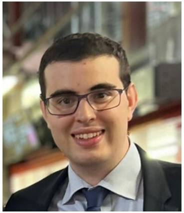
Pathwise Methods and Robust GANs for Pricing and Hedging

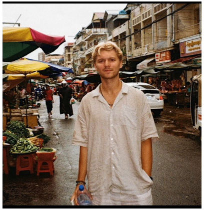
Blanka Horvath
University of Oxford and The Alan Turing Institute

Quantitative Finance Conference
In the Honour of Michael Dempster's 85th Birthday
13th April 2023, Cambridge

This talk is based and inspired by joint works with

Zacharia Issa, Andrew Alden, Yannick Limmer, Owen Futter...





Motivation for our Market Generators

(Architecture, ObjF; TrainData) $\Rightarrow$ **Program**

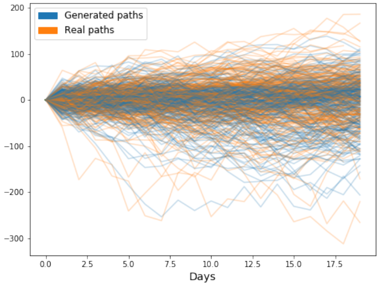
(Program, TestData) $\Rightarrow$ **Output**

Task "Market Generator":
Find (synthetic) **TrainData** for the network, such that performance is optimized when **TestData** $=$ Market Data.

Motivation for our Market Generators

Task "Market Generator": Find (synthetic) **TrainData** for the network, such that performance is optimized when **TestData** = Market Data.

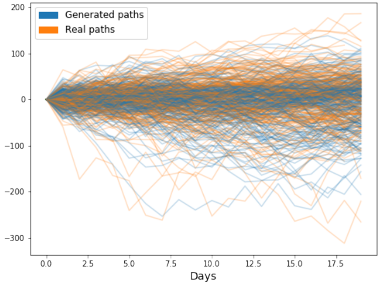
The challenge that market generators face is to produce new data samples that have the same distribution as test data they will later be exposed to.

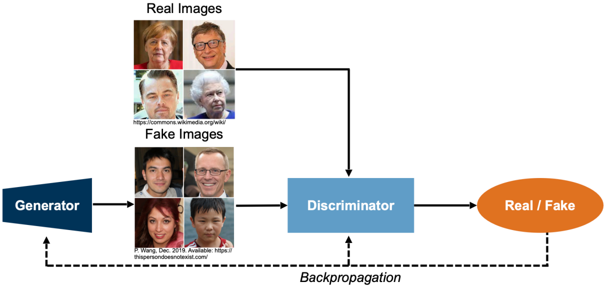


Motivation for our Market Generators

Task "Market Generator": Find (synthetic) **TrainData** for the network, such that performance is optimized when **TestData** = Market Data.

The challenge that market generators face is to produce **new data** samples that have the same distribution as test data they will later be exposed to.

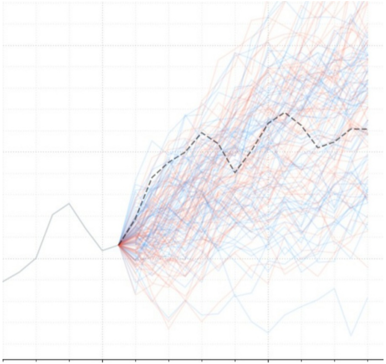




Motivation for our Market Generators

Task "Market Generator": Deliver (synthetic) **TrainData** for the network, such that performance is optimized when **TestData** = Market Data.

The quintessential difficulty that market generators face is to produce **genuinely new data** samples that in aggregation have the same distribution as the test data they will later be exposed to, though we will only see one realisation of the path. (By Zach Issa:)



Challenges with real (historical) data:

(1) **Data availability:** In many real situations, there is very limited data available for training/estimation). ⇒ small datasets may induce higher estimation errors.

