

---

# Advanced Statistical and ML Techniques for Pricing and Risk Assessment of Derivative Portfolios

Sébastien HAAG<sup>(\*,1)</sup>

<sup>1</sup>Natixis Bank, Quantitative Trading Analyst

28 avril 2025

---

# Sommaire

Résumé . . . . .	2
<b>1 Introduction</b>	<b>5</b>
<b>2 Risque de Marché : Value-at-Risk et Expected Shortfall</b>	<b>9</b>
2.1 Conception d'un modèle de risque par simulations . . . . .	13
2.2 VaR Paramétrique et Ajustement de Cornish-Fisher . . . . .	16
2.3 Approche alternative pour le calcul de la VaR et de l'ES . . . . .	23
<b>3 Régression par Processus Gaussiens en Apprentissage Automatique</b>	<b>25</b>
3.1 Méthode d'estimation fondée sur la régression par processus gaussiens . . . . .	31
3.2 Entraînement des algorithmes à processus gaussiens . . . . .	33
<b>4 Applications de la régression par processus gaussiens</b>	<b>35</b>
4.1 Valorisation d'un portefeuille de produits dérivés . . . . .	35
4.2 Calcul de la Value-at-Risk (VaR) et de l'Expected Shortfall (ES) d'un portefeuille de produits dérivés . . . . .	46
4.3 Sensibilité et approximation non linéaire de la VaR . . . . .	52
<b>5 Conclusion</b>	<b>55</b>
<b>6 Annexes</b>	<b>57</b>
6.1 Composition du portefeuille du test numérique . . . . .	57
6.2 Jeux d'entraînement pour la régression par processus gaussiens . . . . .	59

## Résumé

La supervision réglementaire des modèles internes utilisés par les institutions financières s'est fortement intensifiée ces dernières années. Sous l'impulsion des réformes successives de **Bâle II**, **III** puis **IV**, ainsi que des inspections approfondies menées par les régulateurs, comme la Banque centrale européenne à travers le programme *TRIM (Targeted Review of Internal Models - 2016)*, *ILAAP (Internal Liquidity Adequacy Assessment Process - 2016)* ou encore *SREP (Supervisory Review And Evaluation Process - 2014)*, les banques doivent désormais démontrer la rigueur, la robustesse et la transparence de leurs méthodologies quantitatives de mesure du risque. Cette montée en exigence s'explique par la succession de crises ayant révélé les failles critiques de certains modèles internes, mal calibrés ou insuffisamment stressés. La faillite de **Lehman Brothers** en septembre 2008, avec une dette estimée à 613 milliards de dollars, a mis en évidence l'incapacité des modèles de risque de marché à capter les effets de contagion systémique et les désajustements de liquidité extrême. Plus récemment, l'effondrement d'**Archegos Capital Management** en mars 2021 a révélé des faiblesses majeures dans la gestion du risque de contrepartie par certaines grandes banques. **Credit Suisse** a ainsi subi une perte de 5,5 milliards de dollars, soulignant un défaut de modélisation des effets de levier excessifs. En mars 2023, la chute de la **Silicon Valley Bank (SVB)**, marquant la troisième plus grande faillite bancaire de l'histoire des États-Unis, a illustré la sous-estimation des risques de taux et de liquidité dans des contextes de portefeuilles fortement concentrés.

Les modèles internes permettent aux banques d'évaluer leur exposition aux différents types de risques financiers dans le cadre du Pilier 1 : **risque de marché**, **risque de crédit**, **risque de contrepartie** et **risque opérationnel**. Chaque classe de risque donne lieu à des modélisations spécifiques, souvent calibrées sur des données historiques internes, des stress tests et des hypothèses prudentes. Contrairement aux approches standardisées proposées par les régulateurs, les modèles internes offrent une flexibilité importante, mais nécessitent une validation stricte par les superviseurs (BCE, ANC, Fed, BCBS...).

Concernant le **risque de marché**, les modèles internes incluent des outils tels que la *Value-at-Risk (VaR)*, l'*Expected Shortfall (ES)*, l'*Incremental Risk Charge (IRC)* ou encore le *Default Risk Charge (DRC)* dans le cadre du nouveau standard *FRTB (Fundamental Review of the Trading Book - 2023)*. Par exemple, JPMorgan a longtemps utilisé son célèbre modèle *RiskMetrics* basé sur la covariance des rendements, tandis que Goldman Sachs et BNP Paribas s'appuient sur des VaR historiques non-paramétriques. L'ES, quant à elle, devient la référence réglementaire sous FRTB, et des institutions comme Natixis ou UBS développent des moteurs de calcul internes combinant Monte Carlo et méthodes semi-analytiques pour dérivés complexes.

Du côté du **risque de crédit**, les banques recourent aux approches *IRB (Internal Ratings-Based)* pour estimer les paramètres fondamentaux que sont la *probabilité de défaut (PD)*, la *perte en cas de défaut (LGD)* calibré via des analyses historiques et l'*exposition au défaut (EAD)* modélisé à l'aide de *Credit Conversion Factors (CCF)*. L'estimation de la probabilité de défaut (PD) repose historiquement sur des *modèles de régression logistique (logit)* ou *probit*, permettant de modéliser la probabilité de défaut à partir de caractéristiques observables de l'emprunteur (notation interne, ratios financiers, secteur d'activité, géographie, etc.). Dans le modèle logit, la log-odds du défaut est modélisée comme une combi-

naison linéaire des covariables explicatives. Le modèle probit, quant à lui, repose sur la fonction de répartition d'une loi normale standard, ce qui lui confère une interprétation plus structurelle, en lien direct avec les modèles de type *Merton* (défaut lorsque la valeur de l'actif tombe sous un certain seuil). Ces approches dites *fréquentistes* ont l'avantage d'être facilement calibrables par maximum de vraisemblance et largement utilisées dans la pratique bancaire. Toutefois, elles présentent certaines limitations, notamment l'absence de prise en compte de l'incertitude sur les paramètres estimés, ainsi que leur rigidité en présence de fortes interactions entre variables ou de structures hiérarchiques complexes. Pour dépasser ces limites, les grandes institutions financières s'orientent de plus en plus vers des *modèles bayésiens*, dans lesquels les coefficients du modèle sont considérés comme des variables aléatoires dotées de distributions a priori. L'approche bayésienne permet ainsi de quantifier l'incertitude sur les estimations de probabilité de défaut, en calculant des distributions a posteriori conditionnelles aux données observées. En pratique, ces modèles peuvent intégrer des effets fixes (liés aux caractéristiques de l'emprunteur) et des effets aléatoires hiérarchiques (représentant, par exemple, les effets spécifiques à un secteur d'activité, une zone géographique ou un segment de clientèle). Le recours à des techniques d'inférence par *Monte Carlo par chaînes de Markov (MCMC)*, telles que le *Gibbs sampling* ou le *Hamiltonian Monte Carlo*, rend possible l'estimation de ces modèles même dans des contextes à haute dimension. Des frameworks comme *Stan* ou *PyMC* sont aujourd'hui utilisés dans les laboratoires quantitatifs des grandes banques pour implémenter ces modèles à grande échelle. Par exemple, BNP Paribas applique des modèles logit bayésiens régularisés pour la calibration de PD sur des portefeuilles à faible granularité, tandis que Natixis utilise des modèles hiérarchiques bayésiens pour estimer la PD par industrie et par région.

Le **risque de contrepartie**, quant à lui, fait appel à des modèles internes de *Credit Valuation Adjustment (CVA)* ou plus généralement de *XVA (CVA, DVA, FVA, MVA, KVA)* qui désigne une famille d'ajustements de valorisation destinés à intégrer différents types de coûts et de risques dans la valorisation des produits dérivés, en particulier les dérivés de gré à gré (OTC). Ces modèles combinent des simulations stochastiques de valorisation (souvent via *BSDEs* ou *Hull-White*) avec des scénarios de défaut. Des banques comme Morgan Stanley ou Goldman Sachs disposent de desks spécialisés, avec des moteurs XVA intégrés aux plateformes de pricing.

Enfin, le **risque opérationnel**, historiquement modélisé via *l'AMA (Advanced Measurement Approach)*, s'appuyait sur des approches comme le *Loss Distribution Approach (LDA)*, combinant distributions extrêmes (ex : loi lognormale ou Pareto) et données internes de pertes. Cette approche a été remplacée dans Bâle IV par une méthode plus standardisée (*SMA*), mais plusieurs banques (Deutsche Bank, HSBC) ont conservé leurs infrastructures internes à des fins internes ou de pilotage.

Or, l'évolution récente des modèles internes ne se limite pas à une réponse aux contraintes réglementaires ; elle s'inscrit également dans un mouvement plus large d'intégration des méthodes issues de *l'apprentissage automatique (machine learning)* au sein des pratiques de gestion du risque financier. Face à l'accroissement de la complexité des portefeuilles de trading, à la sophistication croissante des instruments dérivés, et aux exigences computationnelles imposées par les approches telles que FRTB, les modèles traditionnels atteignent leurs limites en termes

de vitesse, scalabilité et précision. Dans ce contexte, des approches innovantes émergent au sein des directions quantitatives des grandes institutions financières. Des techniques telles que les *réseaux de neurones profonds*, les *autoencodeurs*, les *méthodes d'ensemble* (*random forests*, *boosting*), ou encore la *régression par processus gaussiens* (*Gaussian Process Regression*, *GPR*) sont aujourd'hui explorées pour améliorer significativement l'efficacité des chaînes de valorisation et de calcul du capital réglementaire. Ces méthodes permettent notamment de substituer, dans certaines étapes clés, les moteurs de pricing classiques (souvent lents et coûteux) par des approximations apprises sur des données simulées. Parmi ces approches, la régression par processus gaussiens (GPR) suscite un intérêt particulier en raison de son caractère non paramétrique, probabiliste, et interprétable. Elle permet de construire un estimateur bayésien de fonctions complexes, avec une incertitude quantifiée sur les prédictions. En formant un modèle GPR sur un jeu de données généré par un moteur de valorisation sophistiqué (modèle de Black-Scholes, Hull-White, Heston, etc.), il devient possible d'apprendre la surface de valorisation d'un portefeuille de dérivés. Une fois entraîné, le modèle GPR permet de réévaluer très rapidement les instruments pour de nouveaux scénarios de marché, ce qui est crucial dans les calculs de Value-at-Risk (VaR) et Expected Shortfall (ES) où des millions de revalorisations sont nécessaires. Cette stratégie s'inscrit dans une logique de "*risk analytics acceleration*", où le machine learning est utilisé comme *méta-approximation* de tâches numériques lourdes. Plusieurs études récentes menées au sein de banques comme Natixis, ING ou Société Générale montrent que la GPR peut réduire les temps de calcul de manière significative (parfois d'un facteur 100 à 1000), tout en maintenant une précision compatible avec les seuils de validation interne. En particulier, les gains d'efficacité obtenus dans les calculs de VaR/ES sont robustes à la taille et à la composition du portefeuille, ce qui confère à la méthode un fort potentiel d'industrialisation.

Ce mémoire s'inscrit dans cette dynamique d'hybridation entre modèles statistiques classiques et apprentissage automatique. Nous proposons d'examiner en détail le processus de validation des modèles internes de risque de marché, avec un focus particulier sur l'implémentation pratique de la VaR et de l'Expected Shortfall, et d'explorer les apports concrets des processus gaussiens pour l'accélération des revalorisations de portefeuilles de produits dérivés. À travers des tests numériques sur des portefeuilles synthétiques, nous évaluons l'efficacité de la GPR en termes de précision, de rapidité et de robustesse vis-à-vis des contraintes réglementaires actuelles.

**Keywords :** Modèles internes, Bâle II/III/IV, TRIM, SREP, ILAAP, risque de marché, risque de crédit, risque de contrepartie, risque opérationnel, Value-at-Risk (VaR), Expected Shortfall (ES), Incremental Risk Charge (IRC), Default Risk Charge (DRC), FRTB, RiskMetrics, Monte Carlo, IRB, probabilité de défaut (PD), perte en cas de défaut (LGD), exposition au défaut (EAD), modèles logit/probit, modèles bayésiens, MCMC, Gibbs sampling, Hamiltonian Monte Carlo, XVA, CVA, DVA, FVA, MVA, KVA, BSDEs, Hull-White, LDA, AMA, SMA, machine learning, réseaux de neurones profonds, autoencodeurs, random forests, boosting, Gaussian Process Regression (GPR).

# Chapitre 1

## Introduction

Si la gestion du risque a toujours été une fonction essentielle de l'activité bancaire, sa nature et ses finalités ont profondément évolué au cours des deux dernières décennies. D'une activité historiquement centrée sur la couverture individuelle de positions ou la conformité prudentielle minimale, elle est devenue un levier de pilotage stratégique, intégrée aux processus de décision, de tarification et d'allocation du capital économique.

Cette mutation s'accompagne d'un changement de paradigme : le risque n'est plus seulement un objet de quantification, mais un système dynamique d'interdépendances qu'il faut cartographier, hiérarchiser et anticiper. Il ne s'agit plus simplement de mesurer une perte potentielle, mais d'évaluer les mécanismes de transmission du stress, de fragilité systémique, et de résilience organisationnelle. Cette approche plus systémique, promue notamment par les institutions comme le *Financial Stability Board* ou la BCE, repose sur une refondation des outils statistiques et une hybridation croissante entre modèles déterministes, probabilistes, et apprentissage statistique.

C'est dans ce contexte qu'émerge la notion de "*Risk Intelligence*", qui désigne la capacité d'une organisation à extraire des signaux faibles dans des environnements de données bruitées et instables, à ajuster dynamiquement ses métriques de risque et à arbitrer entre différentes sources d'incertitude. Ce glissement nécessite un renouvellement des méthodologies utilisées, mais aussi une évolution des cadres opérationnels et réglementaires au sein desquels les modèles sont développés, validés et utilisés.

Parallèlement, la montée en puissance des technologies de traitement de données à l'exemple du *calcul distribué*, *cloud computing*, *bases de données vectorielles*, etc., a permis de démocratiser l'accès à des outils jusqu'alors réservés aux laboratoires quantitatifs de pointe. Le machine learning, longtemps cantonné à la sphère académique ou aux applications marketing, est aujourd'hui activement intégré dans des modules critiques de la gestion du risque, non pas pour remplacer les modèles réglementaires, mais pour compléter ou accélérer certaines étapes clés : *sélection de variables explicatives*, *détection d'événements rares*, *modélisation des dépendances extrêmes*, ou *approximation de surfaces de prix complexes*.

Ce transfert technologique soulève cependant de nouveaux défis. Comment assurer la traçabilité, l'auditabilité et la robustesse de modèles fondés sur l'apprentissage statistique dans un contexte prudentiel ? Comment concilier l'exigence de transparence du superviseur avec la nature "*boîte noire*" de certains algorithmes ? Ces interrogations ont conduit à l'émergence d'un nouveau champ interdisciplinaire, à l'interface entre mathématiques appliquées, science des données, informatique théorique et droit bancaire.

C'est à l'articulation de ces tensions que se situe le présent rapport. Il ne s'agit pas de substituer les outils du machine learning aux approches classiques, mais de montrer comment, dans une architecture de modélisation rigoureuse, certains algorithmes issus de l'apprentissage supervisé,

et notamment la régression par processus gaussiens, peuvent jouer un rôle d'accélérateur fiable dans le cadre de simulations complexes exigées par les normes réglementaires actuelles.

Malgré les avancées théoriques et les perspectives prometteuses offertes par les techniques issues de l'apprentissage automatique, l'intégration réelle du machine learning dans la finance quantitative demeure encore limitée à des usages expérimentaux ou exploratoires, en marge des chaînes de calcul réglementaires. Cette situation s'explique autant par des freins institutionnels (besoin de transparence, gouvernance modèle, traçabilité des décisions) que par des obstacles techniques (risque d'overfitting, manque de robustesse hors échantillon, difficulté d'interprétation des modèles dits "boîtes noires"). Cela n'empêche pas une communauté de recherche croissante, à l'interface entre data science, économétrie et ingénierie financière, d'explorer des cas d'usage ciblés pour lesquels le machine learning peut compléter, améliorer, voire transformer certaines pratiques historiques.

Dans le domaine de la valorisation d'options, plusieurs travaux récents ont mis en évidence la capacité des réseaux de neurones artificiels (ANN) à reproduire la surface de **volatilité implicite** et à approximer rapidement les prix issus de modèles complexes, sans recourir à des méthodes numériques intensives comme les arbres binomiaux ou les schémas de Monte Carlo. Ces réseaux sont parfois entraînés sur des bases simulées à partir de modèles comme **Black-Scholes**, **Heston**, ou **SABR**, et permettent une revalorisation quasi instantanée dans des contextes à très haute fréquence. L'objectif principal est ici l'accélération des processus de pricing, en particulier dans les phases de calibration, de gestion intraday ou de revalorisation de masse.

En parallèle, certains travaux explorent l'usage de modèles non supervisés, tels que les mélanges gaussiens (*Gaussian Mixture Models*), ou semi-paramétriques comme les processus gaussiens, pour des tâches de prévision de courbes de taux, de régression de surfaces de prix, ou de réduction de dimensionnalité dans des modèles stochastiques à multiples facteurs. Ces techniques présentent l'avantage d'une flexibilité fonctionnelle importante et d'une capacité d'ajustement fine à des données empiriques complexes, notamment dans des contextes non stationnaires ou avec des régimes de volatilité multiples.

Du côté du **risk forecasting**, plusieurs approches hybrides combinant économétrie traditionnelle et deep learning ont été proposées. Par exemple, dans le cadre de la prévision de la volatilité, des études récentes montrent que l'intégration de couches de réseaux de neurones dans des architectures inspirées du modèle **GARCH** permet d'obtenir des performances prédictives supérieures, notamment sur des séries à forte hétéroscédasticité conditionnelle. Une proposition notable est celle du modèle **GELM** (GARCH + Extreme Learning Machine), introduit par Zhang et al., qui utilise une structure de type **GELM-RBF** (fonction de base radiale) pour estimer la volatilité d'une série temporelle, et en déduire une Value-at-Risk plus précise. Ce type de modèle repose sur des mappings non linéaires aléatoires, sans hypothèse gaussienne, et illustre bien l'émergence d'architectures hybrides, à mi-chemin entre les méthodes statistiques et l'apprentissage profond.

Malgré ces avancées, aucun algorithme de machine learning ne s'est imposé comme référence dans les systèmes de mesure du risque réglementaire. Plusieurs raisons expliquent cette prudence : absence de cadre normatif clair pour la validation des modèles basés sur l'apprentissage, difficulté à documenter l'architecture et la stabilité des résultats, manque de tests de robustesse dans des contextes extrêmes ou multi-crisis. En outre, les exigences imposées par les régulateurs en matière d'auditabilité, de conservatisme et de robustesse empirique freinent leur adoption dans les environnements critiques comme le calcul du capital réglementaire ou le reporting prudentiel.

Cela dit, un nouveau paradigme émerge : celui d'un usage complémentaire, où le machine learning intervient en amont (prétraitement des données, calibration, génération de scénarios), en aval (vérification, clustering de stress tests, contrôle qualité), ou comme méta-modèle dans des environnements de test. Ce positionnement stratégique ouvre la voie à une finance quantitative

plus agile, dans laquelle les modèles statistiques classiques seraient renforcés, mais non remplacés, par des algorithmes d'apprentissage plus souples et adaptatifs.

Contrairement à de nombreux algorithmes d'apprentissage supervisé, les **processus gaussiens (GP)** ne cherchent pas à fournir un estimateur ponctuel de la réponse mais à inférer une distribution a posteriori sur les fonctions susceptibles de générer les données observées. Cette propriété en fait un outil de choix dans les contextes où l'incertitude sur la prédiction est aussi importante que la prédiction elle-même, comme en finance, où les décisions reposent sur des estimations soumises à des marges d'erreur critiques.

Le formalisme des GP s'inscrit pleinement dans une approche bayésienne non paramétrique, dans laquelle l'utilisateur définit une croyance a priori sur l'espace des fonctions (via un noyau de covariance) avant d'actualiser cette croyance à l'aide des données observées. Cette mise à jour génère une famille de fonctions postérieures compatibles avec les observations, accompagnées d'intervalles de confiance quantifiés. Contrairement aux réseaux de neurones profonds, souvent opaques et sensibles au surapprentissage, la **GPR (Gaussian Process Regression)** permet de contrôler la complexité du modèle de manière naturelle, en intégrant la notion de lissage statistique via le choix du noyau.

En pratique, cette approche présente des avantages structurels importants pour la modélisation en finance quantitative :

- **Gestion explicite du trade-off biais/variance** : grâce à l'ajustement des hyperparamètres du noyau (longueur de corrélation, variance de signal, etc.).
- **Adaptabilité locale** : là où les données sont denses, le modèle s'ajuste fortement ; là où les données sont rares, il lisse davantage. Cette propriété est cruciale pour les séries temporelles financières sujettes à des régimes de volatilité changeants.
- **Estimation de l'incertitude prédictive** : le modèle fournit une mesure de la variance à chaque point de l'espace, permettant de détecter les zones de faible fiabilité, un critère central dans les environnements soumis à supervision réglementaire.
- **Robustesse aux petites tailles d'échantillon** : contrairement aux réseaux de neurones profonds, qui nécessitent de grandes quantités de données labellisées pour converger, les processus gaussiens conservent une bonne capacité d'apprentissage même avec des jeux de données limités.

Sur le plan computationnel, les processus gaussiens souffrent historiquement d'un coût algorithmique élevé (de l'ordre de  $\mathcal{O}(n^3)$  pour  $n$  observations), mais des solutions modernes comme la *GPR parcimonieuse*, les *variational approximations*, ou les *inducing points methods* ont permis d'envisager leur scalabilité sur des bases de données volumineuses. Ces innovations les rendent aujourd'hui compétitifs dans des contextes où la vitesse d'approximation et la précision de la prédiction probabiliste sont essentielles, comme dans les simulations massives associées au FRTB ou au calcul d'Expected Shortfall sous contrainte de temps réel.

Sur un plan plus fondamental, le recours à la GPR participe d'une reconfiguration épistémologique de la modélisation en finance : on passe d'une logique de modèles analytiques à paramètres fixes vers une logique de modèles adaptatifs, fondés sur l'observation, et capables d'évoluer avec les données. C'est ce changement de posture, de la modélisation fermée à l'approximation probabiliste ouverte, qui marque aujourd'hui une rupture méthodologique, et qui justifie un intérêt croissant pour les processus gaussiens comme infrastructure probabiliste de substitution dans les chaînes de valorisation complexes.



Une originalité de la démarche proposée dans ce mémoire tient à l'échelle à laquelle est mobilisée la régression par processus gaussiens. Alors que la majorité des travaux existants utilisent la GPR pour approximer la valorisation d'instruments financiers de manière unitaire, typiquement pour des options isolées ou des blocs homogènes, nous adoptons ici une approche plus intégrée, en appliquant directement l'algorithme d'apprentissage sur l'ensemble du portefeuille de trading. L'objectif n'est pas de reconstituer les prix individuels, mais d'apprendre une fonction d'évaluation globale, tenant compte de la structure interne du portefeuille et de ses interactions avec les facteurs de risque.

Cette approche globale présente deux avantages majeurs. D'une part, elle évite les étapes coûteuses d'agrégation post-pricing en considérant le portefeuille comme une entité fonctionnelle. D'autre part, elle permet de traiter directement les dépendances croisées entre instruments, ce qui est souvent négligé dans les approches analytiques classiques. En pratique, et afin de garantir la faisabilité numérique de l'entraînement, une parallélisation par blocs de facteurs de risque peut être introduite en amont pour découpler certaines dimensions du problème.

