comparaison, une version améliorée de NAIS est conçue. Au contraire du NAIS qui doit être initialisé avec une variable aléatoire qui génère d'emblée des événements rares, le NAIS adaptatif (ANAIS) peut être initialisé avec la variable aléatoire d'origine du système et n'exige donc pas de connaissance *a priori*. Le second cas d'étude est l'estimation de la zone de sécurité vis-à-vis de la chute d'un booster de fusée. Bien que les quantiles extrêmes puissent être estimés par le bais de ANAIS ou AST, ils apparaissent comme inadaptés à une distribution spatiale. A cette fin, le Minimum Volume Set (MVS) est choisi dans la littérature. L'estimateur Monte Carlo (MC) de MVS n'étant pas adapté à l'estimation d'un MVS de niveau extrême, des estimateurs dédiés sont conçus à partir d'ANAIS et d'AST . Ces deux derniers surpassent l'estimateur de type MC.
Mots clés: probabilités d'évènements rares, quantiles extrêmes, cross-entropy, importance sampling, technique de splitting, minimum volume set, ensemble de niveau de densité, aérospatiale, Iridium-Cosmos.