(2) **Computational limitations:** Some limitations imposed on datasets (real & synthetic) by computational- and memory considerations (examples later).

(3) **Data changes over time:** Markets are heteroskedastic and non-stationary
⇒ may be a possible reason for (1) limitations in (training) data availablibIty
⇒ large changes in the data may require retraining / changing the network, **but...**

...small changes in data & estimation errors should not throw the application off track.

Needed: Appropriate (smooth) robustification of tasks towards small changes in input

Challenges with real (historical) data:

(1) **Data availability:** In many real situations, there is very limited data available for training/estimation). $\Rightarrow$ small datasets may induce higher estimation errors.

(2) **Computational limitations:** Some limitations imposed on datasets (real & synthetic) by computational- and memory considerations (examples later).

(3) **Data changes over time:** Markets are heteroskedastic and non-stationary
$\Rightarrow$ may be a possible reason for (1) limitations in (training) data availablibIty
$\Rightarrow$ large changes in the data may require retraining / changing the network, **but...**

...small changes in data & estimation errors should not throw the application off track.

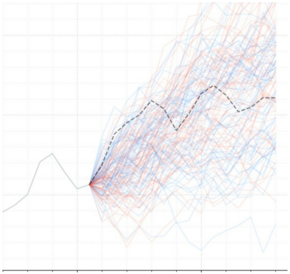
Needed: Appropriate (smooth) robustification of tasks towards small changes in input

Key tool: Quantify similarity (metric) of sets of paths (processes).

Task revisited

- **Now:** **(Architecture, ObjF; TrainData) $\Rightarrow$ Program**
  (Program, TestData) $\Rightarrow$ Output

Task "Market Generator": Find (synthetic) **TrainData** for the network, such that performance is optimized when **TestData** = Market Data. (Zach Issa, Sig-GAN:)



Key: Evaluating the "quality" (→ metrics) of training data for data streams.

- These include: Pathwise signature-based (MMD) metrics, and Wasserstein distances. Older approaches include: Matching quantiles and stylized facts

The signature mapping for data-driven modelling indisputably intriguing: In many cases, working over path space directly is preferable in a financial setting, as the language of financial mathematics revolves around path objects. The signature provides hope in solving some of the more complicated problems in mathematical finance, see [Perez-Arribas PhD Thesis '20]

- model-free pricing and hedging of path-dependent options,
- non-parametric feature extraction with applications to machine learning ....

However, it doesn't come without challenges. At the heart of many applications lies the MMD.

(Signature) Maximum Mean Discrepancy

Maximum mean discrepancy (MMD): μ and ν Borel probability measures on $\mathcal{X}$.

$$MMD_{\mathcal{G}}(\mu, \nu) := \sup_{f \in \mathcal{G}} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\mu(dx) \right|.$$

The $MMD_{\mathcal{G}}$ is a **metric** (point separating) if $\mathcal{G}$ is rich enough (e.g. a RKHS).

(Signature) Maximum Mean Discrepancy

Maximum mean discrepancy (MMD): μ and ν Borel probability measures on $\mathcal{X}$.

$$MMD_{\mathcal{G}}(\mu,\nu) := \sup_{f\in\mathcal{G}} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\mu(dx) \right|.$$

The $MMD_{\mathcal{G}}$ is a **metric** (point separating) if $\mathcal{G}$ is rich enough (e.g. a RKHS).

If $(\mathcal{H}, k)$ is a RKHS with kernel k and $\mathcal{G} := \{f \in \mathcal{H} : ||f||_{\mathcal{H}} \leq 1\}$ then

$$MMD^2_{\mathcal{G}}(\mu,\nu) = \mathbb{E}\left[k(X,X')\right] + \mathbb{E}\left[k(Y,Y')\right] - 2\mathbb{E}\left[k(X,Y)\right]$$

for $X, X' \sim \mu$ independent integrable r.v. and $Y, Y' \sim \nu$ independent integrable r.v.