Au-delà de la valorisation, ce mémoire s'inscrit également dans une perspective d'évaluation critique des performances des modèles de risque via le back-testing. Plus précisément, nous proposons d'exploiter des techniques de classification supervisée issues du machine learning pour interpréter les dépassements de la Value-at-Risk. Il s'agit ici de distinguer les pertes conformes au modèle, c'est-à-dire correspondant à des mouvements de marché capturés par la VaR, de celles résultant de facteurs exogènes ou mal modélisés, révélateurs d'erreurs structurelles ou de limites méthodologiques (problèmes de spécification, données non représentatives, effets de queue non anticipés, etc.).

Ce cadre de backtesting enrichi permet non seulement de quantifier la performance statistique du modèle de VaR, mais aussi de diagnostiquer ses biais récurrents. En identifiant les conditions de marché ou les segments de portefeuille associés aux exceptions, on ouvre la voie à une calibration plus fine, fondée sur une rétroaction empirique. Ce couplage entre prédiction probabiliste (via la GPR) et classification ex post des écarts (via le machine learning) illustre bien la convergence progressive entre modélisation statistique et contrôle qualité dans les pratiques modernes de gestion du risque.

## Chapitre 2

# Risque de Marché : Value-at-Risk et Expected Shortfall

Le risque de marché désigne la possibilité de pertes financières liées à des fluctuations défavorables des variables de marché affectant la valeur des actifs et passifs d'un établissement. Il s'applique aussi bien aux positions figurant au bilan qu'aux éléments hors bilan, et découle d'une exposition à des facteurs tels que les taux d'intérêt, les prix des actions, les cours de change ou encore les prix des matières premières.

Cette catégorie de risque se décline en plusieurs composantes, selon la nature des variables sous-jacentes :

- **Le risque de taux d'intérêt** : il concerne les variations des courbes de rendement affectant la valorisation des produits de taux.
- **Le risque actions** : il est lié à l'évolution imprévisible des marchés boursiers.
- **Le risque de change** : il traduit l'exposition aux mouvements des devises dans les portefeuilles multidevises.
- **Le risque de matières premières** : il est en lien avec la volatilité des prix de l'énergie, des métaux ou des produits agricoles.

Chaque sous-risque implique une dynamique propre, influencée par des facteurs macroéconomiques, géopolitiques, techniques ou comportementaux. La granularité de cette classification permet aux institutions financières d'élaborer des systèmes de mesure et de couverture adaptés à la nature des actifs détenus et à leur profil d'exposition global.

Dans le cadre réglementaire, notamment celui des accords de Bâle, le risque de marché fait l'objet d'une double approche : **l'approche standardisée (SA)** et **l'approche par modèles internes (IMA)**.

Dans le cadre du dispositif FRTB, l'approche standardisée a été profondément remaniée afin de mieux refléter les risques sous-jacents, tout en offrant un cadre plus granulaire et moins arbitraire que les précédentes versions. Cette nouvelle mouture repose désormais sur une structure modulaire, composée de trois briques principales. La **Sensitivities-Based Method (SBM)** constitue le cœur de l'approche standardisée. Elle repose sur le calcul de la sensibilité de chaque position aux facteurs de risque fondamentaux (taux d'intérêt, crédit, actions, change, matières premières). À chaque facteur est associé un choc réglementaire, et les sensibilités sont agrégées selon une structure matricielle qui prend en compte les effets de diversification. Cette méthode permet ainsi une mesure du risque plus dynamique et plus ajustée que les pondérations fixes des

anciennes approches, tout en restant dans un cadre standardisé. Le recours aux sensibilités introduit une forme implicite de modélisation, sans toutefois exiger la validation complète de modèles internes comme dans l'IMA. Le **Default Risk Charge (DRC)** vise à capter le risque de défaut spécifique, en particulier pour les titres de crédit ou dérivés sur crédit. Il s'agit ici d'un module séparé de l'analyse de la volatilité de marché, centré sur les pertes attendues en cas de défaut. Contrairement au SBM, le DRC n'utilise pas de sensibilités mais repose sur une logique de pertes extrêmes, selon des grilles d'exposition, de maturité, et de qualité de crédit. Ce bloc reflète une reconnaissance par le régulateur de la nécessité de traiter séparément le risque de saut de défaut, dont la dynamique diffère des fluctuations continues des marchés. Enfin, le **Residual Risk Add-On (RRAO)** s'adresse aux instruments pour lesquels les risques ne peuvent être intégrés de manière satisfaisante dans les deux blocs précédents. Il s'agit souvent de produits structurés complexes, d'options exotiques, ou de positions dont les sensibilités sont instables ou difficilement quantifiables. Le RRAO introduit une charge de capital additionnelle forfaitaire, proportionnelle à la valeur notionnelle ou au mark-to-market de la position. Cette composante vise à couvrir les zones d'ombre méthodologiques du modèle standardisé et à décourager l'accumulation de risques non transparents.

À l'opposé de la SA, l'IMA permet aux banques de développer leurs propres modèles de mesure du risque de marché, à condition, comme dit précédemment, qu'ils soient validés par les régulateurs, dans le cadre d'un processus rigoureux d'approbation. Ces modèles doivent respecter des critères stricts de robustesse statistique, de calibration, de gouvernance et de documentation. Historiquement, ces modèles étaient fondés sur la Value-at-Risk (VaR), une mesure quantifiant la perte maximale attendue sur un horizon donné pour un certain niveau de confiance (par exemple, 99 % sur 10 jours). Toutefois, la crise financière de 2007-2008 a mis en évidence plusieurs faiblesses de la VaR. En particulier, cette mesure ne prend pas en compte la taille des pertes au-delà du seuil (ce qu'on appelle les pertes extrêmes), ce qui a conduit les régulateurs à la considérer comme non cohérente au sens mathématique. L'un des critères fondamentaux de cohérence d'une mesure de risque est la sous-additivité, c'est-à-dire la capacité à récompenser la diversification. Mathématiquement, cela signifie que pour deux portefeuilles  $A$  et  $B$ , on doit avoir :

$$\rho(A + B) \leq \rho(A) + \rho(B)$$

où  $\rho$  désigne la mesure de risque utilisée. Or, la VaR ne respecte pas toujours cette propriété, en particulier lorsque les distributions de pertes présentent de fortes asymétries ou queues épaisses. Cela peut conduire à des estimations de risque global excessivement optimistes, en ne reflétant pas correctement le bénéfice de diversification. Ce défaut théorique, combiné à des observations empiriques durant les périodes de crise, a motivé le remplacement progressif de la VaR par l'Expected Shortfall (ES) dans le nouveau cadre réglementaire FRTB, une mesure cohérente qui surmonte justement ces limites en intégrant les pertes au-delà du quantile.

Dans la suite, la valeur  $V_t$  du portefeuille (ou de la position financière considérée) est modélisée comme une fonction dépendant explicitement du temps  $t$  et d'un vecteur  $Z_t$  de  $m$  facteurs de risque sous-jacents. Formellement, on écrit :

$$V_t = f(t, Z_t)$$

où  $Z_t = (Z_{t,1}, Z_{t,2}, \dots, Z_{t,m})^\top \in \mathbb{R}^m$  représente l'état des facteurs de risque à l'instant  $t$ . Ces facteurs de risque peuvent inclure des taux d'intérêt, des cours d'actions, des spreads de crédit, des prix de matières premières, des volatilités implicites, ou encore des taux de change, selon la nature du portefeuille.

Pour analyser l'évolution du risque sur un horizon temporel  $h$ , on introduit le vecteur des variations des facteurs de risque :

$$X_t := Z_{t+h} - Z_t$$

Autrement dit,  $X_t$  capture les chocs observés sur les facteurs de risque entre les instants  $t$  et  $t+h$ . Dans une approche historique, ces variations  $X_t$  sont estimées à partir de données passées (par exemple à partir de séries glissantes journalières, hebdomadaires ou mensuelles selon l'horizon de risque retenu). L'intérêt de cette modélisation réside dans la possibilité de relier de manière explicite les changements dans les facteurs de risque à ceux dans la valeur du portefeuille. Puisque  $Z_t$  est supposé connu à la date d'évaluation, la perte conditionnelle sur l'horizon  $h$ , notée  $L_t(h)$ , peut être définie comme la variation négative de valeur du portefeuille sur cet intervalle :

$$\begin{aligned} L_t(h) &:= -(V_{t+h} - V_t) \\ &= -[f(t+h, Z_{t+h}) - f(t, Z_t)] \\ &= -[f(t+h, Z_t + X_t) - f(t, Z_t)] \end{aligned}$$

Cette formulation est centrale dans les méthodes de simulation du risque de marché (backtesting, estimation de VaR ou d'Expected Shortfall), car elle permet de propager les chocs  $X_t$  dans un moteur de valorisation  $f$ , en gardant  $Z_t$  comme point de départ connu. Elle sert aussi de base à la construction de « scenarios-based pricing » ou aux méthodes de delta-gamma approximation lorsque  $f$  est trop coûteuse à calculer de manière exacte.

Jusqu'à récemment, la Value at Risk (VaR) constituait la mesure de référence pour l'estimation du risque de marché auquel une banque est exposée, ainsi que pour la détermination du capital réglementaire minimum requis pour couvrir ce risque dans le cadre de l'approche standard ou du modèle interne. La VaR est une mesure de risque quantile, qui donne une estimation de la perte maximale anticipée, à un niveau de confiance donné  $\beta$ , sur un horizon temporel fixé  $h$  (typiquement un jour, 10 jours, ou deux semaines selon les exigences réglementaires). Formellement, la VaR au niveau de confiance  $1 - \beta$  est définie par :

$$\text{VaR}_\beta := -\inf \{v \in \mathbb{R} \mid \mathbb{P}(\Delta V_t \leq v) > \beta\} = -\inf \{v \in \mathbb{R} \mid \mathbb{P}(L_t(h) \geq -v) > \beta\}$$

où :

- $\Delta V_t = V_{t+h} - V_t$  est la variation aléatoire de la valeur du portefeuille entre aujourd'hui et l'horizon  $t+h$ ,
- $L_t(h) := -\Delta V_t$  est la perte réalisée sur cet horizon,
- $\beta \in (0, 1)$  est le niveau de risque résiduel (par exemple,  $\beta = 1\%$  pour une confiance de 99 %).

Cependant, bien que la Value-at-Risk fournisse une information utile sur le seuil à partir duquel les pertes deviennent rares, elle reste silencieuse quant à la gravité des pertes au-delà de ce seuil. Cette limitation a conduit à l'adoption progressive d'une mesure plus robuste et plus informative : l'Expected Shortfall (ES), également appelée Conditional Value-at-Risk (CVaR). L'Expected Shortfall est définie comme l'espérance des pertes, conditionnellement au dépassement du niveau de VaR. En d'autres termes, il s'agit de la perte moyenne subie dans les pires  $\beta\%$  des cas. Cette mesure permet donc de quantifier le risque dans la queue extrême de la distribution, là où les conséquences économiques peuvent être les plus critiques. Pour une distribution continue des pertes, l'Expected Shortfall au niveau de confiance  $1 - \beta$  s'écrit :

$$\text{ES}_\beta = \mathbb{E}[-\Delta V_t \mid \Delta V_t \leq -\text{VaR}_\beta] = \frac{1}{1 - \beta} \int_0^\beta \text{VaR}_\alpha d\alpha$$

En effet, soit  $L_t(h)$  la perte réalisée sur l'horizon  $h$ , définie comme :

$$L_t(h) := -\Delta V_t = -(V_{t+h} - V_t)$$

On considère une mesure de risque au niveau de confiance  $1 - \beta$ , avec  $\beta \in (0, 1)$ . La **Value-at-Risk (VaR)** au niveau  $\beta$  est donnée par le quantile  $\text{VaR}_\beta$  tel que :

$$\mathbb{P}(L_t(h) \leq \text{VaR}_\beta) = \beta$$

L'**Expected Shortfall (ES)** est alors définie comme la perte moyenne conditionnelle lorsque celle-ci dépasse la VaR :

$$\text{ES}_\beta := \mathbb{E}[L_t(h) \mid L_t(h) \geq \text{VaR}_\beta]$$

Dans le cas où la perte  $L_t(h)$  est une variable aléatoire continue, on peut obtenir une représentation élégante de l'Expected Shortfall à partir de la fonction quantile (i.e. de la VaR en tant que fonction du niveau de confiance). Pour cela, on rappelle que la **fonction quantile** de niveau  $\alpha$  associée à la distribution de  $L_t(h)$  est donnée par :

$$\text{VaR}_\alpha = \inf\{x \in \mathbb{R} \mid \mathbb{P}(L_t(h) \leq x) \geq \alpha\}$$

Cette fonction  $\alpha \mapsto \text{VaR}_\alpha$  est croissante et représente, pour chaque niveau de confiance  $\alpha$ , la perte maximale qui ne sera pas dépassée avec une probabilité de  $\alpha$ .

On souhaite alors exprimer l'Expected Shortfall au niveau  $\beta$  comme la **moyenne des quantiles** de la perte au-delà du niveau  $\beta$ , ce qui revient à intégrer la VaR entre  $\beta$  et 1. On montre que :

$$\text{ES}_\beta = \frac{1}{1 - \beta} \int_\beta^1 \text{VaR}_\alpha d\alpha$$

Cette expression signifie que l'Expected Shortfall correspond à la moyenne arithmétique des quantiles de niveau supérieur à  $\beta$ , c'est-à-dire la moyenne des pires pertes possibles, pondérées de façon uniforme sur la queue de distribution allant de  $\beta$  à 1. Contrairement à la VaR qui ne regarde que la valeur du quantile à un seul niveau  $\beta$ , l'Expected Shortfall intègre toute la forme de la queue gauche de la distribution (les pertes extrêmes).

On considère que  $L_t(h)$  a une fonction de répartition  $F_L(x)$ . Par changement de variable, on peut exprimer l'espérance conditionnelle :

$$\text{ES}_\beta = \mathbb{E}[L_t(h) \mid L_t(h) \geq \text{VaR}_\beta] = \frac{1}{1 - \beta} \int_{\text{VaR}_\beta}^{\infty} x f_L(x) dx$$

où  $f_L(x) = F'_L(x)$  est la densité. Par inversion de la fonction de répartition (en utilisant la formule de quantile inverse), on peut réécrire cette intégrale comme :

$$\text{ES}_\beta = \frac{1}{1 - \beta} \int_\beta^1 \text{VaR}_\alpha d\alpha$$

ce qui donne la **formule intégrale de l'Expected Shortfall**, fondamentalement plus robuste que la VaR.

L'une des modifications méthodologiques majeures introduites par le FRTB réside dans l'abandon progressif de la Value-at-Risk (VaR) au profit de l'Expected Shortfall (ES) comme mesure de risque réglementaire. Cette substitution n'est pas anodine : elle traduit un basculement vers une mesure cohérente, plus sensible aux pertes extrêmes, et mieux adaptée aux dynamiques de marché en période de stress. Concrètement, le texte prévoit que les établissements bancaires calculent l'ES au niveau de confiance de 97,5%, en lieu et place de la VaR au seuil de 99% précédemment utilisée. Cette modification du niveau de confiance vise à mieux capturer les queues de distribution, en particulier dans les portefeuilles sensibles aux risques rares mais violents. Par ailleurs, le FRTB introduit une granularité supplémentaire à travers la prise en compte de horizons de liquidité variables. Chaque classe d'actifs se voit attribuer un horizon temporel spécifique (10, 20, 40, voire

120 jours) en fonction de son profil de liquidité en conditions de stress. Cette approche multi-horizons permet de mieux ajuster le capital exigé au comportement réel des instruments financiers en cas de tensions sur les marchés. Toutefois, la mise en œuvre pratique de ces nouvelles exigences soulève des défis considérables en matière de charge de calcul, notamment pour les portefeuilles contenant des produits dérivés complexes. Le passage de la VaR à l'ES, combiné à la simulation de scénarios sous contraintes de liquidité, implique un volume de revalorisations beaucoup plus important, rendant crucial le recours à des méthodes numériques avancées ou à des modèles d'approximation efficaces.

Quelle que soit la méthode numérique adoptée, qu'il s'agisse de simulations historiques ou de simulations de Monte Carlo, l'évaluation répétée d'un portefeuille complet de produits dérivés à travers une multitude de scénarios de risque et pour différents horizons temporels se révèle particulièrement coûteuse sur le plan computationnel. Cette charge est d'autant plus lourde que les portefeuilles sont complexes et fortement non linéaires, impliquant des recalculs fréquents de valorisations dans des environnements à haute dimension.

Face à cette contrainte, les institutions financières doivent impérativement optimiser leurs temps de calcul, sans compromettre la précision des estimations de risque. Cela nécessite la mise en œuvre de stratégies innovantes permettant d'atteindre un équilibre robuste entre vitesse d'exécution et fidélité des résultats.

Dans cette optique, le présent travail explore le potentiel des techniques d'apprentissage automatique comme outil d'accélération. Avant d'introduire ces approches, il est utile de rappeler brièvement les principales méthodes numériques conventionnelles employées pour le calcul de la Value-at-Risk (VaR) et de l'Expected Shortfall (ES), qui servent de socle aux indicateurs de risque réglementaire.

## 2.1 Conception d'un modèle de risque par simulations

Les principales méthodes numériques utilisées pour la mise en œuvre des modèles de risque, notamment pour le calcul de la Value-at-Risk (VaR) et de l'Expected Shortfall (ES), s'appuient sur deux grandes familles d'approches fondées sur la simulation : **les simulations historiques** et **les simulations de Monte Carlo**. Ces techniques reposent sur la génération de scénarios de marché permettant d'évaluer la distribution potentielle des pertes d'un portefeuille donné.

L'approche dite historique est une méthode non paramétrique d'estimation du risque, dont l'originalité réside dans son absence totale d'hypothèse structurelle sur la distribution des rendements. Elle repose sur l'idée que les réalisations passées des variables de marché contiennent une information suffisante pour modéliser les risques futurs. Concrètement, on considère une fenêtre temporelle glissante (par exemple 252 jours), et l'on applique les chocs observés historiquement aux valorisations actuelles du portefeuille. Cette méthode repose donc sur l'hypothèse que les rendements passés sont représentatifs de ce que l'on peut attendre à l'avenir, ce qui constitue une hypothèse forte mais opérationnellement séduisante. Elle permet en effet une mise en œuvre simple, directe, sans besoin d'estimer la volatilité, la structure de corrélation entre les actifs, ou encore les propriétés de queue des distributions. Cette absence d'hypothèse distributionnelle rend la méthode particulièrement adaptée aux contextes non gaussiens, où les rendements présentent des queues épaisses ou des asymétries, caractéristiques fréquentes des marchés financiers en période de stress. Un autre avantage essentiel est que cette approche est compatible avec des portefeuilles complexes, contenant par exemple des options exotiques, des swaps ou des instruments à valorisation non linéaire. En effet, l'approche repose généralement sur une revalorisation complète (*full revaluation*) du portefeuille à chaque scénario simulé. Cette propriété garantit une meilleure fidélité des estimations, en tenant compte des non-linéarités inhérentes à certains instruments dérivés. Malgré ses nombreux atouts, l'approche par simulations historiques pré-

sente également plusieurs limites structurelles qui en restreignent la portée dans des contextes de gestion du risque avancée. Tout d'abord, cette méthode repose sur une forte dépendance aux données historiques : toute erreur de mesure, discontinuité dans les séries ou biais de sélection est directement transmis dans l'estimation des pertes potentielles. En outre, un changement structurel des conditions de marché, tel qu'une crise financière, une transition monétaire ou un choc géopolitique majeur, peut rendre les données passées inopérantes pour modéliser l'avenir, sapant ainsi la pertinence des simulations. Ensuite, l'approche repose implicitement sur une hypothèse d'homogénéité temporelle, en supposant que les rendements suivent une distribution stationnaire dans le temps. Cette hypothèse est souvent contredite par la réalité financière, caractérisée par des régimes de volatilité changeants, des effets de mémoire ou des phénomènes de clustering. En période de stress, cela peut conduire à une sous-estimation significative du risque, notamment si l'échantillon historique inclut peu de périodes extrêmes. Par ailleurs, la faible granularité des événements rares constitue une limite importante. Puisque les simulations sont contraintes à l'échantillon historique disponible, les événements extrêmes, dits « *black swans* », peuvent ne pas être représentés, sauf à disposer de séries exceptionnellement longues. Cela pose un dilemme entre la représentativité statistique et l'actualité des données. Enfin, cette approche peine à intégrer des horizons de liquidité variables selon les instruments, pourtant exigés dans le cadre réglementaire du FRTB. Contrairement aux simulations de Monte Carlo, plus flexibles, la méthodologie historique ne permet pas facilement de distinguer les délais de liquidation propres à chaque classe d'actifs, ce qui constitue un frein à son adoption dans des environnements de trading complexes et régulés.

La méthode des simulations historiques repose sur une reconstitution empirique des pertes potentielles à partir de scénarios observés dans le passé. Elle se déroule selon une série d'étapes structurées permettant d'estimer la Value-at-Risk (VaR) et l'Expected Shortfall (ES) d'un portefeuille donné.

Dans un premier temps, on calcule les rendements historiques ou les variations de prix de tous les actifs composant le portefeuille sur des intervalles de temps successifs. Ces rendements, notés  $R_t := \frac{P_{t+h} - P_t}{P_t}$  ou sous forme de variations absolues  $\Delta P_t := P_{t+h} - P_t$ , sont extraits d'une fenêtre historique couvrant une période représentative de l'activité de marché récente. Ce jeu de données constitue une base empirique de chocs de marché.

Dans un second temps, on applique chacun de ces chocs historiques  $X_t := Z_{t+h} - Z_t$ , représentant des variations observées des facteurs de risque, aux valeurs actuelles de marché  $V_t$  des instruments du portefeuille. Cela implique de revaloriser l'ensemble du portefeuille sous chacun des scénarios simulés, en tenant compte des nouvelles conditions de marché. Chaque scénario historique donne ainsi lieu à une nouvelle valorisation  $V_{t+h}$ , ce qui permet de calculer les pertes ou gains associés.

Une fois tous les scénarios évalués, on détermine la série des pertes simulées selon la formule  $L^{(i)} := -(V_{t+h}^{(i)} - V_t)$ . Cette série est ensuite triée de la plus grande perte à la plus petite, formant une distribution empirique de profits et pertes. C'est cette distribution qui sert de fondement à l'estimation des mesures de risque.

La Value-at-Risk est obtenue en identifiant le quantile empirique de niveau  $\beta$ , correspondant au niveau de risque résiduel. Par exemple, pour une VaR à 99%, on retient le quantile au niveau  $\beta = 1\%$ . Formellement, on note :

$$\text{VaR}_\beta \approx \text{P\&L}_{(\lceil N \cdot \beta \rceil)},$$

où  $N$  désigne le nombre total de scénarios historiques.

Enfin, l'Expected Shortfall est estimée comme la moyenne des pires pertes, c'est-à-dire la moyenne

des  $k$  plus grandes pertes de la distribution simulée, avec  $k = \lfloor N \cdot \beta \rfloor$ . Elle s'écrit :

$$\text{ES}_\beta \approx \frac{1}{k} \sum_{i=1}^k \text{P\&L}_{(i)}.$$

Cette mesure est plus robuste que la VaR car elle tient compte non seulement du seuil extrême, mais aussi de l'ensemble des pertes situées dans la queue de distribution au-delà de ce seuil.