(Signature) Maximum Mean Discrepancy

Maximum mean discrepancy (MMD): μ and ν Borel probability measures on $\mathcal{X}$.

$$MMD_{\mathcal{G}}(\mu,\nu) := \sup_{f\in\mathcal{G}} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\mu(dx) \right| .$$

The $MMD_{\mathcal{G}}$ is a **metric** (point separating) if $\mathcal{G}$ is rich enough (e.g. a RKHS).

If $(\mathcal{H}, k)$ is a RKHS with kernel k and $\mathcal{G} := \{f \in \mathcal{H} : ||f||_{\mathcal{H}} \leq 1\}$ then

$$MMD^2_{\mathcal{G}}(\mu,\nu) = \mathbb{E}\left[k(X,X')\right] + \mathbb{E}\left[k(Y,Y')\right] - 2\mathbb{E}\left[k(X,Y)\right]$$

for $X, X' \sim \mu$ independent integrable r.v. and $Y, Y' \sim \nu$ independent integrable r.v.

Characteristicness of the signature kernel implies an associated signature MMD, which can be used as a metric on path space [CO '18]. We choose $k(\cdot,\cdot)$ to be the (normalised) *signature* kernel and use $MMD_{\mathcal{G}}(\mu,\nu)$ as a hypothesis test:

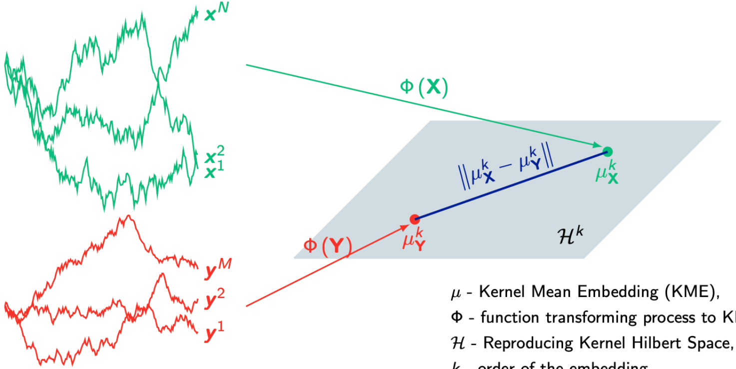
$$H_0 : \quad \nu = \mu$$
$$H_1 : \quad \nu \neq \mu$$

The Signature MMD as a Two-Sample Test

The aforementioned MMD has been one of the most-employed tools in this context. However, only very few results are available on understanding how the signature kernel MMD functions as a statistical tool.

The Signature MMD as a Two-Sample Test

The aforementioned MMD has been one of the most-employed tools in this context. However, only very few results are available on understanding how the signature kernel MMD functions as a statistical tool.



μ - Kernel Mean Embedding (KME),
Φ - function transforming process to KME,
$\mathcal{H}$ - Reproducing Kernel Hilbert Space,
k - order of the embedding.

Image: Andrew Alden

The Signature MMD as a Two-Sample Test

The aforementioned MMD has been one of the most-employed tools in this context. However, only very few results are available on understanding how the signature kernel MMD functions as a statistical tool.

For (independent) samples $X_1, \ldots, X_n \sim \mathbb{P}_X$ and $Y_1, \ldots, Y_n \sim \mathbb{P}_Y$ there is an unbiased estimator $MMD_n^2(X_1, \ldots, X_n; Y_1, \ldots, Y_n)$ and strongly consistent. It for $n \to \infty$ it converges to the (theoretical) MMD

$$MMD_n^2(X_1, \ldots, X_n; Y_1, \ldots, Y_n) \longrightarrow_{n\to\infty} MMD_{\mathcal{G}}^2(\mathbb{P}_X, \mathbb{P}_Y) \quad \text{a.s.}$$

The Signature MMD as a Two-Sample Test

Claim: In this setting, robustification comes (in some parts) naturally through small sample sizes of data:

- Each collection of paths $\{\boldsymbol{x}^k\}_{k=1}^r, \{\boldsymbol{y}^k\}_{k=1}^m$ results in a different distance value.
- Ambiguity in feature vectors.