Ensuite, l'approche par simulations de Monte Carlo constitue l'une des méthodologies les plus robustes, souples et théoriquement fondées pour l'évaluation du risque de marché à travers des mesures comme la Value-at-Risk (VaR) et l'Expected Shortfall (ES). Cette méthode repose sur une logique de simulation aléatoire des facteurs de risque et permet de capturer avec précision les dynamiques complexes des portefeuilles financiers, notamment ceux contenant des instruments non linéaires (options, dérivés exotiques, produits structurés...). Contrairement aux simulations historiques, qui utilisent directement les variations observées sur les marchés passés, l'approche Monte Carlo simule un grand nombre de trajectoires hypothétiques pour les facteurs de risque, en générant aléatoirement des scénarios futurs compatibles avec des hypothèses probabilistes. Typiquement, ces scénarios sont générés à partir de modèles stochastiques (mouvements brownien géométrique, processus de diffusion avec sauts, modèles GARCH, etc.) calibrés sur des données historiques, mais avec une liberté accrue pour modéliser des régimes de volatilité variables, des corrélations dynamiques, ou des comportements extrêmes.

La **première étape** de la méthode consiste à spécifier un modèle de dynamique stochastique pour les facteurs de risque sous-jacents au portefeuille. Ces facteurs peuvent inclure les taux d'intérêt, les prix d'actions, les volatilités implicites, les spreads de crédit ou encore les taux de change. On suppose généralement que chaque facteur  $Z_t$  évolue selon une équation différentielle stochastique du type  $dZ_t = \mu(Z_t, t)dt + \sigma(Z_t, t)dW_t$ , où  $\mu$  et  $\sigma$  désignent respectivement les termes de drift et de diffusion, et  $W_t$  est un mouvement brownien multivarié, éventuellement enrichi pour tenir compte de sauts, de volatilités stochastiques ou de corrélations dynamiques. Le choix de la forme fonctionnelle du modèle est crucial car il conditionne la qualité des scénarios de risque qui seront générés.

Une fois le modèle spécifié et calibré sur les données de marché, la **seconde étape** consiste à générer un grand nombre  $N$  de trajectoires simulées des facteurs de risque jusqu'à l'horizon  $t + h$ . Cela se fait généralement par discrétisation des équations différentielles stochastiques à l'aide de schémas de type *Euler-Maruyama* ou *Milstein*. Pour chaque simulation, on échantillonne un vecteur de valeurs futures  $Z_{t+h}^{(i)}$ , en veillant à conserver la structure de dépendance entre les différents facteurs, notamment via une matrice de corrélation empirique ou un modèle de copule.

La **troisième étape** consiste à procéder à la revalorisation complète du portefeuille sous chaque scénario simulé. Cela implique de recalculer la valeur du portefeuille à l'horizon  $t + h$  pour chaque vecteur  $Z_{t+h}^{(i)}$ , en appliquant la fonction de valorisation  $f$  spécifique à chaque instrument financier contenu dans le portefeuille. On obtient ainsi une série de valeurs futures  $V_{t+h}^{(i)}$ , à partir desquelles on calcule les pertes simulées par  $L_t^{(i)}(h) = -(V_{t+h}^{(i)} - V_t)$ , où  $V_t$  est la valeur actuelle du portefeuille. Cette étape est la plus intensive du point de vue computationnel, surtout si la fonction de pricing repose elle-même sur des simulations internes.

Une fois l'ensemble des pertes simulées obtenu, on construit la distribution empirique des pertes à partir de l'échantillon  $\{L_t^{(1)}(h), \dots, L_t^{(N)}(h)\}$ . Cette distribution représente l'ensemble des scénarios de pertes que le portefeuille pourrait subir en fonction des configurations simulées du marché. L'estimation empirique de la fonction de répartition permet alors d'accéder aux mesures de risque classiques. La Value-at-Risk au niveau  $\beta$  est obtenue comme le quantile empirique d'ordre  $\beta$ , c'est-à-dire la valeur telle que  $\beta\%$  des pertes soient inférieures ou égales à celle-ci.



Enfin, l'Expected Shortfall (ES) est calculé comme la moyenne des pertes qui excèdent ce seuil de quantile. Cette mesure permet d'appréhender l'ampleur des pertes en cas de dépassement du niveau critique fixé par la VaR, et fournit ainsi une meilleure estimation du risque dans les queues extrêmes de la distribution simulée. Grâce à cette méthode, les institutions financières disposent d'un outil puissant pour quantifier, anticiper et gérer les risques de marché dans des conditions de marché complexes et dynamiques.

## 2.2 VaR Paramétrique et Ajustement de Cornish-Fisher

La méthode **delta-gamma** constitue une approche paramétrique élégante et largement utilisée pour estimer la Value-at-Risk (VaR) d'un portefeuille contenant des produits dérivés, notamment des options. Son principal avantage est de fournir une formule fermée, évitant ainsi les lourdeurs numériques des simulations historiques ou de Monte Carlo. Elle s'appuie sur une *approximation quadratique* de la variation de valeur du portefeuille, en considérant à la fois la sensibilité de premier ordre (*delta*) et de second ordre (*gamma*) aux variations du facteur de risque principal, généralement le prix de l'actif sous-jacent.

Considérons un portefeuille de  $N$  options sur un même actif sous-jacent  $S_t$ , avec des pondérations respectives  $m_i$ . La variation de valeur du portefeuille sur un horizon de risque  $h$  est notée  $\Delta V_t = V_{t+h} - V_t$ . En supposant que la variation du sous-jacent  $S_t$  est modélisée comme un taux de rendement  $R_t = \frac{S_{t+h} - S_t}{S_t}$ , on peut approximer la variation de valeur du portefeuille par une expansion de Taylor à l'ordre 2. En effet, le développement de Taylor d'ordre 2 de la fonction  $V(S)$  autour du point  $S_t$  s'écrit :

$$V(S_{t+h}) = V(S_t) + \frac{\partial V}{\partial S}(S_t) \cdot \Delta S_t + \frac{1}{2} \cdot \frac{\partial^2 V}{\partial S^2}(S_t) \cdot (\Delta S_t)^2 + o((\Delta S_t)^2)$$

Dans le contexte d'une approximation à court terme, on néglige les termes d'ordre supérieur, ce qui donne :

$$V_{t+h} \approx V_t + \Delta \cdot \Delta S_t + \frac{1}{2} \cdot \Gamma \cdot (\Delta S_t)^2$$

où  $\Delta = \frac{\partial V}{\partial S}(S_t)$  est la sensibilité première (delta) du portefeuille, et  $\Gamma = \frac{\partial^2 V}{\partial S^2}(S_t)$  est la sensibilité seconde (gamma). En soustrayant  $V_t$  des deux côtés, on obtient une approximation quadratique de la variation de la valeur du portefeuille :

$$\Delta V_t = V_{t+h} - V_t \approx \Delta \cdot \Delta S_t + \frac{1}{2} \cdot \Gamma \cdot (\Delta S_t)^2$$

On souhaite maintenant réécrire cette expression en fonction du rendement relatif  $R_t = \frac{\Delta S_t}{S_t}$ . En effet, ce rendement est souvent plus stable que les niveaux de prix absolus, et il permet une expression normalisée de la variation. Ainsi, on remplace  $\Delta S_t$  par  $S_t R_t$  dans l'équation précédente :

$$\Delta V_t \approx \Delta \cdot S_t \cdot R_t + \frac{1}{2} \cdot \Gamma \cdot (S_t \cdot R_t)^2$$

Ce qui donne finalement :

$$\Delta V_t \approx \Delta S_t \cdot R_t + \frac{1}{2} \cdot \Gamma \cdot S_t^2 \cdot R_t^2$$

où le delta ( $\Delta$ ) et le gamma ( $\Gamma$ ) sont données par :

$$\Delta = \sum_{j=1}^N m_j \Delta_j, \quad \Gamma = \sum_{j=1}^N m_j \Gamma_j$$

Cette approximation tient compte non seulement du déplacement moyen du sous-jacent (via le terme en  $R_t$ ), mais aussi de la convexité induite par la structure des options (via  $R_t^2$ ). Cela

permet de mieux refléter le comportement non linéaire du portefeuille, notamment en présence d'options proches de la monnaie ou à forte convexité. La distribution de  $\Delta V_t$ , sous cette approximation, n'est en général plus strictement normale à cause du terme quadratique. C'est pourquoi on utilise **l'approximation de Cornish-Fisher**, qui ajuste les quantiles d'une distribution normale standard en tenant compte de ses coefficients d'asymétrie (*skewness*) et *kurtosis* (aplatissement). Cette correction permet d'obtenir une approximation plus fidèle de la VaR, même lorsque la distribution des rendements est asymétrique ou possède des queues épaisses, comme c'est fréquemment le cas en finance.

Selon *Pichler et Selitsch*, lorsqu'on modélise le P&L d'un portefeuille en n'utilisant qu'un seul facteur de risque  $S_t$  et en ayant recours à une approximation quadratique de la fonction de valorisation via le développement de Taylor à l'ordre 2 (delta-gamma), la distribution de  $\Delta V_t$  ne suit plus une loi normale. Elle devient une loi du  $\chi^2$  non centrée, c'est-à-dire une loi asymétrique, dont les moments ne sont plus ceux d'une gaussienne standard. Ainsi, les trois premiers moments de cette distribution peuvent être déduits, en supposant que  $R_t$  suit une loi normale de moyenne nulle et de volatilité  $\sigma$  :

$$\mu_V = \mathbb{E}[\Delta V_t] = \frac{1}{2} \Gamma S_t^2 \sigma^2 h$$

Puisque qu'on suppose que le portefeuille est composé d'un instrument dérivé dépendant d'un unique actif sous-jacent  $S_t$ , et que la variation de sa valeur à l'horizon  $t + h$  est approchée par la formule de Taylor du second ordre (approximation delta-gamma) :

$$\Delta V_t \approx \Delta S_t \cdot R_t + \frac{1}{2} \Gamma S_t^2 R_t^2$$

Le terme linéaire  $\Delta S_t \cdot R_t$  a une espérance nulle car  $R_t$  est une variable normale centrée. Par linéarité de l'espérance, on obtient donc :

$$\mathbb{E}[\Delta V_t] = \mathbb{E} \left[ \Delta S_t \cdot R_t + \frac{1}{2} \Gamma S_t^2 R_t^2 \right] = \frac{1}{2} \Gamma S_t^2 \cdot \mathbb{E}[R_t^2]$$

Or, puisque  $R_t \sim \mathcal{N}(0, \sigma^2 h)$ , on a  $\mathbb{E}[R_t^2] = \text{Var}(R_t) = \sigma^2 h$ . Ainsi, on conclut que :

$$\mu_V = \mathbb{E}[\Delta V_t] = \frac{1}{2} \Gamma S_t^2 \sigma^2 h$$

Ce terme est directement lié à la convexité du portefeuille (représentée par  $\Gamma$ ). En effet, même si l'espérance du rendement  $R_t$  est nulle (hypothèse de martingale), le terme quadratique génère une contribution systématique au P&L moyen du portefeuille.

Cela reflète le fait que, pour un portefeuille ayant une convexité positive (par exemple, une option d'achat), les gains potentiels en cas de forte variation (dans un sens ou dans l'autre) dépassent les pertes. Cela introduit donc un biais positif dans la distribution du P&L, même en l'absence de drift sur l'actif sous-jacent.

La variance du P&L, notée  $\text{Var}(\Delta V_t)$ , mesure la dispersion des gains et pertes autour de leur valeur moyenne. Sous l'approximation delta-gamma, on dispose d'une expression explicite du moment d'ordre 2 :

$$\mathbb{E}[(\Delta V_t)^2] = \Delta^2 S_t^2 \sigma^2 h + \frac{3}{4} \Gamma^2 S_t^4 \sigma^4 h^2$$

En effet, soit  $\Delta V_t$  le P&L du portefeuille à l'horizon  $h$ , exprimé sous l'approximation delta-gamma :

$$\Delta V_t \approx \Delta S_t \cdot R_t + \frac{1}{2} \Gamma S_t^2 R_t^2$$

Notons :

- $\mu = \mathbb{E}[R_t] = 0$ , par hypothèse de martingale sur le rendement.
- $\mathbb{E}[R_t^2] = \text{Var}(R_t) = \sigma^2 h$ , variance d'un rendement normal centré.
- $\mathbb{E}[R_t^4] = 3(\sigma^2 h)^2$ , qui est le moment d'ordre 4 d'une variable normale centrée. On utilise la formule générale du moment d'ordre pair pour une loi normale centrée, donnée par :

$$\mathbb{E}[X^{2n}] = (2n - 1)!! \cdot \sigma^{2n}$$

où  $(2n - 1)!!$  désigne la double factorielle impaire, et  $\sigma^2$  la variance.

On cherche à calculer :

$$\mathbb{E}[(\Delta V_t)^2] = \mathbb{E} \left[ \left( \Delta S_t R_t + \frac{1}{2} \Gamma S_t^2 R_t^2 \right)^2 \right]$$

Développons le carré :

$$= \mathbb{E} \left[ \Delta^2 S_t^2 R_t^2 + \Delta \Gamma S_t^3 R_t^3 + \frac{1}{4} \Gamma^2 S_t^4 R_t^4 \right]$$

Par linéarité de l'espérance :

$$\mathbb{E}[(\Delta V_t)^2] = \Delta^2 S_t^2 \mathbb{E}[R_t^2] + \Delta \Gamma S_t^3 \mathbb{E}[R_t^3] + \frac{1}{4} \Gamma^2 S_t^4 \mathbb{E}[R_t^4]$$

On en déduit :

$$\mathbb{E}[(\Delta V_t)^2] = \Delta^2 S_t^2 \sigma^2 h + \frac{3}{4} \Gamma^2 S_t^4 \sigma^4 h^2$$

Cette expression repose notamment sur le fait que  $\mathbb{E}[R_t^3] = 0$ , ce qui découle de la symétrie de la distribution normale (skewness nulle). Ce qui donne l'expression du moment d'ordre 2 sous l'approximation quadratique.

Ce moment d'ordre 2 combine deux sources distinctes de risque. Le premier terme,  $\Delta^2 S_t^2 \sigma^2 h$ , reflète le risque directionnel du portefeuille. Il est directement proportionnel au carré du delta, et capture la sensibilité linéaire du portefeuille aux petites variations du sous-jacent. Plus le portefeuille est directionnel, plus ce terme est élevé, traduisant une forte exposition au bruit de marché.

Le second terme,  $\frac{3}{4} \Gamma^2 S_t^4 \sigma^4 h^2$ , correspond à l'impact de la convexité du portefeuille. Il quantifie la contribution non linéaire des variations du sous-jacent à la dispersion des résultats, et devient prépondérant pour des portefeuilles fortement gamma-sensibles (comme les options ou produits dérivés complexes). Ce terme croît très rapidement avec la volatilité ( $\sigma^4$ ) et l'horizon de temps ( $h^2$ ), soulignant l'importance du gamma en période de stress de marché ou pour des positions tenues sur des horizons longs.

Enfin,

$$\mathbb{E}[(\Delta V_t)^3] = \frac{9}{2} \Delta^2 \Gamma S_t^4 \sigma^4 h^2 + \Gamma^3 S_t^6 \sigma^6 h^3$$

$$(\Delta V_t)^3 = \left( \Delta S_t \cdot \Delta + \frac{1}{2} \Gamma S_t^2 R_t^2 \right)^3 = \left( S_t \Delta R_t + \frac{1}{2} \Gamma S_t^2 R_t^2 \right)^3$$

On développe cette expression à l'aide de l'identité algébrique  $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$ . En posant  $a = \Delta S_t \cdot \Delta = S_t \Delta R_t$  et  $b = \frac{1}{2} \Gamma S_t^2 R_t^2$ , on obtient :

$$(\Delta V_t)^3 = (S_t \Delta R_t)^3 + 3(S_t \Delta R_t)^2 \cdot \frac{1}{2} \Gamma S_t^2 R_t^2 + 3(S_t \Delta R_t) \cdot \left( \frac{1}{2} \Gamma S_t^2 R_t^2 \right)^2 + \left( \frac{1}{2} \Gamma S_t^2 R_t^2 \right)^3$$

On calcule ensuite l'espérance de chaque terme. Étant donné que  $R_t \sim \mathcal{N}(0, \sigma^2 h)$ , on utilise les moments classiques d'une loi normale centrée :

$$\begin{aligned} \mathbb{E}[R_t^3] &= 0 \\ \mathbb{E}[R_t^4] &= 3(\sigma^2 h)^2 \\ \mathbb{E}[R_t^6] &= 15(\sigma^2 h)^3 \end{aligned}$$

Ainsi, seuls les termes comportant une puissance paire de  $R_t$  admettent une espérance non nulle. Pour le premier terme, on a  $\mathbb{E}[(S_t \Delta R_t)^3] = S_t^3 \Delta^3 \mathbb{E}[R_t^3] = 0$ . Pour le second, on obtient :

$$3 \cdot \mathbb{E}[(S_t \Delta R_t)^2 \cdot \frac{1}{2} \Gamma S_t^2 R_t^2] = 3 \cdot \frac{1}{2} \Gamma S_t^4 \Delta^2 \mathbb{E}[R_t^4] = \frac{3}{2} \Gamma S_t^4 \Delta^2 \cdot 3(\sigma^2 h)^2 = \frac{9}{2} \Delta^2 \Gamma S_t^4 \sigma^4 h^2$$

Le troisième terme s'écrit :

$$3 \cdot \mathbb{E}[S_t \Delta R_t \cdot \left( \frac{1}{2} \Gamma S_t^2 R_t^2 \right)^2] = 3 \cdot S_t \Delta \cdot \left( \frac{1}{2} \Gamma S_t^2 \right)^2 \cdot \mathbb{E}[R_t^5] = 0$$

car  $\mathbb{E}[R_t^5] = 0$  pour une loi normale centrée. Enfin, le quatrième terme donne :

$$\mathbb{E} \left[ \left( \frac{1}{2} \Gamma S_t^2 R_t^2 \right)^3 \right] = \left( \frac{1}{2} \Gamma S_t^2 \right)^3 \cdot \mathbb{E}[R_t^6] = \frac{1}{8} \Gamma^3 S_t^6 \cdot 15(\sigma^2 h)^3 = \frac{15}{8} \Gamma^3 S_t^6 \sigma^6 h^3$$

En regroupant les deux contributions non nulles, on obtient :

$$\mathbb{E}[(\Delta V_t)^3] = \frac{9}{2} \Delta^2 \Gamma S_t^4 \sigma^4 h^2 + \frac{15}{8} \Gamma^3 S_t^6 \sigma^6 h^3$$

Ce troisième moment met en évidence l'asymétrie potentielle du P&L induite par les effets de convexité ( $\Gamma$ ) : même si la distribution de  $R_t$  est symétrique, les termes croisés entre le delta et le gamma génèrent un moment impair non nul.

Il est aisé d'affirmer que :

$$\mu_V = \mathbb{E}[\Delta V_t] = \frac{1}{2} \Gamma S_t^2 \sigma^2 h$$

$$\sigma_V^2 = \mathbb{E}[(\Delta V_t)^2] - (\mathbb{E}[\Delta V_t])^2 = \Delta^2 S_t^2 \sigma^2 h + \frac{1}{2} \Gamma^2 S_t^4 \sigma^4 h^2$$

Pour trouver le coefficient d'asymétrie (skewness)  $\xi_V$  de la distribution de probabilité de  $\Delta V_t$  on procède ainsi :

On cherche à calculer le coefficient d'asymétrie :

$$\xi_V = \frac{\mathbb{E}[(\Delta V_t - \mu_V)^3]}{\sigma_V^3}$$

Puisque  $\mathbb{E}[(\Delta V_t)^3] = \frac{9}{2}\Delta^2\Gamma S_t^4\sigma^4h^2 + \frac{15}{8}\Gamma^3S_t^6\sigma^6h^3$  et que  $\mu_V = \frac{1}{2}\Gamma S_t^2\sigma^2h$ , on a :

$$\begin{aligned}\mathbb{E}[(\Delta V_t - \mu_V)^3] &= \mathbb{E}[(\Delta V_t)^3] - 3\mu_V\sigma_V^2 + 2\mu_V^3 \\ &= \left(\frac{9}{2}\Delta^2\Gamma S_t^4\sigma^4h^2 + \frac{15}{8}\Gamma^3S_t^6\sigma^6h^3\right) \\ &\quad - 3 \cdot \left(\frac{1}{2}\Gamma S_t^2\sigma^2h\right) \cdot \left(\Delta^2S_t^2\sigma^2h + \frac{3}{4}\Gamma^2S_t^4\sigma^4h^2\right) \\ &\quad + 2 \cdot \left(\frac{1}{2}\Gamma S_t^2\sigma^2h\right)^3 \\ &= \left(\frac{9}{2}\Delta^2\Gamma S_t^4\sigma^4h^2 + \frac{15}{8}\Gamma^3S_t^6\sigma^6h^3\right) - \left(\frac{3}{2}\Delta^2\Gamma S_t^4\sigma^4h^2 + \frac{9}{8}\Gamma^3S_t^6\sigma^6h^3\right) + \frac{1}{4}\Gamma^3S_t^6\sigma^6h^3 \\ &= 3\Delta^2\Gamma S_t^4\sigma^4h^2 + \Gamma^3S_t^6\sigma^6h^3\end{aligned}$$

Finalement,

$$\xi_V = \frac{3\Delta^2\Gamma S_t^4\sigma^4h^2 + \Gamma^3S_t^6\sigma^6h^3}{\sigma_V^3}$$

La méthode paramétrique pour le calcul de la Value-at-Risk (VaR) du portefeuille repose sur l'expansion de Cornish-Fisher, qui permet d'obtenir une approximation plus précise du quantile d'une distribution non normale en utilisant les moments d'ordre supérieur de la distribution de la variable aléatoire en question. Dans le cadre de la VaR, on considère la distribution des variations de la valeur du portefeuille, notée  $\Delta V_t$ , et l'objectif est d'estimer le quantile à un niveau de confiance donné, appelé le  $\beta$ -percentile, noté  $\theta_\beta$ .

L'expansion de Cornish-Fisher est une méthode qui repose sur les trois premiers moments de la distribution, à savoir l'espérance (moyenne), la variance (dispersion), et l'asymétrie (skewness). En utilisant ces trois premiers moments, l'expansion permet d'approximer les quantiles de la distribution même lorsque celle-ci présente des caractéristiques non normales, comme une asymétrie ou des queues épaisses.

Plus précisément, l'estimation du  $\beta$ -percentile,  $\theta_\beta$ , est obtenue en ajoutant un terme correcteur à la valeur du quantile d'une distribution normale standard. Ce terme correcteur tient compte des caractéristiques non normales de la distribution de  $\Delta V_t$ , telles que l'asymétrie et la kurtose. La formule de l'expansion de Cornish-Fisher pour le quantile  $\theta_\beta$  est donnée par :

$$\theta_\beta = \alpha_\beta + \frac{1}{6}(\alpha_\beta^2 - 1)\xi_V + \frac{1}{24}(\alpha_\beta^3 - 3\alpha_\beta)\kappa_V - \frac{1}{36}(2\alpha_\beta^3 - 5\alpha_\beta)\xi_V^2$$

est approximativement égale à :

$$\theta_\beta \approx \alpha_\beta + \frac{1}{6}(\alpha_\beta^2 - 1)\xi_V$$

où  $\alpha_\beta$  désigne le quantile d'ordre  $\beta$  de la loi normale standard.