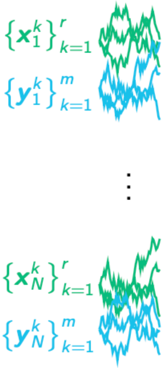


Image: Andrew Alden

The Signature MMD Two-Sample Test (Base Case, truncated signatures)

To asses whether a generative model generate "realistic" paths, sample real paths $Y_1, \ldots, Y_n$, for some $n \in \mathbb{N}$, and sample generated paths $X_1, \ldots, X_n$ and apply the two-sample test in [Chevyrev and Oberhauser '18]. Signature-based MMD test statistic $T(X_1, \ldots, X_n; Y_1, \ldots, Y_n)$ where $k(\cdot, \cdot)$ is the *signature kernel*:

$$T(X_1, \ldots, X_n; Y_1, \ldots, Y_n) := \frac{1}{n(n-1)} \sum_{i,j;i \neq j} k(X_i, X_j) - \frac{2}{n^2} \sum_{i,j} k(X_i, Y_j) + \frac{1}{n(n-1)} \sum_{i,j;i \neq j} k(Y_i, Y_j),$$

Then, given a confidence level $\alpha \in (0, 1)$, compute $c_\alpha(n) := 4\sqrt{-n^{-1} \log \alpha}$ (**threshold**). Generated paths are **realistic with a confidence** α if $T_n^2 < c_\alpha(n)$.
Note how threshold depends on the number n of samples considered.

The Signature MMD Two-Sample Test (Base Case, truncated signatures)

To asses whether a generative model generate "realistic" paths, sample real paths $Y_1, \ldots, Y_n$, for some $n \in \mathbb{N}$, and sample generated paths $X_1, \ldots, X_n$ and apply the two-sample test in [Chevyrev and Oberhauser '18]. Signature-based MMD test statistic $T(X_1, \ldots, X_n; Y_1, \ldots, Y_n)$ where $k(\cdot, \cdot)$ is the *signature kernel*:

$$T(X_1, \ldots, X_n; Y_1, \ldots, Y_n) := \frac{1}{n(n-1)} \sum_{i,j; i \neq j} k(X_i, X_j) - \frac{2}{n^2} \sum_{i,j} k(X_i, Y_j) + \frac{1}{n(n-1)} \sum_{i,j; i \neq j} k(Y_i, Y_j),$$

Then, given a confidence level $\alpha \in (0, 1)$, compute $c_\alpha(n) := 4\sqrt{-n^{-1} \log \alpha}$ (**threshold**). Generated paths are **realistic with a confidence** α if $T_n^2 < c_\alpha(n)$. Note how threshold depends on the number n of samples considered.

This base-case has been considered in "Signature-based validation of real-world economic scenarios" [Andres, Boumezoued, Jourdain '23] for $n \neq m$.

The Signature MMD as a Two-Sample Test

The aforementioned MMD has been one of the most-employed tools in this context. However, only very few results are available on understanding how the signature kernel MMD functions as a statistical tool.

Claim:

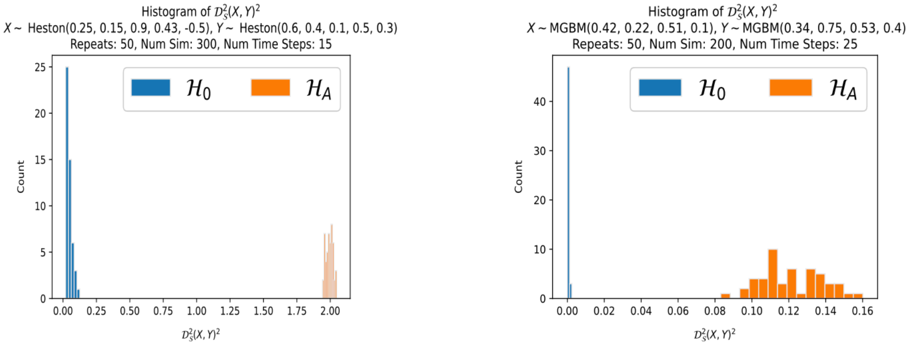
In this setting, robustification comes (in some parts) naturally through small sample sizes of data:

- To demonstrate this, we will first revisit the signature-kernel two sample test in [Chevyrev and Oberhauser '18] as it illustrates well some of our claims.
- **Though there were several further developments of related tests since then (e.g. the kernel trick proposed in [Cass, Foster, Lyons, Salvi, Yang '21] and it's higher rank version proposed in [H. Lemercier, Liu, Lyons, Salvi '22]) which we will later discuss from this aspect for MMD-tests using the kernel-trick (1st and 2nd order) as well.**

The Signature MMD as a Two-Sample Test

The aforementioned MMD has been one of the most-employed tools in this context. However, only few results are available on understanding the MMD as a statistical tool.

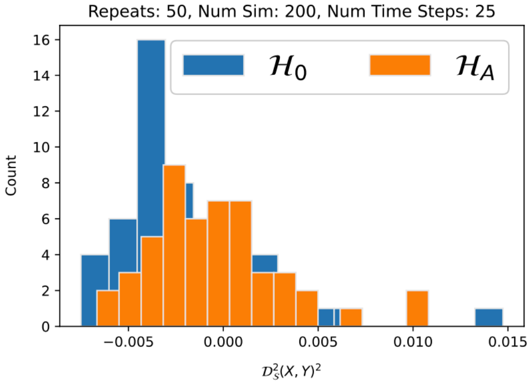
- Perform a two-sample test: $\mathcal{H}_0 \colon \mathbb{P}_{X|\mathcal{F}_X} = \mathbb{P}_{Y|\mathcal{F}_Y}$ and $\mathcal{H}_A \colon \mathbb{P}_{X|\mathcal{F}_X} \neq \mathbb{P}_{Y|\mathcal{F}_Y}$.



- **Empirical estimate** of 2^{nd}-order MMD has a long **run-time**.

The Signature MMD as a Two-Sample Test

Claim: In this setting, robustification comes (in some parts) naturally through small sample sizes of data:



Statistical power of test

"Signature-based validation of real-world economic scenarios" [A.B.J. '23]

- numerical analysis with synthetic data in order to measure the statistical power
- then use it on historical data to study the ability of the test to discriminate between models and demonstrate the potential of the MMD-based validation for real-world economic scenarios and applications requiring to exhibit the consistency of a stochastic model with historical paths
- consider an asymmetric setting $n \neq m$ in which a large (m) sample of simulated real-world scenarios is compared to a small (n) sample of real/target data
- by performing specific transformations of the signature, we can reach statistical powers close to 1 in this framework. However such transformations can have a flip-side and are to be handled with care...(see later)

We also see challenges arising related to the numerical implementation, and limitations in its domain of validity in terms of the distance between models and the volume of data at hand.

Path scalings and type II errors

1. In practice one works with empirical estimators $\Lambda_k(\overline{\mathbb{P}}, \overline{\mathbb{Q}})$ of the Sig-MMD $d(\mathbb{P}, \mathbb{Q})$ with $\overline{\mathbb{P}} = (X_1, \ldots, X_N), X_i \sim \mathbb{P}$ and $\overline{\mathbb{Q}} = (Y_1, \ldots, Y_M), Y_j \sim \mathbb{Q}$
2. The variance of these estimators $\mathrm{Var}(\Lambda_k(\overline{\mathbb{P}}, \overline{\mathbb{Q}})) \to 0$ as $N, M \to \infty$[1].
3. However for fixed sample size N, especially in a low data environment, type II error may occur. Lower order terms contribute more, meaning higher-order dynamics are discounted (or truncated)
4. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
5. Working with the signature kernel (**??**) via the kernel trick means that the original (expected) signature $\mathbb{E}_{\mathbb{P}}[S(X)]$ is no longer recoverable (cannot scale signature directly)
6. Solution: path scaling function $\sigma_\lambda(X) = \lambda X$ for some $\lambda \in \mathbb{R}^d$