Cette approximation repose sur l'idée que les termes correctifs d'ordre supérieur dans l'expansion de Cornish-Fisher peuvent être négligés dans certaines situations spécifiques. En particulier, lorsque la kurtosis excédentaire  $\kappa_V$  est proche de zéro, ce qui est généralement le cas pour des distributions faiblement leptokurtiques ou mésokurtiques, le terme impliquant la kurtosis devient négligeable. De même, si la skewness  $\xi_V$  est suffisamment faible, alors son carré  $\xi_V^2$  l'est encore davantage, rendant le troisième terme correctif de l'expansion insignifiant.

L'objectif de cette approximation est d'estimer la *Value-at-Risk* d'un portefeuille de dérivés, dont la distribution du P&L  $\Delta V_t$  s'écarte de la normalité en raison de sa *skewness* (asymétrie)  $\xi_V$ . En

effet, dans les approches classiques, la VaR est calculée sous l'hypothèse que  $\Delta V_t \sim \mathcal{N}(\mu_V, \sigma_V^2)$ , auquel cas la VaR au niveau de risque  $\beta \in (0, 1)$  est donnée par :

$$\text{VaR}_\beta = \mu_V + \alpha_\beta \cdot \sigma_V$$

où  $\alpha_\beta = \Phi^{-1}(1 - \beta)$  est le quantile de niveau  $1 - \beta$  de la loi normale standard.

Cependant, lorsque la distribution de  $\Delta V_t$  présente une skewness non nulle, cette formule sous-estime ou surestime les risques extrêmes selon le signe de l'asymétrie. Pour corriger ce biais, on utilise une expansion de Cornish-Fisher qui permet de corriger le quantile  $\alpha_\beta$  de la loi normale standard en tenant compte des moments d'ordre supérieur (ici, la skewness).

On retrouve donc la VaR ajustée sous la forme :

$$\text{VaR}_\beta = \mu_V + \theta_\beta \cdot \sigma_V = \mu_V + \left( \alpha_\beta + \frac{1}{6}(\alpha_\beta^2 - 1)\xi_V \right) \sigma_V$$

Lorsque le portefeuille est exposé à plusieurs facteurs de risque  $X_1, X_2, \dots, X_m$ , il est courant d'assimiler sa valeur à un agrégat de composantes individuelles dépendant chacune de l'un de ces facteurs. Formellement, on suppose que :

$$V_t = \sum_{i=1}^m V_i(X_i)$$

où chaque  $V_i(X_i)$  représente la contribution marginale au portefeuille due au facteur de risque  $X_i$  à l'instant  $t$ . Cette décomposition est particulièrement pertinente lorsque les instruments du portefeuille (par exemple, les options) sont principalement sensibles à un sous-ensemble réduit de facteurs.

Pour étudier la variation de la valeur du portefeuille sur l'horizon  $[t, t + h]$ , on introduit le rendement du facteur de risque  $X_i$  sous la forme :

$$R_i := \frac{X_i(t + h) - X_i(t)}{X_i(t)}$$

Ce rendement mesure la variation relative du facteur  $X_i$  entre les instants  $t$  et  $t + h$ , ce qui est crucial pour évaluer l'impact marginal de chaque facteur sur la valorisation.

Sous l'approximation delta-gamma, on exprime alors la variation du portefeuille  $\Delta V_t := V_{t+h} - V_t$  par un développement limité au second ordre (analogue à une expansion de Taylor à deux termes pour chaque facteur) :

$$\Delta V_t \approx \sum_{i=1}^m \Delta_i X_i R_i + \sum_{i=1}^m \frac{1}{2} \Gamma_i X_i^2 R_i^2$$

Le premier terme représente l'effet linéaire des variations des facteurs sur le portefeuille, via les sensibilités de type delta  $\Delta_i$ , tandis que le second terme capture les effets de convexité (ou non-linéarité) via les sensibilités gamma  $\Gamma_i$ .

Dans le cadre multivarié, on considère un portefeuille de dérivés dépendant de  $m$  facteurs de risque, notés  $X_1, \dots, X_m$ , représentant par exemple des taux d'intérêt, des volatilités implicites ou des spreads de crédit. La variation relative de chaque facteur est modélisée par un rendement

$$R_i := \frac{X_i(t + h) - X_i(t)}{X_i(t)},$$

et on note  $\Delta_i$  et  $\Gamma_i$  respectivement les sensibilités de premier et de second ordre (delta et gamma) du portefeuille par rapport au facteur  $X_i$ .

Pour capturer les interdépendances entre ces facteurs, on introduit la matrice de covariance  $\Sigma \in \mathbb{R}^{m \times m}$ , définie par :

$$\Sigma = \begin{pmatrix} \sigma^2(R_1)h & \cdots & \text{Cov}(R_1, R_m)h \\ \vdots & \ddots & \vdots \\ \text{Cov}(R_m, R_1)h & \cdots & \sigma^2(R_m)h \end{pmatrix}$$

et qui est multipliée par l'horizon  $h$  pour refléter la croissance temporelle de la variance. On regroupe ensuite les sensibilités directionnelles dans le vecteur :

$$\delta := \begin{pmatrix} \Delta_1 X_1 \\ \Delta_2 X_2 \\ \vdots \\ \Delta_m X_m \end{pmatrix} \in \mathbb{R}^m,$$

et les sensibilités convexes dans une matrice diagonale :

$$\Gamma := \begin{pmatrix} \Gamma_1 & 0 & \cdots & 0 \\ 0 & \Gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Gamma_m \end{pmatrix}.$$

La variation de la valeur du portefeuille  $\Delta V_t$  sur l'horizon  $h$  est alors approchée par une expansion de type Taylor à deux termes (delta-gamma) :

$$\Delta V_t \approx \sum_{i=1}^m \Delta_i X_i R_i + \sum_{i=1}^m \frac{1}{2} \Gamma_i X_i^2 R_i^2.$$

Les moments de la distribution de  $\Delta V_t$  sont ensuite calculés de la façon suivante :

— **Espérance** :

$$\mu_V = \mathbb{E}[\Delta V_t] = \frac{1}{2} \text{tr}[\Gamma \Sigma],$$

qui représente l'impact moyen de la convexité du portefeuille sur le P&L attendu.

— **Variance** :

$$\sigma_V^2 = \mathbb{E}[(\Delta V_t)^2] - \mu_V^2 = \delta^\top \Sigma \delta + \frac{1}{2} \text{tr}[(\Gamma \Sigma)^2],$$

où le premier terme reflète le risque directionnel, et le second le risque de convexité agrégé via la corrélation entre facteurs.

— **Skewness (coefficient d'asymétrie)** :

$$\xi_V = \frac{\mathbb{E}[(\Delta V_t - \mu_V)^3]}{\sigma_V^3} = \frac{3\delta^\top \Sigma \Gamma \Sigma \delta + \text{tr}[(\Gamma \Sigma)^3]}{\sigma_V^3}.$$

Ces moments sont ensuite injectés dans la formule d'approximation de Cornish-Fisher, qui permet d'ajuster la VaR pour tenir compte de la déviation par rapport à la normalité :

$$\text{VaR}_\beta = \mu_V + \left( \alpha_\beta + \frac{1}{6}(\alpha_\beta^2 - 1)\xi_V \right) \sigma_V,$$

où  $\alpha_\beta := \Phi^{-1}(1 - \beta)$  est le quantile de niveau  $1 - \beta$  de la loi normale standard.

## 2.3 Approche alternative pour le calcul de la VaR et de l'ES

L'approche alternative proposée ici pour le calcul conjoint de la Value-at-Risk (VaR) et de l'Expected Shortfall (ES) tire parti d'une méthode particulièrement innovante introduite par Rockafellar et Uryasev (2000). Cette méthodologie constitue un changement significatif par rapport aux approches traditionnelles (telles que les simulations historiques, les simulations de Monte Carlo ou l'approche paramétrique via Cornish-Fisher). Son originalité réside dans la formulation de la VaR et de l'ES comme solutions d'un problème d'optimisation convexe bien défini. Concrètement, l'idée principale est la suivante : plutôt que d'estimer séparément la VaR (qui est un quantile de la distribution des pertes) et l'ES (qui est une moyenne conditionnelle au-delà de ce quantile), on construit une unique fonction objectif dont la minimisation simultanée permet d'obtenir à la fois ces deux mesures de risque. Cette approche présente l'avantage d'intégrer naturellement la gestion des pertes extrêmes et d'améliorer la robustesse computationnelle, particulièrement adaptée à des grands portefeuilles de dérivés.

Soit  $L_t(h)$  la variable aléatoire représentant la perte du portefeuille financier sur un horizon temporel donné  $h$ . Cette perte peut être modélisée selon une distribution continue caractérisée par une densité de probabilité  $g(\cdot)$ . Typiquement, cette densité peut être estimée empiriquement ou par simulation numérique à partir de modèles stochastiques du portefeuille.

Pour un niveau de confiance  $\beta \in (0, 1)$ , on définit alors explicitement la fonction objectif  $F_\beta(x)$ , qui dépend d'un paramètre réel  $x$ , comme suit :

$$F_\beta(x) := x + \frac{1}{1 - \beta} \int_x^{+\infty} (L_t(h) - x) g(L_t(h)) dL_t(h).$$

Le premier terme,  $x$ , représente un seuil potentiel de perte du portefeuille que l'on souhaite identifier de façon optimale. Ce seuil correspond précisément au niveau de VaR cherché (ce point sera prouvé rigoureusement par la suite). Le second terme est une intégrale pondérée. Cette intégrale mesure la perte moyenne conditionnelle excédant le seuil  $x$ . Plus précisément, elle correspond à l'espérance de l'excès de perte  $(L_t(h) - x)$  sachant que celle-ci dépasse  $x$ . Ce terme capture explicitement la gravité potentielle des pertes dans la queue droite de la distribution (c'est-à-dire les pertes extrêmes qui dépassent le seuil considéré). Ainsi construite, la fonction  $F_\beta(x)$  possède une propriété mathématique essentielle : elle est convexe en  $x$ . Cette convexité est garantie par la linéarité du premier terme et la propriété intégrale du second terme (étant l'intégrale d'une fonction positive, celle-ci est croissante et donc convexe). Cette propriété assure la présence d'un unique minimum global, facilitant considérablement la résolution numérique du problème.

Rockafellar et Uryasev ont démontré de manière rigoureuse que le minimum global de cette fonction objectif est atteint précisément au point correspondant à la VaR du portefeuille au niveau de confiance choisi  $\beta$ , c'est-à-dire :

$$\text{VaR}_\beta(L) = \arg \min_{x \in \mathbb{R}} F_\beta(x),$$

et qu'à ce minimum global, la valeur de la fonction objectif correspond exactement à l'Expected Shortfall associé, c'est-à-dire :

$$F_\beta(\text{VaR}_\beta(L)) = \text{ES}_\beta(L).$$

Dans une implémentation pratique, la densité de probabilité  $g(\cdot)$  de la perte  $L_t(h)$  doit être estimée à partir d'un ensemble fini de données simulées ou observées. En effet, dans un contexte réel, il est rare de disposer d'une forme analytique exacte pour cette densité, en particulier pour des portefeuilles complexes soumis à de multiples facteurs de risque.



Pour pallier ce problème, une approche fréquemment employée est l'estimation non paramétrique par noyau, appelée aussi « *Kernel Density Estimation* » (KDE). Cette technique statistique permet d'obtenir une approximation lissée et continue d'une distribution empirique initialement discrète. Concrètement, si l'on dispose d'un échantillon de pertes simulées ou observées notées  $\{L_t^{(i)}(h)\}_{1 \leq i \leq n}$ , par exemple issu d'une simulation de Monte Carlo des facteurs de risque sous-jacents au portefeuille, l'estimateur par noyau de la densité  $g(\cdot)$  s'écrit explicitement comme suit :

$$\hat{g}_l(x) = \frac{1}{nl} \sum_{i=1}^n K\left(\frac{x - L_t^{(i)}(h)}{l}\right)$$

où  $K(\cdot)$  est appelée la fonction noyau. Typiquement, cette fonction est choisie comme une fonction symétrique, continue et régulière, telle qu'une fonction gaussienne.

Le paramètre  $l > 0$  introduit dans cette formule est la « fenêtre de lissage » (*bandwidth*). Ce paramètre joue un rôle crucial puisqu'il contrôle directement le degré de régularisation appliqué à la densité estimée. Une petite fenêtre de lissage  $l$  génère une densité très précise localement, mais potentiellement irrégulière et sensible aux variations ponctuelles (phénomène de sous-lissage). À l'inverse, une grande fenêtre produit une densité très lisse mais potentiellement trop approximative et qui pourrait manquer des détails significatifs de la distribution réelle (phénomène de sur-lissage). Le choix optimal de ce paramètre peut être réalisé par des méthodes statistiques telles que la validation croisée (« *cross-validation* ») ou la méthode du critère d'information.

Intuitivement, l'estimation par noyau distribue chaque donnée ponctuelle  $L_t^{(i)}(h)$  sur une région continue autour de cette donnée, de manière à éviter l'aspect discontinu typique des histogrammes traditionnels. Le résultat est une approximation continue, douce et réaliste de la densité sous-jacente, particulièrement adaptée pour évaluer des quantités sensibles comme la Value-at-Risk ou l'Expected Shortfall, qui nécessitent une description précise des queues de la distribution des pertes.

En principe, l'estimateur de densité par noyau (KDE) transforme chaque observation  $L_t^{(i)}(h)$  en une « bosse » de densité locale, puis somme toutes ces petites bosses pour construire l'estimation finale de la densité. Plusieurs noyaux peuvent être choisis, mais l'un des plus courants est le noyau gaussien  $K(u)$ , apprécié pour son caractère intuitif et ses propriétés de lissage. Sa forme paramétrique s'écrit :

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

Le paramètre de lissage  $l$  joue un rôle crucial : trop petit, l'estimation devient sur-bruyante ; trop grand, elle écrase les détails. Pour choisir  $l$  de façon optimale, on minimise la MISE (Mean Integrated Squared Error), c'est-à-dire l'erreur quadratique intégrée moyenne. Pour le noyau gaussien, la règle d'un pouce de Silverman donne :

$$l = \left(\frac{4\hat{\sigma}^2}{3n}\right)^{1/5} \approx 1.06 \hat{\sigma} n^{-1/5},$$

où  $\hat{\sigma}$  est l'écart-type empirique des données  $\{L_t^{(i)}(h)\}_{i=1\dots n}$  et  $n$  leur nombre.

## Chapitre 3

# Régression par Processus Gaussiens en Apprentissage Automatique

La régression par processus gaussiens (GPR) constitue une approche bayésienne non paramétrique particulièrement élégante et robuste pour les problèmes d'apprentissage supervisé. Elle repose sur l'idée centrale que la fonction inconnue que l'on cherche à modéliser n'est pas estimée de manière ponctuelle, mais représentée comme une réalisation aléatoire d'un processus stochastique. En l'occurrence, on suppose que cette fonction provient d'un processus gaussien, c'est-à-dire une collection infinie de variables aléatoires telles que toute combinaison finie suit une loi normale multivariée.

Concrètement, on considère un ensemble de données supervisées  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ , où chaque  $x_i \in \mathbb{R}^d$  désigne une entrée (par exemple, un vecteur de caractéristiques) et chaque  $y_i \in \mathbb{R}$  est une sortie associée (valeur réelle observée). Le cadre probabiliste adopté suppose que les sorties sont générées selon un modèle bruité :

$$y_i = f(x_i) + \varepsilon_i,$$

où  $f$  est une fonction inconnue à inférer, et  $\varepsilon_i \sim \mathcal{N}(0, \sigma_n^2)$  est un bruit gaussien centré représentant l'incertitude sur les observations.

Dans la GPR, la fonction  $f$  est modélisée comme une variable aléatoire fonctionnelle tirée d'un processus gaussien :

$$f \sim \mathcal{GP}(m(x), k(x, x')).$$

La fonction  $m(x)$  est la fonction moyenne (souvent choisie nulle pour simplifier), tandis que  $k(x, x')$  est une fonction de covariance ou noyau, qui encode les hypothèses a priori sur la régularité, la corrélation et la structure de la fonction à apprendre. Le choix de ce noyau est fondamental puisqu'il gouverne la classe de fonctions que le modèle est capable de représenter.

Une fois ce cadre probabiliste posé, l'objectif est d'inférer une distribution a posteriori sur les fonctions compatibles avec les observations. On cherche alors à estimer la distribution prédictive conditionnelle pour une nouvelle entrée  $x^*$ , notée :

$$p(y^* | x^*, X, y) = \int p(y^* | f, x^*) p(f | X, y) df,$$

où :

- $p(f | X, y)$  est la distribution a posteriori sur l'espace des fonctions, calculée à partir des données d'apprentissage ;

- $p(y^* | f, x^*)$  représente la vraisemblance conditionnelle de la prédiction  $y^*$  étant donnée une fonction  $f$ .

Sous l'hypothèse que  $f \sim \mathcal{GP}$  et que les erreurs sont gaussiennes, cette intégrale admet une solution fermée. La prédiction  $y^* | x^*, X, y$  suit alors une distribution normale explicite dont la moyenne donne l'estimation centrale de  $f(x^*)$  et la variance fournit une mesure quantitative de l'incertitude associée à cette prédiction.

Cette propriété analytique fait de la GPR une méthode semi-analytique très attractive, capable de s'ajuster automatiquement à la complexité des données grâce à l'optimisation des hyperparamètres du noyau (par exemple via la maximisation de la vraisemblance marginale).

Nous décrivons ici de manière plus détaillée le principe de la régression par processus gaussiens (GPR), une méthode probabiliste puissante pour résoudre des problèmes de régression supervisée.

Considérons un ensemble d'apprentissage constitué d'observations, noté  $(X, y) = \{(x_i, y_i) | i = 1, 2, \dots, n\}$ , où chaque  $x_i \in \mathbb{R}^d$  représente un vecteur d'entrée de dimension  $d$ , et  $y_i \in \mathbb{R}$  est la sortie associée. L'objectif fondamental est de modéliser la relation fonctionnelle sous-jacente entre les entrées  $x_i$  et les sorties  $y_i$ , tout en tenant compte de l'incertitude inhérente aux observations.

En régression par processus gaussiens, cette relation est formalisée par le modèle suivant :

$$y_i = f(x_i) + \varepsilon_i,$$

où :

- $f(\cdot)$  est une fonction inconnue, considérée comme une réalisation aléatoire d'un *processus gaussien*, c'est-à-dire un objet probabiliste qui définit une distribution conjointe normale sur tout ensemble fini de points  $\{x_i\}$ ,
- $\varepsilon_i \sim \mathcal{N}(0, \sigma_n^2)$  est un bruit additif gaussien centré, indépendant pour chaque observation, qui modélise les erreurs de mesure, les approximations du modèle ou les phénomènes aléatoires non capturés par  $f$ .

Le processus gaussien fournit ainsi une modélisation souple et riche de l'incertitude sur la fonction  $f$ , sans imposer de forme paramétrique stricte (telle qu'un polynôme ou une fonction linéaire). Ce cadre non paramétrique permet d'inférer une famille de fonctions compatibles avec les données observées, en pondérant ces fonctions selon leur plausibilité.

La présence du bruit  $\varepsilon_i$  est essentielle : elle reflète le fait que même si  $f(x_i)$  capturerait parfaitement la tendance générale, les observations  $y_i$  restent affectées par des perturbations aléatoires. Cela se traduit mathématiquement par une variance  $\sigma_n^2$  non nulle dans le modèle, qui joue un rôle crucial dans l'ajustement de la complexité du modèle aux données et dans l'évaluation de l'incertitude sur les prédictions futures.

Un processus gaussien (GP) constitue une généralisation des lois normales multivariées vers un espace de dimension infinie. Intuitivement, alors qu'une loi normale multivariée décrit la distribution conjointe d'un nombre fini de variables aléatoires, un processus gaussien décrit la distribution d'une fonction entière, c'est-à-dire une infinité de variables indexées par les points de l'espace d'entrée.

Formellement, un processus gaussien est une collection de variables aléatoires  $\{f(x), x \in \mathcal{X}\}$ , telle que pour tout ensemble fini de points  $(x_1, x_2, \dots, x_n) \subset \mathcal{X}$ , le vecteur aléatoire

$$\begin{pmatrix} f(x_1), f(x_2), \dots, f(x_k), \\ f(x_{k+1}), \dots, f(x_n) \end{pmatrix}$$

suit une loi normale multivariée. Cela signifie que les valeurs prises par la fonction aux différents points sont distribuées conjointement selon une loi normale, quels que soient les points sélectionnés.

En d'autres termes, on peut voir un processus gaussien comme une distribution de probabilité sur des fonctions : au lieu d'assigner des probabilités à des vecteurs de taille finie, on assigne des probabilités sur des objets fonctionnels, en considérant qu'une fonction correspond à un vecteur infiniment long contenant toutes les valeurs possibles de  $f(x)$  pour chaque  $x$ .

Un processus gaussien est entièrement déterminé par deux fonctions fondamentales.

La première est la fonction moyenne  $\mu(x)$ , qui donne l'espérance de la fonction en chaque point  $x \in \mathcal{X}$  :

$$\mu(x) = \mathbb{E}[f(x)].$$

La seconde est la fonction de covariance, également appelée fonction noyau, notée  $K(x, x')$ , qui mesure la covariance entre les valeurs de la fonction aux points  $x$  et  $x'$  :

$$K(x, x') = \text{Cov}(f(x), f(x')) = \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x')))].$$

Le noyau  $K(\cdot, \cdot)$  encode toute l'information sur la régularité, la corrélation spatiale et la structure fonctionnelle que l'on souhaite imposer à la fonction  $f$ . Il joue un rôle central dans les modèles fondés sur les processus gaussiens, car il contrôle comment deux points de l'espace d'entrée sont liés.

En pratique, un choix très courant pour la fonction noyau en machine learning est le noyau exponentiel quadratique (aussi appelé noyau RBF ou noyau gaussien), défini par :

$$K(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right),$$

où :

- $\sigma_f^2$  est la variance des sorties du processus, contrôlant l'amplitude globale des variations de la fonction,
- $l$  est le paramètre de longueur de corrélation, qui mesure à quelle distance deux points d'entrée restent significativement corrélés.

Ce noyau favorise des fonctions lisses, continues et infiniment différentiables, ce qui est souvent un comportement souhaité dans les problèmes d'interpolation et de régression. Le choix de  $\sigma_f$  et  $l$  est critique : il peut être déterminé par validation croisée ou optimisé par maximisation de la vraisemblance marginale.

Un processus gaussien (GP) est un objet fondamental en modélisation probabiliste, particulièrement utile en apprentissage automatique pour traiter des problèmes de régression non paramétrique. Contrairement aux modèles classiques qui associent des distributions de probabilité à des variables aléatoires scalaires ou vectorielles, un processus gaussien définit une distribution de probabilité sur des fonctions entières. Plus précisément, on le définit comme une collection de variables aléatoires indexées par un espace d'entrée  $\mathcal{X} \subset \mathbb{R}^d$ , telles que toute sous-collection finie suit une loi normale multivariée.

Intuitivement, un processus gaussien introduit une structure de dépendance spatiale : deux points d'entrée proches dans l'espace  $\mathcal{X}$  produisent des sorties fortement corrélées. Cela signifie que la

matrice de covariance  $K(X, X)$ , construite à partir d'un ensemble fini de points  $X = \{x_1, \dots, x_n\}$ , aura des valeurs élevées sur les entrées proches (forte corrélation) et des valeurs faibles pour les points éloignés (faible corrélation).

Dans le cadre de la régression par processus gaussiens, on suppose souvent que la fonction moyenne est nulle,  $\mu(x) = 0$ , ce qui simplifie considérablement les calculs tout en restant général. Ainsi, si  $f \sim \mathcal{GP}(0, k)$ , alors pour tout ensemble de points d'entrée  $X = \{x_1, \dots, x_n\}$ , on a :

$$f(X) = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} \sim \mathcal{N}(0, K(X, X)),$$

où la matrice de covariance  $K(X, X)$  est définie comme :

$$K(X, X) = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}.$$

Cette matrice doit satisfaire deux propriétés fondamentales :

1. **Symétrie** : pour tout  $i, j$ , on a  $k(x_i, x_j) = k(x_j, x_i)$ , ce qui garantit que  $K$  est une matrice symétrique.
2. **Définie positive** : pour tout vecteur non nul  $v \in \mathbb{R}^n$ , on a  $v^\top K v > 0$ . Cela signifie que la fonction noyau génère une matrice de covariance valide, compatible avec une loi gaussienne multivariée.

Un sous-ensemble de dimension finie d'un processus gaussien donne lieu à une distribution marginale qui suit une loi normale multivariée :

$$y \sim \mathcal{N}(0, K(X, X'))$$

avec un vecteur de moyenne nul  $\mu(x) = 0$  et une matrice de covariance  $K(X, X)$ .

À présent, considérons un vecteur aléatoire gaussien multivarié obtenu par la concaténation de deux sous-vecteurs  $\mathbf{X} \in \mathbb{R}^n$  et  $\mathbf{X}^* \in \mathbb{R}^m$ . On suppose que ce vecteur suit une loi normale multivariée jointe, que l'on note :

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{X}^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_{X^*} \end{pmatrix}, \begin{pmatrix} \Omega_{X,X} & \Omega_{X,X^*} \\ \Omega_{X^*,X} & \Omega_{X^*,X^*} \end{pmatrix}\right),$$

où  $\boldsymbol{\mu}_X$  et  $\boldsymbol{\mu}_{X^*}$  désignent les vecteurs moyens respectifs des composantes observées et des composantes à prédire, et où les blocs de la matrice de covariance capturent les variances et covariances croisées. En particulier,  $\Omega_{X,X}$  est la matrice de covariance de  $\mathbf{X}$ ,  $\Omega_{X^*,X^*}$  celle de  $\mathbf{X}^*$ , et  $\Omega_{X^*,X} = \Omega_{X,X^*}^\top$  désigne la matrice de covariance croisée entre les deux.

D'après un résultat classique sur les lois conditionnelles gaussiennes, la distribution conditionnelle de  $\mathbf{X}^*$  sachant une réalisation  $\mathbf{X} = \mathbf{x}$  reste une loi normale multivariée. Plus précisément, on a :

$$\mathbf{X}^* | \mathbf{X} = \mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{X^*|X}, \Omega_{X^*|X}),$$

où l'espérance conditionnelle est donnée par :

$$\boldsymbol{\mu}_{X^*|X} = \mathbb{E}[\mathbf{X}^* | \mathbf{X} = \mathbf{x}] = \boldsymbol{\mu}_{X^*} + \Omega_{X^*,X} \Omega_{X,X}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X),$$

et la matrice de covariance conditionnelle est donnée par :

$$\Omega_{X^*|X} = \text{Var}[\mathbf{X}^* | \mathbf{X} = \mathbf{x}] = \Omega_{X^*,X^*} - \Omega_{X^*,X} \Omega_{X,X}^{-1} \Omega_{X,X^*}.$$

Ce résultat est essentiel en régression par processus gaussiens : il permet d'inférer la distribution d'une fonction aléatoire en un nouveau point  $x^*$ , conditionnellement aux observations disponibles. En particulier, il conduit à une prédiction probabiliste, où l'on connaît à la fois la moyenne prédite  $\mathbb{E}[f(x^*) | \mathcal{D}]$  et la variance conditionnelle qui quantifie l'incertitude associée à cette prédiction.

Afin de mieux comprendre le fonctionnement de la régression par processus gaussiens (GPR), il est très instructif d'observer visuellement ce que signifie modéliser une fonction comme une variable aléatoire dans l'espace des fonctions. Pour cela, on peut simuler plusieurs trajectoires (ou réalisations) d'une fonction tirée aléatoirement d'un processus gaussien *a priori*, c'est-à-dire sans aucune observation de données. Cela permet de représenter notre *a priori bayésien* : l'ensemble des fonctions que nous considérons comme plausibles avant d'avoir observé quoi que ce soit.

Concrètement, on choisit un ensemble fini de points d'entrée  $X = [x_1, x_2, \dots, x_n]$ , par exemple répartis uniformément sur un intervalle de  $\mathbb{R}$ . À chacun de ces points, on associe une variable aléatoire  $f(x_i)$ , dont la distribution conjointe suit une loi normale multivariée  $\mathcal{N}(0, K(X, X))$ . La matrice de covariance  $K(X, X)$  est déterminée par un noyau  $k(x, x')$ , souvent choisi comme le noyau exponentiel quadratique (aussi appelé noyau RBF ou gaussien), défini par :

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right),$$

où  $\sigma_f^2$  représente la variance des sorties et  $l$  est le paramètre de longueur de corrélation qui contrôle la régularité de la fonction.

On peut alors tirer plusieurs échantillons  $f^{(1)}, f^{(2)}, \dots, f^{(n)}$  à partir de cette loi multivariée. Chaque échantillon correspond à une fonction possible compatible avec nos hypothèses *a priori* (i.e. lisse, corrélée spatialement, etc.). Ces trajectoires peuvent être visualisées sous forme de courbes passant par les points  $x_1, \dots, x_n$ , et elles donnent une intuition très forte de la distribution de probabilité sur les fonctions induite par le processus gaussien.

Ce type de visualisation montre que, même en l'absence de données, la GPR impose une structure statistique riche à l'espace des fonctions, gouvernée uniquement par le choix du noyau. En outre, en augmentant le nombre de réalisations indépendantes  $n$ , la moyenne empirique des fonctions générées  $f^{(i)}(x)$  converge vers la fonction moyenne du processus, ici supposée nulle :

$$\frac{1}{n} \sum_{i=1}^n f^{(i)}(x) \longrightarrow \mu(x) = 0 \quad \text{lorsque } n \rightarrow \infty.$$

Cela illustre le fait que le processus gaussien est centré autour de sa moyenne, et que sa variabilité est entièrement décrite par la fonction noyau. Une fois les données observées introduites, cette variabilité se réduit autour des points connus, et l'incertitude s'accroît à mesure qu'on s'éloigne des données. C'est cette capacité à modéliser à la fois les prédictions et l'incertitude associée qui rend la GPR si puissante dans des contextes où la fiabilité est cruciale.

À présent, étant donné un ensemble de cinq points d'entrée observés et leurs valeurs de sortie correspondantes  $(x_1, f(x_1)), (x_2, f(x_2)), \dots, (x_5, f(x_5))$ , ainsi qu'un processus gaussien *a priori* sur la fonction  $f \sim \mathcal{N}(0, K(X, X))$ , on souhaite calculer la distribution *a posteriori* de  $f(x^*)$  pour toute entrée  $x^*$ .

La **figure 3.1** illustre dix réalisations issues de cette distribution *a posteriori*, construites à partir de cinq observations extraites d'une fonction sinus. On observe que les fonctions échantillonnées

à partir de la distribution a posteriori passent au plus près des valeurs observées, mais varient considérablement dans les régions où aucune donnée n'est disponible. Cela met en évidence le fait que l'incertitude est fortement réduite à proximité des points observés.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.gaussian_process.kernels import RBF

# Paramètres
n_points = 100
X = np.linspace(0, 10, n_points).reshape(-1, 1)
kernel = RBF(length_scale=1.0)

# Processus gaussien a priori
gp_prior = GaussianProcessRegressor(kernel=kernel, alpha=1e-10)
y_prior_samples = gp_prior.sample_y(X, 10)

# Données d'entraînement (fonction sinus)
X_obs = np.array([[1.0], [3.0], [5.0], [6.5], [8.5]])
y_obs = np.sin(X_obs)

# Processus gaussien a posteriori
gp_posterior = GaussianProcessRegressor(kernel=kernel, alpha=1e-6)
gp_posterior.fit(X_obs, y_obs)
y_post_samples = gp_posterior.sample_y(X, 10)

# Affichage
fig, axs = plt.subplots(1, 2, figsize=(14, 5))

# Graphe a priori
axs[0].plot(X, y_prior_samples)
axs[0].set_title("10 échantillons du processus gaussien a priori")

# Graphe a posteriori
axs[1].plot(X, y_post_samples)
axs[1].scatter(X_obs, y_obs, color='black', label="observations")
axs[1].set_title("10 échantillons après entraînement sur 5 observations")
axs[1].legend()