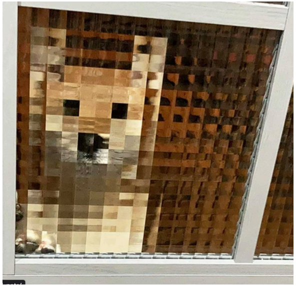
[1] We assume the population MMD is the sum of the (expected) level contributions.

Path scalings and type II errors

1. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
2. Can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
3. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different resolutions / zoom settings on a camera.

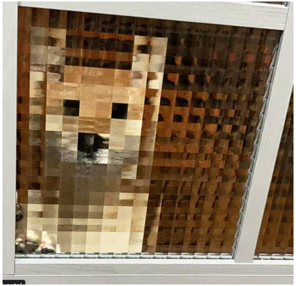
Path scalings and type II errors

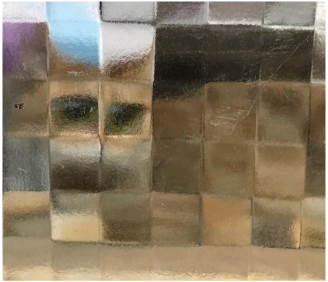
1. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
2. Can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
3. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different resolutions / zoom settings on a camera.



Path scalings and type II errors

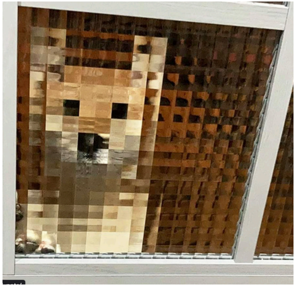
1. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
2. Can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
3. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different resolutions / zoom settings on a camera.

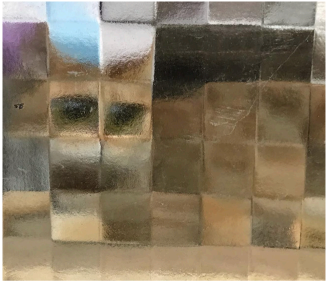


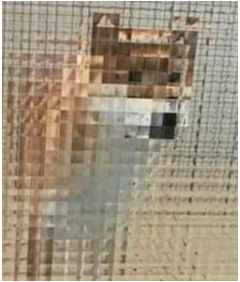


Path scalings and type II errors

1. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
2. Can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
3. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different resolutions / zoom settings on a camera.







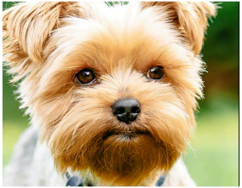
Path scalings and type II errors

1. Path scalings can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
2. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different resolutions / zoom settings on a camera.



Path scalings and type II errors

1. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
2. Can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
3. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different zoom settings on a camera.



Path scalings and type II errors

1. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
2. Can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
3. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different resolutions / zoom settings on a camera.

Answer: May need multiple scalings for use-cases.

Path scalings and type II errors

1. Since $\varphi(x) = x^k/k!$ is increasing in x, (larger) scaling has the effect of increasing the numerical size of higher-order signature terms.
2. Can help reduce the incidence of type II error for small batch sizes $\Rightarrow$ Rejection of the paradigm "higher order terms can be ignored due to factorial decay": numerical size $\neq$ information value.
3. **Issue:** what is the (most) appropriate scaling to tell apart two different stochastic proceses? **Analogy:** different resolutions / zoom settings on a camera.

Answer: May need multiple scalings for use-cases.