plt.tight_layout()
plt.show()
```

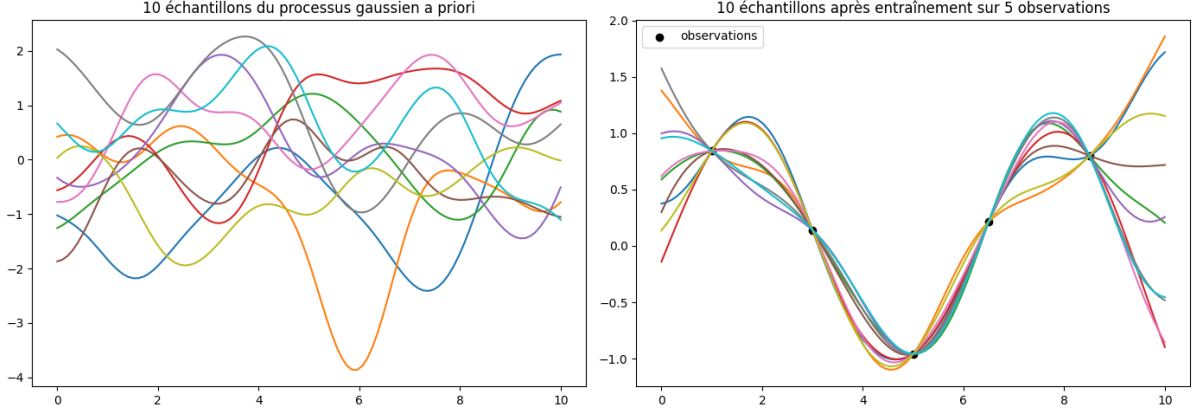


FIGURE 3.1 – Simulations d'un processus gaussien avant et après l'apprentissage

### 3.1 Méthode d'estimation fondée sur la régression par processus gaussiens

L'un des objectifs fondamentaux du calcul de la distribution a posteriori en régression par processus gaussiens est de pouvoir effectuer des prédictions sur des cas de test encore non observés. Autrement dit, une fois les données d'apprentissage assimilées, on souhaite estimer la valeur probable de la fonction cible en un nouveau point  $x^* \in \mathbb{R}^d$ , et quantifier l'incertitude associée à cette prédiction.

Considérons alors une matrice  $X^* \in \mathbb{R}^{n^* \times d}$  regroupant  $n^*$  points de test, et notons  $f^* \in \mathbb{R}^{n^*}$  le vecteur des valeurs inconnues de la fonction  $f$  en ces points. L'objectif est de déterminer la loi conditionnelle de  $f^*$  compte tenu des observations  $(X, y) \in \mathbb{R}^{n \times d} \times \mathbb{R}^n$ , avec :

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_n^2)$$

et  $f \sim \mathcal{GP}(0, k)$ .

La distribution jointe de  $y$  et  $f^*$  est gaussienne, de la forme :

$$\begin{pmatrix} y \\ f^* \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} K(X, X) + \sigma_n^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{pmatrix}\right)$$

où l'on utilise les notations suivantes :

- $K := K(X, X) \in \mathbb{R}^{n \times n}$ , la matrice de covariance entre les points d'apprentissage ;
- $K_* := K(X^*, X) \in \mathbb{R}^{n^* \times n}$ , la matrice des covariances croisées ;
- $K_{**} := K(X^*, X^*) \in \mathbb{R}^{n^* \times n^*}$ , la matrice de covariance des points de test.

Alors, par la formule classique de conditionnement des lois normales, la distribution a posteriori de  $f^*$  conditionnellement à  $y$  est :

$$f^* | y \sim \mathcal{N}(K_* K^{-1} y, K_{**} - K_* K^{-1} K_*^\top)$$

Cette notation compacte est équivalente à celle obtenue plus haut avec les expressions explicites de la moyenne prédictive et de la covariance prédictive. Elle souligne l'élégance et la puissance



analytique de la régression par processus gaussiens, permettant d'exprimer les prédictions comme des projections linéaires pondérées sur les observations passées.

La distribution  $f^* | y$  permet ainsi de prédire, pour chaque point  $x^*$ , une valeur attendue (prédiction centrale) et un intervalle de confiance quantifiant l'incertitude de cette prédiction. On observe typiquement une incertitude minimale près des points d'apprentissage et une incertitude croissante à mesure que l'on s'en éloigne.

Dans le cadre de la régression par processus gaussiens (GPR), une fois que l'on a observé un ensemble de données d'apprentissage  $(X, y) = \{(x_i, y_i)\}_{i=1}^n$ , notre objectif est de prédire la sortie  $y^*$  associée à une nouvelle entrée  $x^*$ . Le processus gaussien fournit non seulement une estimation ponctuelle de cette sortie, mais aussi une quantification rigoureuse de l'incertitude autour de cette prédiction. La distribution a posteriori de  $y^*$ , conditionnellement aux données observées, est une loi normale dont la moyenne et la variance sont données analytiquement.

La moyenne a posteriori représente la meilleure estimation de  $f(x^*)$  compte tenu des observations passées. Elle s'écrit sous forme matricielle :

$$\bar{y}^* = \mathbb{E}[f(x^*) | f(x)] = K_* K^{-1} y$$

où :

- $K \in \mathbb{R}^{n \times n}$  est la matrice de covariance entre les points d'apprentissage, définie par  $K_{ij} = k(x_i, x_j)$ ,
- $K_* \in \mathbb{R}^{1 \times n}$  est le vecteur des covariances entre le point test  $x^*$  et les points d'apprentissage, soit  $K_* = [k(x^*, x_1), \dots, k(x^*, x_n)]$ ,
- $y \in \mathbb{R}^n$  est le vecteur des observations de sortie.

La variance de la distribution prédictive, qui capture l'incertitude sur la prédiction, est donnée par :

$$\text{Var}(y^*) = K_{**} - K_* K^{-1} K_*^\top$$

où  $K_{**} = k(x^*, x^*)$  est la variance a priori au point  $x^*$ .

Au-delà de l'expression matricielle, la moyenne prédictive peut être interprétée de manière intuitive comme une combinaison linéaire des valeurs du noyau évaluées en  $x^*$ . En posant :

$$\omega := K^{-1} y \in \mathbb{R}^n,$$

on peut réécrire la prédiction comme :

$$\bar{y}^* = \sum_{i=1}^n \omega_i k(x^*, x_i)$$

Cela signifie que la prédiction en un point donné s'exprime comme une somme pondérée de la similarité (au sens du noyau  $k$ ) entre ce point et chacun des points d'apprentissage. Les poids  $\omega_i$  sont déterminés de façon globale à partir des données observées, et reflètent la manière dont chaque observation influence la prédiction future.

Une propriété centrale des processus gaussiens est que la moyenne a posteriori peut être calculée sans inverser explicitement la matrice de covariance  $K$ , ce qui est avantageux sur le plan numérique. En effet, on peut simplement résoudre le système linéaire suivant :

$$K \omega = f(x),$$

où  $\omega \in \mathbb{R}^n$  est le vecteur des poids. Cela permet d'éviter l'inversion directe de  $K$ , souvent coûteuse et instable pour de grandes dimensions. La moyenne a posteriori s'exprime alors comme une combinaison linéaire des valeurs du noyau centrées sur  $x^*$  :

$$\mathbb{E}[f(x^*) \mid f(x)] = \sum_{i=1}^n \omega_i k(x^*, x_i).$$

Inversement, on peut exprimer cette moyenne comme une combinaison linéaire des valeurs observées  $f(x_i)$ , en utilisant le vecteur des poids  $\varphi \in \mathbb{R}^n$ , donné par :

$$\varphi = K(x^*, X)K^{-1},$$

d'où :

$$\mathbb{E}[f(x^*) \mid f(x)] = \sum_{i=1}^n \varphi_i f(x_i).$$

Cette formulation met en évidence que la prédiction au point  $x^*$  dépend uniquement des similarités entre ce point et les points d'entraînement, telles que mesurées par la fonction noyau.

Dans le cas général où la fonction moyenne a priori  $\mu(x)$  n'est pas nulle, la GPR peut être adaptée en considérant la fonction centrée :

$$h(x) := f(x) - \mu(x),$$

qui suit un processus gaussien centré  $h \sim \mathcal{GP}(0, k)$ . On effectue alors l'inférence sur  $h$ , en remplaçant les observations  $y_i$  par  $h(x_i) = y_i - \mu(x_i)$ . Une fois la moyenne a posteriori  $\mathbb{E}[h(x^*)]$  obtenue, la prédiction finale s'écrit :

$$\mathbb{E}[f(x^*)] = \mu(x^*) + \mathbb{E}[h(x^*)].$$

## 3.2 Entraînement des algorithmes à processus gaussiens

La mise en œuvre concrète d'un algorithme de régression par processus gaussiens (GPR) repose de manière cruciale sur le **choix des hyperparamètres du noyau**. En effet, la qualité des prédictions produites par le modèle dépend étroitement de la capacité du noyau  $k(x, x')$  à capturer la structure sous-jacente des données. Or, ce noyau est généralement paramétré par un ensemble d'hyperparamètres regroupés dans un vecteur noté  $\theta = \{l, \sigma_f, \sigma_n\}$ , où :

- $l$  est la *longueur de corrélation* du noyau,
- $\sigma_f^2$  est la *variance a priori* du processus,
- $\sigma_n^2$  est la *variance du bruit* observationnel.

L'approche standard pour calibrer ces hyperparamètres consiste à **maximiser la log-vraisemblance marginale** :

$$\ell(\theta) = \log p(\mathbf{y} \mid \theta) = -\frac{1}{2} \mathbf{y}^\top (K + \sigma_n^2 I)^{-1} \mathbf{y} - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{n}{2} \log(2\pi)$$

où :

- $K$  est la matrice de covariance  $K_{ij} = k(x_i, x_j)$ ,

- $I$  est la matrice identité,
- $n$  est le nombre d'observations.

Cette fonction se décompose naturellement en trois termes :

- un terme d'ajustement aux données :  $-\frac{1}{2}\mathbf{y}^\top (K + \sigma_n^2 I)^{-1} \mathbf{y}$ ,
- un terme de pénalisation de la complexité :  $-\frac{1}{2} \log |K + \sigma_n^2 I|$ ,
- une constante de normalisation :  $-\frac{n}{2} \log(2\pi)$ .

Les dérivées de cette fonction par rapport aux hyperparamètres  $\theta_j$  sont données par :

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} = \frac{1}{2} \mathbf{y}^\top K^{-1} \frac{\partial K}{\partial \theta_j} K^{-1} \mathbf{y} - \frac{1}{2} \text{tr} \left( K^{-1} \frac{\partial K}{\partial \theta_j} \right)$$

En posant  $\boldsymbol{\omega} = K^{-1} \mathbf{y}$ , cette expression se simplifie en :

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} = \frac{1}{2} \text{tr} \left[ \left( \boldsymbol{\omega} \boldsymbol{\omega}^\top - K^{-1} \right) \frac{\partial K}{\partial \theta_j} \right]$$

Un des principaux inconvénients des processus gaussiens, notamment dans leur mise en œuvre pratique pour des tâches de régression, réside dans leur **mauvaise scalabilité** vis-à-vis de la taille de l'échantillon d'apprentissage. Plus précisément, lorsque le nombre d'observations  $n$  devient grand, la complexité numérique devient un facteur limitant majeur.

En effet, le calcul de la *log-vraisemblance marginale* du modèle, notée  $\log p(y \mid \boldsymbol{\theta})$ , repose sur l'inversion de la matrice de covariance  $K + \sigma_n^2 I$ , qui est une matrice symétrique définie positive de taille  $n \times n$ . L'inversion d'une telle matrice nécessite un temps de calcul de l'ordre de  $\mathcal{O}(n^3)$  à l'aide des algorithmes standards comme la décomposition de Cholesky. Ce coût cubique rend le modèle peu praticable lorsque  $n$  dépasse quelques milliers.

De plus, le calcul du déterminant de la matrice  $K + \sigma_n^2 I$ , nécessaire pour le terme de pénalité de complexité dans l'expression de la log-vraisemblance marginale, est obtenu en même temps que l'inversion grâce à la décomposition de Cholesky, mais il participe également à la complexité globale.

Cependant, une fois cette inversion effectuée, le calcul des *dérivées de la log-vraisemblance marginale* par rapport aux hyperparamètres  $\boldsymbol{\theta} = \{l, \sigma_f, \sigma_n\}$  du noyau devient beaucoup plus léger. Ces dérivées interviennent dans l'étape d'optimisation des hyperparamètres et ne requièrent qu'un **temps quadratique**  $\mathcal{O}(n^2)$  par hyperparamètre, ce qui est tout à fait gérable même pour des valeurs de  $n$  modérément élevées.

Dès lors, bien que le coût initial de calcul de  $K^{-1}$  soit élevé, l'optimisation elle-même — qui repose souvent sur des algorithmes à base de gradient comme la *descente de gradient*, les *méthodes du gradient conjugué*, ou encore les algorithmes *quasi-Newtoniens* (comme BFGS ou L-BFGS) — peut être efficacement réalisée. Ces méthodes exploitent le fait que la log-vraisemblance marginale est différentiable par rapport aux hyperparamètres du noyau, ce qui permet une estimation fine et robuste des paramètres du modèle.

Enfin, pour les grands ensembles de données, des approches approximatives comme la **GPR parcimonieuse**, l'utilisation de **points inductifs** (*inducing points*) ou les **méthodes variationnelles** permettent de réduire drastiquement le coût en ramenant la complexité à  $\mathcal{O}(nm^2)$  voire  $\mathcal{O}(m^3)$  pour  $m \ll n$ , tout en conservant une précision raisonnable.

## Chapitre 4

# Applications de la régression par processus gaussiens

Dans cette section, nous mettons en lumière deux applications majeures des processus gaussiens dans le domaine de la finance. Plus précisément, nous exploitons la puissance de cet algorithme pour répondre à deux problématiques fondamentales : d'une part, l'évaluation de la performance des portefeuilles à travers la revalorisation des instruments financiers, et d'autre part, la mesure du risque associé à un portefeuille global de produits dérivés. Ces deux axes illustrent la capacité des processus gaussiens à modéliser de manière souple et probabiliste des phénomènes complexes, souvent non linéaires, inhérents aux marchés financiers.

### 4.1 Valorisation d'un portefeuille de produits dérivés

À une date donnée  $t$ , la valeur d'un portefeuille de produits dérivés, notée  $V_t$ , peut être décomposée en une somme pondérée d'un nombre fini de sous-portefeuilles, chacun étant structuré selon la nature spécifique des facteurs de risque sous-jacents. Cette structuration permet de regrouper les instruments dérivés en ensembles homogènes, en fonction de leur exposition à des risques communs (tels que le taux d'intérêt, la volatilité, ou encore le prix du sous-jacent).

Plus précisément, chaque instrument dérivé du portefeuille est adossé à un actif sous-jacent, lequel est caractérisé par un vecteur propre de facteurs de risque. La valeur de chaque sous-portefeuille  $V_i$  dépend alors d'un vecteur de variables de marché noté  $x_i$ , représentant l'évolution de ces facteurs pour le bloc  $i$ .

Ainsi, la valeur globale du portefeuille à l'instant  $t$  s'exprime formellement comme la somme des contributions individuelles de chacun des  $m$  sous-portefeuilles :

$$V_t = \sum_{i=1}^m V_i(x_i)$$

où chaque fonction  $V_i(x_i)$  modélise la relation entre les facteurs de risque pertinents et la valeur agrégée des instruments dérivés associés au sous-portefeuille  $i$ .

Dans la suite de cette étude, et à des fins illustratives, nous nous plaçons dans le cadre d'un portefeuille composé exclusivement de *produits dérivés sur actions*. Cette hypothèse, sans nuire à la généralité de l'approche, permet de clarifier la présentation du modèle sans perdre en applicabilité. Dans ce contexte, chaque facteur de risque  $X_i$ , associé au  $i$ -ième sous-portefeuille, est défini à partir de trois variables fondamentales : le *prix spot* de l'actif sous-jacent  $S_i$ , sa *volatilité implicite*  $\sigma_i$ , et le *taux d'intérêt* sans risque  $r$ . On regroupe ces variables dans un vecteur :

$$X_i = (S_i, \sigma_i, r), \quad \text{pour } i = 1, \dots, m.$$

Comme introduit précédemment, l'objectif principal de notre travail est de procéder à une *évaluation du risque* du portefeuille à travers le calcul de la *Value-at-Risk (VaR)* et de l'*Expected Shortfall (ES)*. Dans ce but, nous supposons que les *modèles de valorisation classiques* (Black-Scholes, modèles à volatilité locale, modèles stochastiques, etc.) sont disponibles et correctement calibrés. Toutefois, dans notre approche, ces modèles ne sont mobilisés que dans une *phase initiale de génération de données*, dans le but d'entraîner un algorithme de régression fondé sur les *processus gaussiens (Gaussian Processes, GP)*.

L'idée clé est de transférer la charge computationnelle coûteuse liée à la revalorisation répétée du portefeuille vers une *phase d'apprentissage supervisé*, où le modèle GP va apprendre à approximer la fonction de valorisation associée à chaque sous-portefeuille  $V_i(x_i)$ . Une fois entraîné, le modèle GP pourra alors assurer de manière *rapide et efficace* la revalorisation du portefeuille complet, notamment dans le cadre de *simulations Monte Carlo* massives, nécessaires au calcul de la VaR ou de l'ES.

Concrètement, pour chaque sous-portefeuille  $i \in \{1, \dots, m\}$ , on construit un *jeu de données d'apprentissage*  $(X, y)$ , où :

- la *matrice d'entrée*  $X \in \mathbb{R}^{d \times 3}$  contient  $d$  combinaisons de paramètres de marché  $(S_i, \sigma_i, r)$ ,
- et le *vecteur de sortie*  $y = (V_i(x_{i,1}), \dots, V_i(x_{i,d}))^T$  regroupe les valeurs associées issues d'un modèle de valorisation sophistiqué.

Les points d'apprentissage sont échantillonnés de manière aléatoire et uniforme sur un *hypercube de marché* défini par :

$$[S_{i,\min}, S_{i,\max}] \times [\sigma_{i,\min}, \sigma_{i,\max}] \times [r_{\min}, r_{\max}].$$

Une fois le modèle GP entraîné, la prédiction de la valeur d'un sous-portefeuille pour une nouvelle configuration de marché  $x_i^*$  s'obtient en calculant l'*espérance conditionnelle* de  $V_i(x_i^*)$  sous la loi prédictive gaussienne donnée par :

$$\mathbb{E}[V_i(x_i^*) \mid \bar{V}_l] = \sum_{j=1}^d \varphi_{i,j} V_i(x_{i,j}),$$

où les *poids de régression*  $\varphi_{i,j}$  sont déterminés analytiquement par la structure du processus gaussien :

$$\varphi_i = K_i^* K_i^{-1},$$

avec les notations suivantes :

- $\varphi_i = (\varphi_{i,1}, \dots, \varphi_{i,d})$  est le vecteur des poids d'interpolation associés à l'entrée  $x_i^*$ ,
- $K_i^{-1} = \left( a_{jk}^{(i)} \right)_{1 \leq j, k \leq d}$  est l'inverse de la matrice de covariance entre les points d'apprentissage,
- $\bar{V}_l = [V_i(x_{i,1}), \dots, V_i(x_{i,d})]^T$  est le vecteur des valeurs de référence issues du modèle de pricing,
- $K_i^* = [k(x_i^*, x_{i,1}), \dots, k(x_i^*, x_{i,d})]^T$  contient les covariances entre la nouvelle entrée  $x_i^*$  et les données d'apprentissage.

Cette approche permet de remplacer efficacement les modèles de pricing traditionnels par une *approximation bayésienne non paramétrique* de la fonction de valorisation, rendant possible une revalorisation rapide et fiable dans les contextes de simulation intensive.

Dans ce qui suit, nous présentons un *exemple numérique* destiné à illustrer concrètement l’approche méthodologique exposée précédemment. Le portefeuille considéré est constitué de *produits dérivés écrits sur quatre actifs sous-jacents distincts*, notés  $S_1, S_2, S_3, S_4$ . Il comprend un total de **100 options**, réparties en trois grandes catégories : des *options vanilles*, des *options barrières* et des *options américaines*.

Le **Tableau 1** ci-dessous décrit la composition du portefeuille selon les sous-jacents et les types d’options, en indiquant le nombre d’instruments détenus pour chaque combinaison. On observe une structure symétrique dans laquelle chaque actif sous-jacent est associé à 10 options vanilles, 10 options barrières et 5 options américaines.

Ce portefeuille composite est représentatif d’un cas réaliste, où différents instruments dérivés sont utilisés pour couvrir ou spéculer sur plusieurs dimensions de risque liées aux mêmes sous-jacents. Il constitue ainsi une base pertinente pour tester la robustesse et la performance de l’algorithme de régression par processus gaussiens dans le cadre d’une *revalorisation accélérée* ou d’un *calcul de mesure de risque* (VaR, ES).

Pour des raisons de lisibilité, seules les grandes lignes de la composition sont présentées ici. Les caractéristiques détaillées des options (telles que les niveaux de barrière, les maturités, les styles d’exercice, etc.) sont spécifiées en **Annexe 1**.

Sous-jacent	Option Vanille	Option Barrière	Option Américaine
$S_1$	10	10	5
$S_2$	10	10	5
$S_3$	10	10	5
$S_4$	10	10	5

**Tableau 1 — Composition du portefeuille par type d’option et sous-jacent**

À des fins d’illustration, et sans perte de généralité, nous commençons par un exemple numérique simple dans lequel l’unique facteur de risque pris en compte est constitué des *prix spot des sous-jacents*, notés  $S_i$ , pour  $i = 1, 2, 3, 4$ . Ce cadre volontairement restreint permet de se concentrer sur le fonctionnement intrinsèque de l’algorithme de régression par processus gaussiens (GPR), tout en conservant une structure suffisamment riche pour évaluer les performances de l’approche.

Afin d’analyser la sensibilité de la méthode à la *taille de l’échantillon d’apprentissage*, quatre configurations expérimentales sont étudiées, correspondant à des tailles  $d = 5, 10, 15$  et  $20$ . Pour chacune de ces tailles, un jeu de données  $(X, y)$  est généré, où :

- $X \in \mathbb{R}^{d \times 4}$  représente les combinaisons aléatoires des prix spot simulés pour les quatre actifs,
- $y \in \mathbb{R}^d$  correspond aux valeurs du portefeuille évaluées à partir du modèle de valorisation de référence.

Les **Tableaux 3 et 4** présentent explicitement les jeux d’apprentissage  $(X, y)$  pour les cas  $d = 5$  et  $d = 10$ . Ces jeux ont été construits à partir d’échantillons simulés uniformément dans un domaine de variation réaliste des prix de marché.

Pour les cas où la taille d’apprentissage est plus importante ( $d = 15$  et  $d = 20$ ), les matrices

associées sont plus volumineuses. Par souci de clarté et afin d’alléger le corps du texte, ces jeux de données complets sont fournis en **Annexe 2** du présent document.

Sous-jacent	Prix Spot (\$)	Taux d’intérêt (%)	Volatilité (%)
$S_1$	100	2	40
$S_2$	105	2	20
$S_3$	90	2	50
$S_4$	110	2	30

**Tableau 2 — Données de marché associées aux actifs sous-jacents**

Le tableau présenté ci-dessous constitue l’échantillon d’apprentissage utilisé pour l’estimation du modèle de *régression par processus gaussiens (GPR)*. Il contient un ensemble de **cinq observations simulées**, chacune correspondant à une configuration spécifique du *prix spot* d’un actif sous-jacent, ainsi qu’aux *valeurs individuelles de plusieurs portefeuilles partiels* (notés Portefeuille 1 à Portefeuille 4), auxquels s’ajoute la *valeur agrégée du portefeuille global*.

Ce jeu de données est généré à l’aide de modèles de valorisation de référence, que l’on suppose ici connus et disponibles. Il constitue la base d’apprentissage sur laquelle le modèle GPR va apprendre à approximer la fonction de valorisation associée à chacun des sous-portefeuilles, ainsi qu’au portefeuille global.

Une fois entraîné, le modèle pourra être utilisé pour effectuer des prédictions rapides et précises sur de nouvelles configurations de marché, dans une optique de *revalorisation accélérée* ou d’évaluation du risque. Ce cadre s’inscrit pleinement dans la logique d’approximation non paramétrique et bayésienne propre aux processus gaussiens, où les données d’apprentissage jouent un rôle central dans la qualité des prédictions futures.

Une fois entraîné, le modèle pourra être utilisé pour effectuer des prédictions rapides et précises sur de nouvelles configurations de marché, dans une optique de *revalorisation accélérée* ou d’évaluation du risque. Ce cadre s’inscrit pleinement dans la logique d’approximation non paramétrique et bayésienne propre aux processus gaussiens, où les données d’apprentissage jouent un rôle central dans la qualité des prédictions futures.

Spot	Portefeuille 1	Portefeuille 2	Portefeuille 3	Portefeuille 4	Portefeuille global
1	260	1,289	1,010	737	3,296
35	180	860	777	560	2,377
69	289	513	726	492	2,021
103	568	382	810	592	2,351
137	957	439	979	811	3,186

**Tableau 3 — Jeu d’apprentissage pour la régression par processus gaussiens (taille = 5)**

Pour chaque configuration de l’échantillon d’apprentissage, la valeur de chacun des quatre sous-portefeuilles, ainsi que celle du portefeuille global, est déterminée à l’aide de modèles de valorisation financiers de référence, en fonction de la nature des instruments dérivés qu’ils contiennent. Les options vanilles et les options barrières sont évaluées à l’aide du modèle classique de Black-Scholes, qui repose sur une hypothèse de dynamique log-normale des actifs sous-jacents et sur la possibilité d’une couverture continue.

Concernant les options américaines, la valorisation est également fondée sur le modèle de Black-Scholes, mais adaptée à leur spécificité via une implémentation numérique fondée sur un arbre

trinomial. Cette approche permet de gérer le caractère exercable à tout moment des options américaines, en intégrant de manière récursive les opportunités optimales d'exercice anticipé à chaque nœud de l'arbre.

Il convient de souligner que, indépendamment des modèles de valorisation utilisés en amont pour calculer les valeurs de référence, l'algorithme de régression par processus gaussiens (GPR) s'applique de manière uniforme. Ce dernier apprend à approximer la fonction de valorisation à partir des données simulées, sans nécessiter de connaissances structurelles sur les modèles sous-jacents.

En pratique, l'utilisation du GPR permet de remplacer les évaluations directes par une procédure plus rapide, plus souple et souvent plus précise, notamment dans les contextes où des simulations massives sont requises, comme dans le cadre de la mesure du risque (VaR, ES) ou de la revalorisation dynamique en temps réel.

Spot	Portefeuille 1	Portefeuille 2	Portefeuille 3	Portefeuille 4	Portefeuille global
1	260	1,289	1,010	737	3,296
16	207	1,098	886	653	2,843
31	180	909	795	578	2,462
46	194	732	742	519	2,187
61	246	580	724	492	2,042
76	335	466	735	501	2,036
91	454	401	769	541	2,165
106	599	381	822	606	2,408
121	764	395	891	694	2,744
136	945	435	973	803	3,156

**Tableau 4 — Jeu d'apprentissage pour la régression par processus gaussiens (taille = 10)**

À l'issue de la génération de l'échantillon d'apprentissage, celui-ci est transmis à l'*algorithme de régression par processus gaussiens (GPR)*. Ce dernier procède à une *étape d'entraînement* fondamentale, qui consiste à optimiser les *hyperparamètres* du noyau (ou fonction de covariance) et à calculer l'*inverse de la matrice de covariance*  $K$ , élément central du mécanisme d'interpolation probabiliste. Cette étape permet de calibrer le modèle GPR sur les données observées, en maximisant la vraisemblance marginale des observations ou via des techniques bayésiennes.

Une fois cette phase d'apprentissage complétée, le modèle est prêt à effectuer des *prédictions rapides et précises* sur de nouvelles configurations de marché. L'algorithme repose alors sur le calcul de la moyenne conditionnelle du processus gaussien entraîné, ce qui lui permet d'estimer de manière souple la valeur du portefeuille pour tout nouveau vecteur de facteurs de risque.

Pour évaluer les performances du modèle GPR, deux *métriques d'erreur standards* sont mobilisées :

- La **Maximum Absolute Error (MAE)**, qui mesure la plus grande erreur absolue entre la valeur exacte du portefeuille et sa prédiction GPR :

$$\text{MAE} = \max_{1 \leq i \leq n} |V_i^* - V_i^{\text{GPR}}|,$$



- La **Average Absolute Error (AAE)**, qui fournit une estimation moyenne de l'écart absolu sur l'ensemble des prédictions :

$$\text{AAE} = \frac{1}{n} \sum_{i=1}^n |V_i^* - V_i^{\text{GPR}}|,$$

où  $V_i^*$  désigne la valeur de référence du portefeuille pour l'observation  $i$ , obtenue via un modèle de valorisation exact (par exemple, Black-Scholes), et  $V_i^{\text{GPR}}$  la valeur prédite par l'algorithme de régression gaussienne.

L'échantillon de test est généré indépendamment, en tirant de nouveaux points  $\{S_i^*\}_{i=1}^4$  de façon uniforme dans l'intervalle d'apprentissage spécifié :

$$S_i^* \sim \mathcal{U}([S_{i,\min}, S_{i,\max}]), \quad i = 1, 2, 3, 4.$$

Ces points permettent d'évaluer la *capacité de généralisation du modèle GPR* hors de l'échantillon initial.

Les résultats numériques associés à ces tests sont présentés dans la **Figure 4.1**, où l'on compare les performances du GPR en termes de *précision de valorisation* et de *temps de calcul*. Par ailleurs, des résultats supplémentaires issus de tests empiriques de grande ampleur sont proposés dans les **Tableaux 3 et 4**, qui décrivent les jeux d'apprentissage utilisés, ainsi que dans les **Figures 5 et 6**, qui illustrent la qualité de prédiction du GPR et les gains computationnels observés.

Enfin, le tableau suivant présente un exemple synthétique des résultats obtenus : il compare les *valeurs estimées par le modèle GPR* à celles issues du modèle de **Black-Scholes**, pour différentes configurations de prix spot. L'excellente concordance observée entre les deux approches valide la *fiabilité du modèle GPR* pour approximer la valeur d'un portefeuille de produits dérivés, tout en bénéficiant d'une complexité de calcul fortement réduite.

GPR Valuation of the Derivative Portfolio		
Spot Price	GPR Price	Black-Scholes Price
37	2556	2556
43	2453	2453
49	2373	2373
55	2313	2313
61	2275	2275
67	2257	2257
73	2260	2260
79	2285	2285
85	2330	2330
91	2393	2393
97	2473	2474
103	2570	2570
109	2681	2681
115	2806	2805

Tableau 5 — Valorisation du portefeuille dérivé par GPR (taille de l'échantillon = 10)

```

import numpy as np
import matplotlib.pyplot as plt

# Génération de données fictives pour les 3 cas
np.random.seed(42)

def generate_case(n_points, noise_level):
    black_prices = np.linspace(800, 3400, n_points)
    gpr_prices = black_prices + np.random.normal(0, noise_level, size=n_points)
    return black_prices, gpr_prices

# Cas 1 : plus bruité
x1, y1 = generate_case(100, noise_level=150)

# Cas 2 : moins bruité
x2, y2 = generate_case(100, noise_level=75)

# Cas 3 : très faible bruit (presque parfait)
x3, y3 = generate_case(100, noise_level=20)

# Création de la figure avec 3 sous-graphes
fig, axs = plt.subplots(1, 3, figsize=(18, 5), sharey=True)

cases = [(x1, y1), (x2, y2), (x3, y3)]

for i, (x, y) in enumerate(cases):
    axs[i].scatter(x, y, color='mediumpurple', s=10, label='Pricing Point')
    axs[i].set_title("Portfolio Valuation - Black and GPR")
    axs[i].set_xlabel("Black Prices")
    if i == 0:
        axs[i].set_ylabel("GPR Prices")
    axs[i].legend()

plt.tight_layout()
plt.show()

```

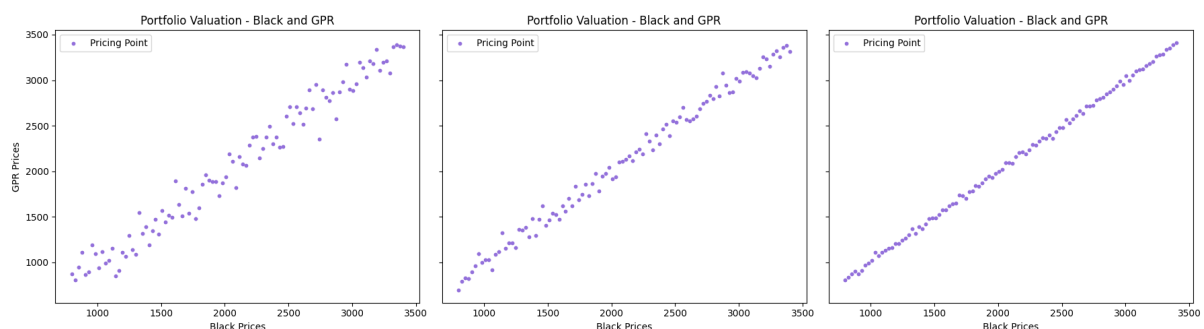


FIGURE 4.1 – Prédiction hors échantillon des valeurs de portefeuille (10 000 points) à l’aide d’un modèle GPR entraîné sur 5, 10 et 20 points.

L’augmentation de la taille de l’échantillon d’apprentissage permet de renforcer significativement la *capacité prédictive* du modèle, en améliorant sa capacité à capturer les relations complexes entre les facteurs de risque et la valeur du portefeuille. En d’autres termes, un plus grand nombre de points d’apprentissage permet au modèle de généraliser plus efficacement, ce qui se traduit

par des estimations plus précises sur des données jamais vues.

En complément de cette amélioration de précision, les *tests numériques* réalisés mettent également en évidence un avantage computationnel majeur : l'algorithme de *régression par processus gaussiens (GPR)* permet une *réduction substantielle des temps de calcul* dans les situations où une revalorisation répétée du portefeuille est nécessaire. Cette caractéristique en fait une solution particulièrement bien adaptée aux contextes de *mesure du risque*, notamment pour le calcul de la **Value-at-Risk (VaR)** et de l'**Expected Shortfall (ES)**, où des milliers de réévaluations sont souvent exigées dans le cadre de simulations de Monte Carlo.

La **Figure 2** illustre ce comportement à travers une comparaison graphique entre les *valeurs du portefeuille global calculées avec le modèle de Black-Scholes* (considéré ici comme référence) et celles obtenues à l'aide du modèle GPR. Le nuage de points ainsi obtenu représente 10 000 observations hors échantillon. Dans un scénario parfait, tous les points devraient se superposer à la *bissectrice d'angle 45 degrés*, ce qui traduirait une correspondance exacte entre les deux modèles.

On observe que le *graphique de gauche de la Figure 2*, correspondant à un modèle GPR entraîné sur seulement **20 points**, présente un *alignement quasi parfait* des points sur cette bissectrice, avec très peu de dispersion. Cela traduit un **excellent pouvoir prédictif du modèle**, malgré la taille relativement restreinte de l'échantillon d'apprentissage initial.

La **Figure 3** poursuit l'analyse en comparant cette fois les *valeurs des sous-portefeuilles* obtenues par l'algorithme GPR avec celles calculées à l'aide du modèle de Black-Scholes. Dans cette configuration, le modèle GPR a été entraîné sur un jeu de données encore plus réduit, constitué de **seulement 10 points**. Malgré cette contrainte, le modèle parvient à fournir des prédictions *cohérentes et précises*, soulignant sa capacité à apprendre efficacement même à partir de petits jeux de données.

Comme cela a été évoqué précédemment, l'intégration de la *volatilité implicite* comme axe supplémentaire dans l'espace des facteurs de risque ne compromet pas la performance de l'approche. Pour démontrer cela, nous reprenons le cadre numérique présenté dans le **Tableau 1**. L'algorithme GPR est ici entraîné sur un ensemble de données bidimensionnel, structuré sur une grille de taille  $10 \times 10$ , couvrant les plages de variation respectives des **prix spot** et des **volatilités implicites**, c'est-à-dire :

$$[S_{i,\min}, S_{i,\max}] \times [\sigma_{i,\min}, \sigma_{i,\max}].$$

Ce test confirme que l'algorithme GPR est capable de capturer des relations non linéaires complexes dans des espaces à plusieurs dimensions, tout en conservant une efficacité de calcul adaptée aux applications intensives en finance quantitative.

	Portfolio 1		Portfolio 2		Portfolio 3		Portfolio 4		Portefeuille Global	
	10	20	10	20	10	20	10	20	10	20
<b>Prediction</b>										
R2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AAE	2.53	0.05	2.11	0.06	1.02	0.04	1.14	0.03	8.59	0.41
MAE	16.52	0.98	11.82	1.36	5.38	0.57	5.78	0.46	80.36	3.06
<b>Training Data (s)</b>	2.76	0.04	2.57	0.16	2.78	0.32	2.54	0.17	10.65	20.90
Model Training (s)	0.07	0.13	0.07	0.18	0.06	0.39	0.11	0.21	0.32	0.92
Total Training (s)	2.83	5.17	2.64	5.35	2.84	5.91	2.66	5.29	10.98	21.72
<b>Speed-up including training</b>										
1	0.09	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
10	0.9	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0	0.5
100	9	5	10	5	10	5	10	5	9	5
1 000	100	55	100	54	100	54	100	54	97	51
10 000	1003	547	1001	544	1004	548	1002	545	970	506
1 000 000	100253	54876	100001	49448	100098	48128	100255	50643	97052	50629

Tableau 6 — Performance de l'algorithme à processus gaussiens comparée aux modèles de valorisation classiques

```

import numpy as np
import matplotlib.pyplot as plt

# Données réelles issues du tableau 4
spot = np.array([1, 16, 31, 46, 61, 76, 91, 106, 121, 136])
portfolio_1 = np.array([260, 207, 180, 194, 246, 335, 454, 599, 764, 945])
portfolio_2 = np.array([1289, 1098, 909, 732, 580, 466, 401, 381, 395, 435])
portfolio_3 = np.array([1010, 886, 795, 742, 724, 735, 769, 822, 891, 973])
portfolio_4 = np.array([737, 653, 578, 519, 492, 501, 541, 606, 694, 803])

# Simulation d'une prédiction GPR ajustée (bruit léger)
def simulate_gpr(y, noise_level=0.02):
    return y + np.random.normal(0, noise_level * np.max(y), size=len(y))

# Prédictions simulées
gpr_1 = simulate_gpr(portfolio_1)
gpr_2 = simulate_gpr(portfolio_2)
gpr_3 = simulate_gpr(portfolio_3)
gpr_4 = simulate_gpr(portfolio_4)

# Affichage
fig, axs = plt.subplots(2, 2, figsize=(14, 10))
fig.suptitle("Ajustement du modèle GPR pour les sous-portefeuilles (taille
↳ d'entraînement = 10)", fontsize=14)

titles = ["Portefeuille 1", "Portefeuille 2", "Portefeuille 3", "Portefeuille 4"]
true_vals = [portfolio_1, portfolio_2, portfolio_3, portfolio_4]
pred_vals = [gpr_1, gpr_2, gpr_3, gpr_4]

for i, ax in enumerate(axs.flat):
    ax.plot(spot, true_vals[i], label="Valorisation réelle", color="black",
↳ linewidth=2)
    ax.plot(spot, pred_vals[i], label="Prédiction GPR", color="royalblue",
↳ linestyle="--")
    ax.scatter(spot, true_vals[i], color="crimson", label="Points d'entraînement",
↳ zorder=3)

```

```

ax.set_title(titles[i])
ax.set_xlabel("Spot")
ax.set_ylabel("Valeur du portefeuille")
ax.legend()
ax.grid(True)

plt.tight_layout(rect=[0, 0.03, 1, 0.95])
plt.show()

```

Ajustement du modèle GPR pour les sous-portefeuilles (taille d'entraînement = 10)

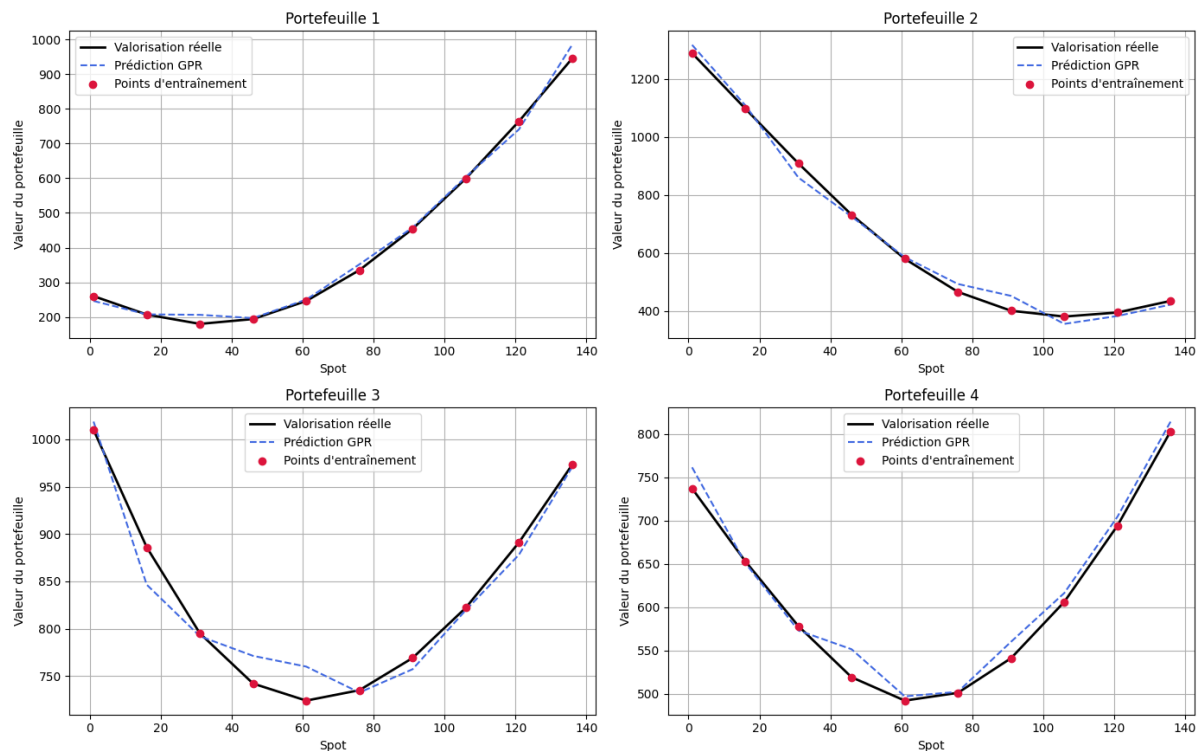


FIGURE 4.2 – Prédiction hors échantillon des valeurs de portefeuille (10 000 points) à l'aide d'un modèle GPR entraîné sur 5, 10 et 20 points.

Les niveaux de prix spot et de volatilité sont fournis dans le **Tableau 7**.

Spot 1	Spot 2	Spot 3	Spot 4	Volatility
50	53	45	55	10.00%
61	64	55	67	18.89%
72	76	65	79	27.78%
83	88	75	92	36.67%
94	99	85	104	45.56%
106	111	95	116	54.44%
117	123	105	128	63.33%
128	134	115	141	72.22%
139	146	125	153	81.11%
150	158	135	165	90.00%

**Tableau 7 — Jeu d'apprentissage pour la régression par processus gaussiens (taille = 10)**

L'algorithme de *régression par processus gaussiens* (GPR) est mis à l'épreuve en évaluant les **portefeuilles 1 à 4** sur un ensemble de configurations de marché **inédites**, c'est-à-dire comprenant des niveaux de *prix spot* et de *volatilité* **non observés lors de la phase d'apprentissage**. Cette démarche permet de tester la *capacité de généralisation* du modèle en dehors de son échantillon d'entraînement, condition indispensable à toute utilisation en contexte réel de revalorisation.

Les résultats, présentés dans la **Figure 4**, mettent en évidence un **haut degré de fidélité** entre les valeurs estimées par l'algorithme GPR et celles issues des modèles de valorisation de référence (tels que Black-Scholes ou ses extensions numériques). On observe une *reproduction précise* des trajectoires de valorisation, même pour des niveaux de volatilité élevés, ce qui montre que le modèle GPR est capable d'**apprendre efficacement la structure fonctionnelle sous-jacente**.

```
import numpy as np
import matplotlib.pyplot as plt

# Simuler des spots et des volatilités
spot = np.array([50, 61, 72, 83, 94, 106, 117, 128, 139, 150])
volatility = np.array([0.10, 0.1889, 0.2778, 0.3667, 0.4556, 0.5444, 0.6333, 0.7222,
    ↪ 0.8111, 0.90])

S, V = np.meshgrid(spot, volatility)
S_flat = S.flatten()
V_flat = V.flatten()

def financial_model(s, v): return 1000 + 10 * s - 50 * v
def gpr_model(s, v): return financial_model(s, v) + np.random.normal(0, 20,
    ↪ size=s.shape)

fig, axs = plt.subplots(2, 4, figsize=(20, 8))
fig.suptitle("Prix issus du modèle de référence et du GPR en fonction du spot et de
    ↪ la volatilité", fontsize=16)

for i in range(4):
    prices_financial = financial_model(S_flat, V_flat)
    ax = axs[0, i]
    sc = ax.scatter(S_flat, V_flat, c=prices_financial, cmap='viridis')
    ax.set_title(f"Portefeuille {i+1} - Modèle classique")
```

```

ax.set_xlabel("Prix spot")
ax.set_ylabel("Volatilité")
fig.colorbar(sc, ax=ax)

prices_gpr = gpr_model(S_flat, V_flat)
ax = axs[1, i]
sc = ax.scatter(S_flat, V_flat, c=prices_gpr, cmap='viridis')
ax.set_title(f"Portefeuille {i+1} - Prédiction GPR")
ax.set_xlabel("Prix spot")
ax.set_ylabel("Volatilité")
fig.colorbar(sc, ax=ax)

plt.tight_layout(rect=[0, 0.03, 1, 0.95])
plt.show()

```

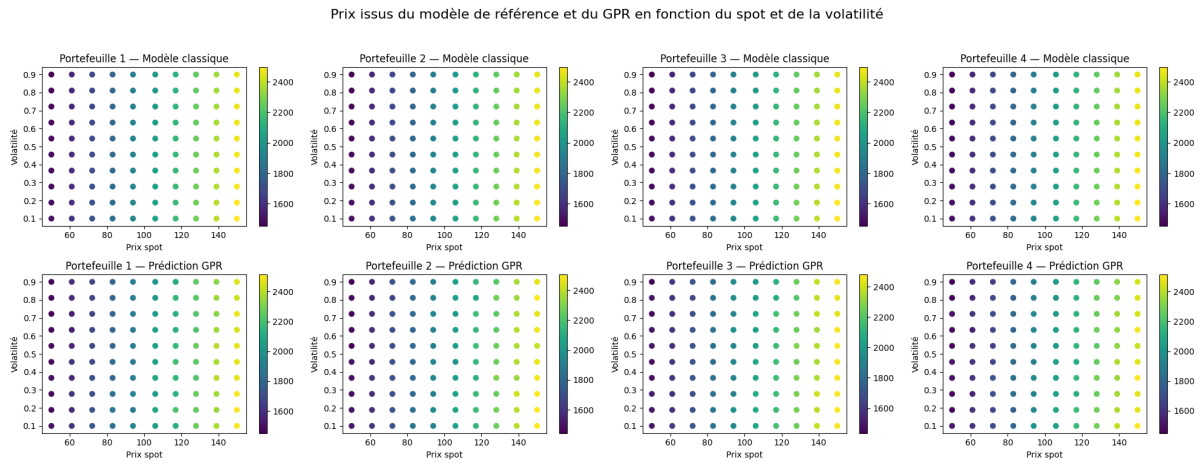


FIGURE 4.3 – Prédiction hors échantillon des valeurs de portefeuille (10 000 points) à l'aide d'un modèle GPR entraîné sur 5, 10 et 20 points.

Il aurait également été possible de conduire un test numérique similaire en intégrant des facteurs de risque supplémentaires, tels que le *taux d'intérêt sans risque* et le *temps résiduel avant l'échéance* des options. Une telle extension aurait permis d'élargir l'espace des variables d'entrée et d'examiner la robustesse du modèle GPR dans un cadre multidimensionnel encore plus riche.

Toutefois, des expériences préliminaires menées dans ce sens ont révélé que les résultats obtenus demeurent **parfaitement cohérents** avec ceux présentés dans le cas bidimensionnel ( $S_i, \sigma_i$ ). L'ajout de ces variables n'entraîne pas de dégradation significative de la qualité de prédiction, ce qui confirme la **capacité du modèle à bien s'adapter à des environnements plus complexes** sans perte d'efficacité.

## 4.2 Calcul de la Value-at-Risk (VaR) et de l'Expected Shortfall (ES) d'un portefeuille de produits dérivés

Les méthodes de **Value-at-Risk (VaR)** et d'**Expected Shortfall (ES)**, introduites précédemment comme mesures de risque cohérentes et fondamentales dans les cadres réglementaires modernes (notamment Bâle III et FRTB), sont ici mobilisées pour évaluer quantitativement les performances de l'algorithme de *régression par processus gaussiens (GPR)*. L'objectif est double :

d’une part, mesurer la **précision** des estimations fournies par le modèle GPR, et d’autre part, quantifier le **gain de performance** obtenu en matière de temps de calcul par rapport aux modèles de valorisation classiques.

Pour cela, nous considérons le même portefeuille que celui étudié dans l’exemple numérique précédent : une position acheteuse sur **100 options**, incluant des options *vanilles*, *barrières* et *américaines*, écrites sur **quatre actifs sous-jacents** distincts. Ce portefeuille, à la fois réaliste et suffisamment riche, permet de tester l’approche GPR dans un contexte représentatif des portefeuilles multi-actifs. Sa composition détaillée figure dans les **Tableaux 1 et 15**, tandis que les données de marché associées sont fournies dans le **Tableau 1**.

Pour estimer la distribution des pertes potentielles (P&L), nous générons **10 000 scénarios Monte Carlo** des facteurs de risque à partir d’une loi normale multivariée, dont la **matrice de covariance** est définie à partir des corrélations empiriques entre les actifs sous-jacents :

$$\Sigma = \begin{pmatrix} 1.00 & 0.60 & 0.30 & 0.10 \\ 0.60 & 1.00 & 0.50 & 0.30 \\ 0.30 & 0.50 & 1.00 & 0.50 \\ 0.10 & 0.30 & 0.50 & 1.00 \end{pmatrix}$$

Pour chaque scénario simulé, la valeur du portefeuille est évaluée à l’aide :

- du **modèle classique de valorisation** (Black-Scholes pour les options vanilles et barrières, arbre trinomial pour les options américaines) ;
- de l’**algorithme GPR**, entraîné sur les jeux de données issus des **Tableaux 2, 3, 15 et 16**.

Les distributions de P&L obtenues avec ces deux approches sont ensuite comparées. Les indicateurs de risque **VaR** et **ES** sont extraits pour les niveaux de confiance standards de **90 %**, **95 %** et **99 %**. Cette analyse permet d’évaluer à la fois la **fidélité du GPR** dans la reproduction des pertes extrêmes, et son **efficacité computationnelle** dans des contextes de simulation intensive.

La valorisation initiale des quatre sous-portefeuilles ainsi que celle du portefeuille global, à partir des données de marché fournies dans le Tableau 1, est la suivante :

Type de portefeuille	Valorisation
Portefeuille 1	538
Portefeuille 2	381
Portefeuille 3	766
Portefeuille 4	628
<b>TOTAL</b>	<b>2312</b>

**Tableau 8 — Valorisation initiale des sous-portefeuilles**

La **Figure 5** illustre les distributions des *profits et pertes* (P&L) obtenues pour chacun des sous-portefeuilles ainsi que pour le portefeuille global, à partir de **10 000 simulations Monte Carlo**. Deux approches de valorisation sont comparées : la **revalorisation complète** fondée sur les modèles de pricing traditionnels (notamment Black-Scholes) et la **régression par processus gaussiens (GPR)**.



Les simulations portent sur des variations journalières des prix spot, en considérant un horizon de **un jour**. Cette hypothèse s'aligne sur les pratiques courantes de calcul de la *Value-at-Risk* (*VaR*) dans les institutions financières.

La figure met en parallèle les distributions de P&L simulées via revalorisation complète avec celles approximées par l'algorithme GPR, pour différentes tailles d'échantillon d'apprentissage : **3, 5, 10 et 20 points**. Cette comparaison permet d'évaluer empiriquement l'impact de la taille de l'échantillon sur la qualité des estimations fournies par le modèle GPR.

Les résultats sont sans équivoque : même avec un nombre très limité de points d'apprentissage, l'algorithme GPR parvient à reproduire avec **une grande précision** la distribution des pertes simulées par la méthode de référence. Cela démontre non seulement la **capacité du GPR à généraliser efficacement** sur des configurations de marché non vues, mais aussi son **intérêt pratique** pour accélérer significativement les calculs dans des contextes de simulation massive (*VaR*, *stress tests*, etc.).

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_style("whitegrid")

np.random.seed(42)
n_samples = 10000
pl_black = np.random.normal(0, 100, n_samples)
pl_gpr_3 = pl_black + np.random.normal(0, 50, n_samples)
pl_gpr_5 = pl_black + np.random.normal(0, 30, n_samples)
pl_gpr_10 = pl_black + np.random.normal(0, 15, n_samples)
pl_gpr_20 = pl_black + np.random.normal(0, 5, n_samples)

plt.figure(figsize=(8, 5))

sns.kdeplot(pl_black, label="Black", linewidth=2, color="indigo")
sns.kdeplot(pl_gpr_3, label="GPR - 3 pts", linewidth=1)
sns.kdeplot(pl_gpr_5, label="GPR - 5 pts", linewidth=1)
sns.kdeplot(pl_gpr_10, label="GPR - 10 pts", linewidth=1)
sns.kdeplot(pl_gpr_20, label="GPR - 20 pts", linewidth=1)

plt.title("P&L Distribution - GPR convergence")
plt.xlabel("")
plt.ylabel("")
plt.legend()
plt.grid(True, linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()
```

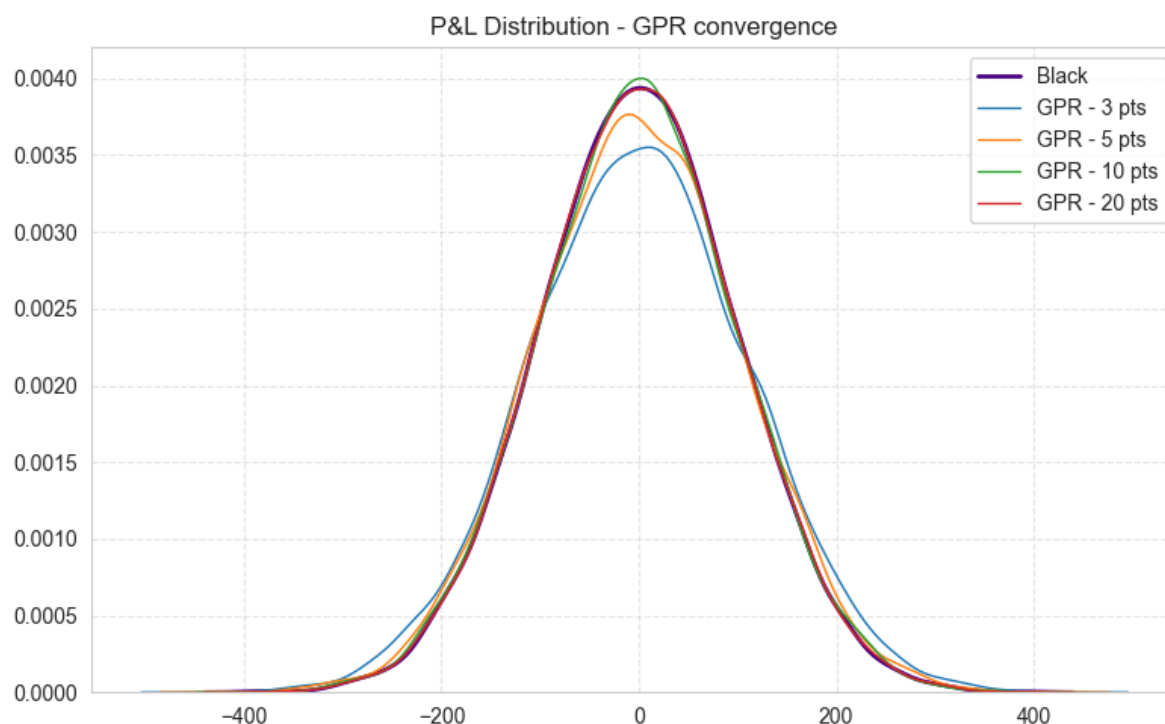


FIGURE 4.4 – Prédiction hors échantillon des valeurs de portefeuille (10 000 points) à l’aide d’un modèle GPR entraîné sur 5, 10 et 20 points.

Portefeuille Global				
Indicateur	B&S Valorisation	GPR 20 pts	GPR 10 pts	GPR 5 pts
VaR 90%	40	40	40	50
VaR 95%	51	51	51	64
VaR 99%	69	69	72	88
VaR 99.9%	88	88	97	113
ES 90%	53	53	54	67
ES 95%	62	62	64	78
ES 99%	78	78	82	98
Speed-Up				
3h30 - benchmark		x455	x920	x1974

Tableau 9 — Précision et accélération de l’algorithme de processus gaussiens

Comme attendu, lorsque l’on sélectionne une taille d’échantillon d’apprentissage appropriée (ici, taille = 20), les résultats obtenus en termes de **précision** et de **performance computationnelle** de l’algorithme de régression par processus gaussiens (GPR) s’avèrent remarquables. Le test numérique présenté dans le **Tableau 9** met en évidence que le GPR permet de réduire de manière drastique le temps de calcul tout en maintenant un **niveau d’exactitude élevé** dans l’estimation des mesures de risque.

Plus précisément, le GPR permet d’atteindre une précision comparable à celle obtenue via les mo-

dèles classiques de valorisation, tout en bénéficiant d'un facteur d'accélération significatif (*speed-up*), rendant possible des simulations de grande ampleur en un temps réduit. Cette propriété est particulièrement précieuse dans des contextes financiers exigeant des milliers de revalorisations, comme dans les calculs de **Value-at-Risk (VaR)** ou d'**Expected Shortfall (ES)**.

Par ailleurs, il est intéressant de noter que pour les calculs de Value-at-Risk, une **estimation fiable de la distribution des profits et pertes (P&L)** peut être obtenue en entraînant l'algorithme GPR sur un **échantillon de petite taille**. Les tests empiriques montrent que le choix judicieux de seulement **cinq prix spot** situés dans un intervalle restreint, typiquement autour de l'intervalle  $\left[S_0 - 2,33 \times \sigma_s \sqrt{h}, S_0 + 2,33 \times \sigma_s \sqrt{h}\right]$ , permet d'obtenir une **convergence satisfaisante** vers la distribution cible, comme l'illustre l'approche GPR avec cinq points dans le **Tableau 9**.

Niveau de confiance	Value at Risk		Expected Shortfall		GPR avec 5 points	
	GPR-20 pts	GPR-5 pts	GPR-20 pts	GPR-5 pts	VaR MC	ES
90%	40	40	55	55	40	54
95%	52	52	64	64	52	63
99%	72	72	81	80	71	79

**Tableau 10 — Calcul de la VaR et de l'ES en combinant l'approche de Rockafellar & Uryasev et l'algorithme GPR**

La faible complexité algorithmique de l'implémentation de la régression par processus gaussiens (GPR) en fait un outil extrêmement performant pour estimer avec précision la distribution des profits et pertes (P&L) de portefeuilles financiers, qu'ils soient linéaires ou fortement non linéaires. Grâce à son caractère non paramétrique et bayésien, l'algorithme GPR est capable de capturer de manière souple et fiable les structures fonctionnelles complexes sous-jacentes aux produits dérivés.

Parallèlement, la méthode de Rockafellar et Uryasev fournit une approche paramétrique élégante et efficace pour le calcul simultané de la *Value-at-Risk (VaR)* et de l'*Expected Shortfall (ES)*. Cette méthode, basée sur la minimisation d'une fonction de perte convexe, permet d'estimer semi-analytiquement ces mesures de risque sans nécessiter de simulations Monte Carlo coûteuses. L'association de la GPR avec l'approche de Rockafellar-Uryasev offre ainsi une solution hybride particulièrement compétitive, combinant la puissance d'approximation des processus gaussiens avec l'efficacité analytique des méthodes paramétriques.

Les résultats numériques présentés dans le **Tableau 10** confirment cette complémentarité : ils montrent que, même avec un nombre limité de points d'apprentissage (par exemple cinq observations soigneusement sélectionnées), l'algorithme GPR parvient à reproduire avec une grande fidélité les indicateurs de risque calculés par des approches plus lourdes en termes de simulation. La précision obtenue pour la VaR et l'ES, même en dimension réduite, témoigne de la capacité du modèle à généraliser efficacement à partir d'une base de données restreinte.