Pricing (path dependent) payoffs with the help of the MMD

Pricing (path dependent) payoffs with the help of the MMD

Typically (due to conditioning) 2^{nd}-order MMD needed.

Issue:

Empirical estimate of 2^{nd}-order MMD has a long run-time.
Computational limitations (batch size) due to memory considerations (e.g. 8GN memory, paths of length=64, dim=2, size=128)
$\Rightarrow$ smaller samples in estimates $\Rightarrow$ bigger variance.

Pricing (path dependent) payoffs with the help of the MMD

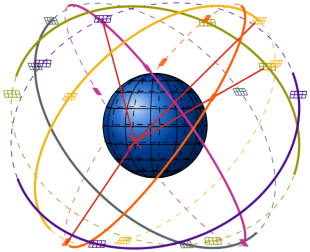
Typically (due to conditioning) 2^{nd}-order MMD needed.

Issue:
Empirical estimate of 2^{nd}-order MMD has a long run-time.
Computational limitations (batch size) due to memory considerations (e.g. 8GN memory, paths of length=64, dim=2, size=128)
$\Rightarrow$ smaller samples in estimates $\Rightarrow$ bigger variance.
Solution:
estimate simultaneously from multiple angles to sharpen perspective (estimate MMD to several reference points/models):

Pricing path dependent payoffs with the help of the MMD

To obtain the price of a derivative priced under the stochastic process $\mathbb{Y}$:

- Select N base processes as reference models (satellites).
- Compute the distance from $\mathbb{Y}$ to each base process (distance to satellites).
- Use these distances to price (GPS location).



Adapted from Walcott, K. (2012) *Three-dimensional graphics with PGF/TiKZ* [67] and Trzeciak, T. (2008) *Example: Stereographics and cylindrical map projections* (https://texample.net/tikz/examples/map-projections/) [64]

Image: Andrew Alden

Pricing path dependent payoffs with the help of the MMD

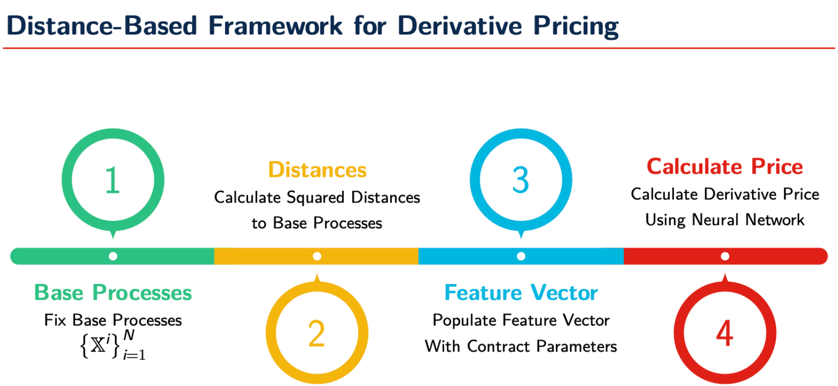


Image: Andrew Alden

Pricing path dependent payoffs with the help of the MMD

The method has been (successfully) applied to price path dependent options[2]:

- Down-and-In Barrier Options
- Best-Call Rainbow Options
- Autocallable Options

[2] in multidimensional classical (Heston, BS) and mixture model settings

Pricing path dependent payoffs with the help of the MMD

The method has been (successfully) applied to price path dependent options[2]:

- Down-and-In Barrier Options
- Best-Call Rainbow Options
- Autocallable Options

Derivative	MC Run-Time (s)	NN Run-Time (s)
Barrier	2.86	0.0165
Rainbow	0.325	0.0138
Autocallable	48.4	0.0231

Table: Run-Times.

Table: Andrew Alden

[2] in multidimensional classical (Heston, BS) and mixture model settings

Introducing robustness into Deep Hedging/ Deep Trading with the help of the MMD

Introducing robustness into DH/DT with the MMD

In the context hedging and trading strategies, the MMD can be helpful to introduce a smooth ambiguity-aversion effect into the hedging / trading objective.