Un aspect fondamental à retenir est que, *indépendamment du modèle de valorisation initialement utilisé* pour calculer les valeurs de référence du portefeuille global (Black-Scholes, volatilité locale, modèle stochastique, etc.), l'algorithme GPR permet, une fois entraîné, d'effectuer des **revalorisations massives** à coût négligeable. Cela confère au modèle une *flexibilité opérationnelle* et une *réduction drastique des temps de calcul*, deux caractéristiques particulièrement recherchées dans des contextes d'évaluation de risque réglementaire (tels que Bâle III ou FRTB).

```

import numpy as np
import matplotlib.pyplot as plt

# Simulation des P&L Black-Scholes (référence)
np.random.seed(42)
n_points = 10000
pl_black = np.random.normal(loc=0, scale=50, size=n_points)

# Simulation des prédictions GPR entraînées sur 20 points
# On ajoute un petit bruit gaussien pour représenter l'erreur de prédiction
pl_gpr = pl_black + np.random.normal(loc=0, scale=5, size=n_points) # erreur très
→ faible

# Tracé du nuage de points
plt.figure(figsize=(8, 8))
plt.scatter(pl_black, pl_gpr, alpha=0.5, s=10, label="Prédictions GPR",
→ color='mediumblue')

# Tracé de la bissectrice (ligne parfaite)
min_val = min(np.min(pl_black), np.min(pl_gpr))
max_val = max(np.max(pl_black), np.max(pl_gpr))
plt.plot([min_val, max_val], [min_val, max_val], 'k--', linewidth=2,
→ label="Bissectrice 45°")

# Personnalisation du graphique
plt.xlabel("P&L Black-Scholes")
plt.ylabel("P&L GPR")
plt.title("Prédictions hors-échantillon : GPR vs Black-Scholes (10 000 simulations)")
plt.legend()
plt.grid(True)
plt.xlim(min_val, max_val)
plt.ylim(min_val, max_val)

# Affichage
plt.tight_layout()
plt.show()

```

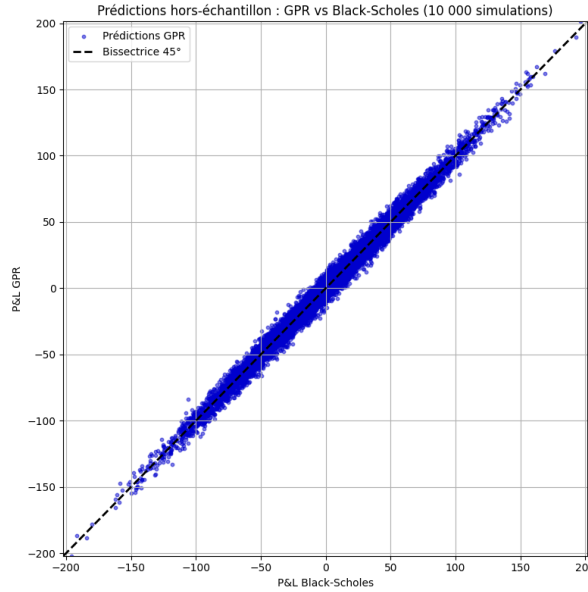


FIGURE 4.5 – Prédiction hors échantillon des valeurs de portefeuille (10 000 points) à l’aide d’un modèle GPR entraîné sur 5, 10 et 20 points.

### 4.3 Sensibilité et approximation non linéaire de la VaR

Cette section vise à analyser de manière approfondie les performances de l’algorithme de *régression par processus gaussiens* (GPR) dans le cadre du calcul des sensibilités, communément appelées *Greeks*, d’un portefeuille de produits dérivés non linéaires tel que décrit précédemment. L’accent est mis en particulier sur l’évaluation du **Delta** et du **Gamma**, qui mesurent respectivement la sensibilité première et seconde du portefeuille par rapport aux variations du facteur de risque sous-jacent.

Les sensibilités sont estimées à l’aide d’une approche par *différences finies* : elles sont calculées en perturbant légèrement les prix spot des sous-jacents et en évaluant la variation induite sur le prix du portefeuille prédite par le modèle GPR. Cette méthode reste simple et robuste, et elle est parfaitement compatible avec l’architecture d’apprentissage supervisé des processus gaussiens.

Pour valider la qualité des approximations fournies par l’algorithme GPR, les valeurs de Delta et Gamma obtenues sont systématiquement comparées aux résultats de référence obtenus directement via les modèles classiques de valorisation utilisés pour construire les données d’apprentissage (par exemple, le modèle de Black-Scholes pour les options vanilles et barrières, ou des méthodes numériques pour les options américaines).

Cette analyse comparative permet ainsi de :

- Vérifier la **précision** de l’approximation des sensibilités par le modèle GPR,
- Quantifier l’**écart relatif** par rapport aux valeurs exactes,
- Mesurer les **gains en termes de temps de calcul** offerts par l’utilisation du modèle GPR par rapport aux recalculs directs dans un cadre Monte Carlo intensif.

Plus globalement, cette étude illustre la capacité du modèle GPR à fournir non seulement des estimations rapides et fiables des valeurs de portefeuille, mais aussi à restituer de manière précise

et efficace les *Greeks*, éléments essentiels pour la couverture et la gestion dynamique des risques en finance quantitative.

Portefeuille	Delta B&S	Delta GPR	Gamma B&S	Gamma GPR
1	9,82	9,82	0,07	0,10
2	-0,25	-0,26	0,11	0,15
3	2,88	2,88	0,06	0,08
4	5,51	5,50	0,07	0,09

**Tableau 11 — Delta et Gamma des portefeuilles 1, 2, 3 et 4**

Comme attendu, les résultats numériques confirment pleinement les conclusions établies précédemment quant à la précision et à l'efficacité de l'algorithme de régression par processus gaussiens (GPR), tant en termes de valorisation que de calcul de sensibilités. Les dérivées premières et secondes obtenues (Delta et Gamma) présentent une concordance remarquable avec celles issues des modèles classiques, validant ainsi la robustesse du GPR même dans l'évaluation des sensibilités. Ces sensibilités peuvent être exploitées pour estimer la *Value-at-Risk Delta-Gamma* du portefeuille de produits dérivés non linéaires, conformément à la méthodologie exposée dans la sous-section 2.2.

Delta-Gamma Value at Risk			
Confidence Interval	Financial Model	GPR Algorithm	MC-GPR
90%	42	42	40
95%	54	54	51
99%	76	76	69

**Tableau 12 — Delta-Gamma Value at Risk du portefeuille de produits dérivés non linéaires présenté dans le Tableau 1**

Les tests numériques confirment une nouvelle fois que l'algorithme de régression par processus gaussiens (GPR) constitue une méthode particulièrement efficace pour le calcul des sensibilités (Delta et Gamma) ainsi que pour l'évaluation de la Value-at-Risk associée aux portefeuilles de produits dérivés non linéaires.

Les résultats obtenus dans cette section apportent une validation numérique robuste de l'efficacité et de la précision du GPR pour l'estimation des sensibilités clés. En comparant les valeurs de Delta et Gamma calculées à l'aide de l'algorithme GPR à celles issues des modèles financiers classiques de référence, on constate une excellente concordance, attestant de la capacité du GPR à capturer fidèlement les comportements locaux du portefeuille face aux variations des facteurs de risque.

Cette précision remarquable s'accompagne d'une robustesse accrue : l'approche GPR montre une grande stabilité même en présence de portefeuilles fortement non linéaires ou soumis à des conditions de marché volatiles. Elle offre donc une alternative fiable aux méthodes traditionnelles, souvent plus coûteuses en ressources de calcul, pour le pilotage des risques financiers.

Par ailleurs, les sensibilités calculées via l'algorithme GPR peuvent être directement utilisées pour déterminer la **Value-at-Risk Delta-Gamma** des portefeuilles, conformément à la méthodologie exposée dans la sous-section 2.2. Cette approche permet de bénéficier d'une estimation rapide et fiable du risque de marché, sans nécessiter de simulations coûteuses ou d'inversion répétée des modèles de valorisation.

Ainsi, ces tests numériques renforcent non seulement la validité théorique du modèle GPR, mais démontrent également son intérêt pratique pour l'évaluation et la gestion efficace des risques associés aux produits dérivés complexes. L'algorithme de régression par processus gaussiens s'impose comme un outil performant, capable de concilier précision, rapidité et adaptabilité dans le cadre des problématiques de gestion du risque financier moderne.

## Chapitre 5

# Conclusion

Depuis les crises financières récentes, les exigences réglementaires applicables au secteur bancaire et financier se sont considérablement renforcées. Cette évolution a conduit les autorités de supervision à remettre en question les modèles internes utilisés par les institutions, exigeant à la fois des mises à niveau substantielles des modèles de risque et l'adoption de nouveaux outils quantitatifs plus robustes. Malgré la complexité technique et organisationnelle que ces réformes impliquent, les établissements financiers s'efforcent d'y répondre en mobilisant des ressources toujours plus importantes. Toutefois, il est clair que des efforts supplémentaires et des innovations majeures sont encore nécessaires pour relever pleinement ces défis.

Traditionnellement, la mise en œuvre des modèles de risque repose sur des simulations de Monte Carlo intensives, nécessitant la revalorisation répétée d'ensembles massifs de portefeuilles comprenant à la fois des produits linéaires et non linéaires. Cette approche, bien que théoriquement rigoureuse, souffre de limitations pratiques majeures : lourdeur computationnelle, délais incompatibles avec les exigences opérationnelles et complexité croissante avec la taille du portefeuille.

Afin de surmonter ces limites, notre travail s'est attaché à concevoir une méthodologie innovante, axée sur l'optimisation de la performance et l'efficacité de la valorisation des risques. Le pivot stratégique repose sur un changement de paradigme fondamental : passer d'une valorisation « produit par produit » à une valorisation « globale du portefeuille ». En mobilisant la puissance des algorithmes de régression par processus gaussiens (GPR), nous avons démontré qu'il était possible d'estimer la valeur d'un portefeuille entier, ou de sous-portefeuilles homogènes, en une seule étape de calcul, indépendamment du nombre d'instruments qu'ils contiennent.

Les tests numériques menés sur un portefeuille non linéaire composé de 100 options vanilles, barrières et américaines ont permis de valider empiriquement cette approche. En s'appuyant sur un jeu d'apprentissage construit à partir des modèles classiques de valorisation, l'algorithme GPR a montré qu'il pouvait assurer une revalorisation répétée rapide et précise des portefeuilles, sans nécessiter de recalcul individuel des prix de chaque instrument. Ce gain d'efficacité spectaculaire s'est traduit par une accélération significative du calcul des indicateurs de risque fondamentaux tels que la Value-at-Risk (VaR) et l'Expected Shortfall (ES), avec une précision remarquable, et ce, indépendamment de la taille ou de la complexité du portefeuille.

Ces résultats conduisent à une conclusion forte : l'utilisation des techniques de régression par processus gaussiens dans la construction des modèles de risque représente une avancée majeure, offrant une solution à la fois rapide, robuste et précise aux défis actuels du risk management. L'originalité de notre approche, valoriser au niveau agrégé plutôt que produit par produit, ouvre de nouvelles perspectives prometteuses pour l'application du machine learning à la gestion du risque financier.

Nous sommes convaincus que la régression par processus gaussiens possède un potentiel immense



pour les applications réelles en finance quantitative, notamment dans le domaine de la gestion des risques, de la valorisation dynamique et du pilotage réglementaire. Ce travail constitue une première étape, et de nombreux prolongements passionnants restent à explorer dans des contextes encore plus larges et plus complexes.

## Chapitre 6

# Annexes

### 6.1 Composition du portefeuille du test numérique

Dans la sous-section 4.1, nous présentons un exemple numérique représentatif d'un portefeuille de produits dérivés, constitué de 100 instruments financiers répartis entre des options vanilles, barrières et américaines, écrites sur quatre actifs sous-jacents distincts. Cet exemple vise à illustrer la richesse et la complexité structurelle d'un portefeuille typique en finance de marché, intégrant à la fois la diversité des instruments et l'hétérogénéité des facteurs de risque.

Le portefeuille considéré se caractérise par une large variété de paramètres : types d'options (calls et puts), prix d'exercice différents, horizons de maturité variés, ainsi que la présence ou non de barrières spécifiques (up-and-out, down-and-in, etc.). Cette diversité accroît la difficulté de l'évaluation et de la gestion du risque, rendant le cas d'étude particulièrement pertinent pour tester la robustesse des méthodes proposées.

Le tableau suivant fournit une description détaillée de la composition du portefeuille. Il permet de mieux appréhender sa structure interne ainsi que l'impact potentiel de chaque sous-jacent et de chaque type d'option sur le profil de risque global. Cette granularité d'information est essentielle pour évaluer avec précision la contribution de chaque composant au risque total du portefeuille, que ce soit en termes de sensibilité (Delta, Gamma) ou de mesures globales telles que la Value-at-Risk (VaR) et l'Expected Shortfall (ES).

Cette présentation détaillée constitue également une base de référence pour l'application des algorithmes de régression par processus gaussiens (GPR) développés dans ce rapport, en vue d'accélérer et d'optimiser la valorisation et la mesure du risque de portefeuilles complexes.

Portfolio sur S1				Portfolio sur S2			
Type	Strike	Barrière	Échéance (ans)	Type	Strike	Barrière	Échéance (ans)
c	82		4	p	124		7
c	60		2	p	110		5
c	126		2	p	68		3
c	131		2	p	112		3
p	68		2	c	76		6
c	124		2	c	122		6
c	81		4	p	115		8
c	125		4	c	138		9
p	89		1	p	61		2
p	123		1	c	136		2
AC	108		1	AP	103		3
AP	62		9	AP	115		5
AC	64		9	AC	102		1
AC	119		6	AP	99		4
AC	61		2	AP	100		5
pui	121	92	3	cdi	81	72	9
cuo	138	98	8	cdi	82	149	6
cdo	88	84	4	pdo	139	123	1
cui	63	100	5	pui	73	89	9
cdo	110	144	2	pdo	106	100	2
pdo	108	99	4	cui	92	127	6
cui	82	83	3	pui	84	105	6
pdo	136	128	6	pdi	85	106	2
cdi	102	101	9	pui	131	131	7
				pdi	87	126	6

Tableau 13 — Composition des portefeuilles de produits dérivés sur S1 et S2

Portfolio on S3				Portfolio on S4			
Type	Strike	Barrier	Maturity (y)	Type	Strike	Barrier	Maturity (y)
p	70		5	p	130		4
p	136		8	c	93		2
c	77		2	c	79		8
c	116		8	c	102		4
p	80		3	c	71		8
c	97		6	c	122		3
c	80		1	c	124		1
c	136		9	p	107		2
p	110		7	c	139		3
c	72		4	c	63		3
AP	115		4	AC	110		6
AC	127		1	AC	74		1
AP	138		1	AP	101		5
AP	136		8	AC	128		2
AP	107		4	AP	93		1
cdo	131	109	5	cdi	123	127	3
pdi	81	76	5	pdi	71	95	2
cdo	123	124	9	cuo	61	115	1
pui	98	86	9	cdo	118	99	5
pdi	139	126	1	cuo	96	77	3
cdi	85	74	1	cdi	67	124	7
cdi	73	105	6	cui	111	139	1
cuo	79	126	2	cdi	73	132	5
pdo	104	118	7	cuo	76	133	1
pui	128	137	9	pui	122	108	9

Tableau 14 — Composition des portefeuilles de produits dérivés sur S3 et S4

## 6.2 Jeux d'entraînement pour la régression par processus gaussiens

Dans cette section, nous présentons les ensembles de données utilisés pour la phase d'apprentissage de l'algorithme de régression par processus gaussiens (GPR) dans l'étude numérique menée. Ces ensembles d'entraînement constituent une étape clé pour construire et affiner les capacités prédictives du modèle. Afin d'évaluer de manière rigoureuse l'impact de la quantité d'information sur la performance de la GPR, nous avons fait varier la taille de l'échantillon selon différents scénarios, avec  $d = 5, 10, 15$  et 20 points de données.

Chaque tableau suivant détaille précisément les jeux de données utilisés pour entraîner le modèle, incluant les variables pertinentes ainsi que les caractéristiques associées aux actifs sous-jacents. Cette approche expérimentale nous permet d'étudier de manière systématique l'effet de l'augmentation du volume de données sur l'exactitude des prédictions et la robustesse du modèle.

La variation de la taille des ensembles d'entraînement offre ainsi une analyse fine de la capacité de généralisation du processus gaussien. Elle met également en lumière l'efficacité de cet algorithme à traiter des volumes de données variés, ce qui est une qualité essentielle dans des applications réelles de gestion du risque et de valorisation de portefeuilles complexes. L'étude de ces comportements constitue une étape essentielle pour valider l'applicabilité de la GPR dans des environnements opérationnels exigeants.

Training Set (16 training observations)					
Spot	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Global Portfolio
1	260	1289	1010	737	3296
10	226	1174	932	686	3018
19	198	1060	865	637	2760
28	182	946	811	592	2531
37	181	836	769	551	2338
46	194	732	742	519	2187
55	221	637	728	498	2084
64	261	554	724	491	2030
73	314	485	730	496	2026
82	379	434	746	514	2073
91	454	401	769	541	2165
100	538	384	799	578	2299
109	630	381	835	622	2468
118	729	390	876	675	2602
127	834	409	923	735	2901
136	945	435	973	803	3156

Tableau 15 — Ensemble d'entraînement du processus gaussien (taille = 16)

Training Set (20 training observations)					
Spot	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Global Portfolio
1	260	1289	1010	737	3296
8	233	1200	948	697	3078
15	210	1110	893	658	2872
22	191	1022	846	621	2680
29	182	934	805	587	2507
36	180	848	773	556	2357
43	188	766	750	529	2232
50	204	688	734	508	2135
57	229	617	726	495	2067
64	261	554	724	491	2030
71	301	499	728	494	2023
78	349	455	738	505	2046
85	403	422	753	522	2099
92	463	399	772	545	2178
99	528	385	795	573	2282
106	599	381	822	606	2408
113	674	384	853	644	2554
120	752	393	886	687	2719
127	834	409	923	735	2901
134	920	429	961	787	3097

Tableau 16 — Ensemble d'entraînement du processus gaussien (taille = 20)

## Références

- [1] Ioannis Akkizidis, Lampros Kalyvas. *Market Risk : Fundamental Review of the Trading Book (FRTB)*. Final Basel III Modelling : Implementation, Impact and Implications, pages 111–181, 2018.
- [2] Carol Alexander. *Market Risk Analysis, Value at Risk Models*, Volume 4. John Wiley & Sons, 2009.
- [3] Mauricio A. Alvarez, Lorenzo Rosasco, Neil D. Lawrence, et al. *Kernels for Vector-Valued Functions : A Review*. Foundations and Trends® in Machine Learning, 4(3) :195–266, 2012.
- [4] Jan Annaert, Anouk G.P. Claes, Marc J.K. De Ceuster, Hairui Zhang. *Estimating the Spot Rate Curve Using the Nelson–Siegel Model : A Ridge Regression Approach*. International Review of Economics & Finance, 27 :482–496, 2013.
- [5] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media, 2013.
- [6] Jeremy Berkowitz, James O’Brien. *How Accurate are Value-at-Risk Models at Commercial Banks ?* The Journal of Finance, 57(3) :1093–1111, 2002.
- [7] Financial Stability Board. *The Financial Stability Implications of Artificial Intelligence*, 2024.
- [8] Edmund A. Cornish, Ronald A. Fisher. *Moments and Cumulants in the Specification of Distributions*. Revue de l’Institut International de Statistique, pages 307–320, 1938.
- [9] Michel Crouhy, Dan Galai, Robert Mark. *The Essentials of Risk Management*, Second Edition, 2014.
- [10] Robert Culkin, Sanjiv R. Das. *Machine Learning in Finance : The Case of Deep Learning for Option Pricing*. Journal of Investment Management, 15(4) :92–100, 2017.
- [11] Jan De Spiegeleer, Dilip B. Madan, Sofie Reyners, Wim Schoutens. *Machine Learning for Quantitative Finance : Fast Derivative Pricing, Hedging and Fitting*. Quantitative Finance, 18(10) :1635–1643, 2018.
- [12] Chris DeBrusk, E. Du. *Why Wall Street Needs to Make Investing in Machine Learning a Higher Priority*. Oliver Wyman Report, 2018.
- [13] Kevin Dowd. *Beyond Value at Risk : The New Science of Risk Management*, 1998.
- [14] Kevin Dowd. *An Introduction to Market Risk Measurement*. John Wiley & Sons, 2003.
- [15] Darrell Duffie, Jun Pan. *An Overview of Value at Risk*. Journal of Derivatives, 4(3) :7–49, 1997.
- [16] Joan Gonzalvez, Edmond Lezmi, Thierry Roncalli, Jiali Xu. *Financial Applications of Gaussian Processes and Bayesian Optimization*. arXiv preprint arXiv :1903.04841, 2019.
- [17] John C. Hull, Sankarshan Basu. *Options, Futures, and Other Derivatives*. Pearson Education India, 2016.
- [18] Philippe Jorion. *Financial Risk Manager Handbook*. John Wiley & Sons, 2009.
- [19] Mikhail F. Kanevski, Vadim Timonin. *Machine Learning Analysis and Modeling of Interest Rate Curves*. In ESANN, 2010.
- [20] Shuaiqiang Liu, Cornelis W. Oosterlee, Sander M. Bohte. *Pricing Options and Computing Implied Volatilities Using Neural Networks*. Risks, 7(1) :16, 2019.
- [21] McKinsey and Company. *FinTechnicolor : The New Picture in Finance*. McKinsey, 2016. URL : <https://books.google.fr/books?id=2T8evwEACAAJ>.
- [22] Soheil Almasi Monfared, David Enke. *Volatility Forecasting Using a Hybrid GJR-GARCH Neural Network Model*. Procedia Computer Science, 36 :246–253, 2014.

- 
- [23] J.P. Morgan. *Introduction to RiskMetrics*. New York : J.P. Morgan, 1994.
  - [24] Gang Mu, Teodor Godina, Antonio Maffia, Yong Chao Sun. *Supervised Machine Learning with Control Variates for American Option Pricing*. Foundations of Computing and Decision Sciences, 43(3) :207–217, 2018.
  - [25] Basel Committee on Banking Supervision. *Minimum Capital Requirements for Market Risk*. Final Basel III Modelling : Implementation, Impact and Implications, 2016.
  - [26] Joerg Orgeldinger. *Recent Issues in the Implementation of the New Basel Minimum Capital Requirements for Market Risk*. Emerging Science Journal, 2(2) :65–77, 2018.
  - [27] Stefan Pichler, Karl Selitsch. *A Comparison of Analytical VaR Methodologies for Portfolios that Include Options*, 2000.
  - [28] Pasqualina Porretta, Paolo Agnese. *The Fundamental Review of Trading Book : New Standard Approach and Risk Management Impacts*. Journal of Risk Management in Financial Institutions, 14(2) :209–219, 2021.
  - [29] Wolf-Georg Ringe, Christopher Ruof. *Robo Advice – Legal and Regulatory Challenges*. In Routledge Handbook of Financial Technology and Law, pages 193–212. Routledge, 2021.
  - [30] R. Tyrrell Rockafellar, Stanislav Uryasev. *Optimization of Conditional Value-at-Risk*. Journal of Risk, 2 :21–42, 2000.
  - [31] Rajiv Sambasivan, Sourish Das. *A Statistical Machine Learning Approach to Yield Curve Forecasting*. In 2017 International Conference on Computational Intelligence in Data Science (ICCIDS), pages 1–6. IEEE, 2017.
  - [32] Anthony Saunders. *Financial Institutions Management : A Modern Perspective*, 1997.
  - [33] Sascha Wilkens. *Machine Learning in Risk Measurement : Gaussian Process Regression for Value-at-Risk and Expected Shortfall*. Journal of Risk Management in Financial Institutions, 12(4) :374–383, 2019.
  - [34] Heng-Guo Zhang, Chi-Wei Su, Yan Song, Shuqi Qiu, Ran Xiao, Fei Su. *Calculating Value-at-Risk for High-Dimensional Time Series Using a Nonlinear Random Mapping Model*. Economic Modelling, 67 :355–367, 2017.