Introducing robustness into DH/DT with the MMD

The original objective of a trading can be (most generally) given as

$$\max_{(\xi_t)_{t\in[0,T]}} \mathbb{E}\Big[U(V_T)\Big], \quad \text{where} \quad V_T = \sum_{m=1}^{d} \int_0^T \xi_t^m dX_t^m,$$

where V_T is the terminal profit and loss of our trading strategy.

Introducing robustness into DH/DT with the MMD

The original objective of a trading can be (most generally) given as

$$\max_{(\xi_t)_{t\in[0,T]}} \mathbb{E}\Big[U(V_T)\Big], \quad \text{where} \quad V_T = \sum_{m=1}^{d} \int_0^T \xi_t^m dX_t^m,$$

where V_T is the terminal profit and loss of our trading strategy. To introduce robustness in a feasible way, we need to extend the set of possible models (from $\mathbb{P}$), but restrict this to a permissible ambiguity set: Popular approaches to robustification include allowing alternative models within a δ-ball $B_\delta(\mathbb{P})$ around $\mathbb{P}$.

$$\max_{(\xi_t)_{t\in[0,T]}} \min_{\mathbb{Q}\in B_\delta(\mathbb{P})} \mathbb{E}^{\mathbb{Q}}\Big[U(V_T)\Big],$$

where–often–the viewpoint is taken that all alternative models $\mathbb{Q} \in B_\delta(\mathbb{P})$ are equally likely to materialize.

Introducing robustness into DH/DT with the MMD

The original objective of a trading can be (most generally) given as

$$\max_{(\xi_t)_{t\in[0,T]}} \mathbb{E}\Big[U(V_T)\Big], \quad \text{where} \quad V_T = \sum_{m=1}^{d} \int_0^T \xi_t^m dX_t^m,$$

where V_T is the terminal profit and loss of our trading strategy. To introduce robustness in a feasible way, we need to extend the set of possible models (from $\mathbb{P}$), but restrict this to a permissible ambiguity set.

Our approach to robustification: In the context of deriving (robust) optimal trading strategies, the MMD is helpful to introduce a smoother version of model ambiguity

$$\max_{(\xi_t)_{t\in[0,T]}} \min_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\Big\{U(V_T) + \tfrac{1}{\eta} d(\mathbb{P},\mathbb{Q})\Big\},$$

where $d(\mathbb{P},\mathbb{Q})$ denotes the distance of the alternative model $\mathbb{Q} \in \mathcal{Q}$ to our reference model $\mathbb{P}$ and $\frac{1}{\eta}$ is a scaling parameter representing the investor's aversion to model ambiguity.

Introducing robustness into DH/DT with the MMD

The original objective of a trading can be (most generally) given as

$$\max_{(\xi_t)_{t\in[0,T]}} \mathbb{E}\Big[U(V_T)\Big], \quad \text{where} \quad V_T = \sum_{m=1}^{d} \int_0^T \xi_t^m dX_t^m,$$

where V_T is the terminal profit and loss of our trading strategy. To introduce robustness in a feasible way, we need to extend the set of possible models (from $\mathbb{P}$), but restrict this to a permissible ambiguity set.

Our approach to robustification: In the context of deriving (robust) optimal trading strategies, the MMD is helpful to introduce a smoother version of model ambiguity

$$\max_{(\xi_t)_{t\in[0,T]}} \min_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\Big\{U(V_T) + \tfrac{1}{\eta} d(\mathbb{P},\mathbb{Q})\Big\},$$

where $d(\mathbb{P},\mathbb{Q})$ denotes the distance of the alternative model $\mathbb{Q}\in\mathcal{Q}$ to our reference model $\mathbb{P}$ and $\frac{1}{\eta}$ is a scaling parameter representing the investor's aversion to model ambiguity. Taking $d(\cdot,\cdot)$ as the Sig-MMD allows to take a fully pathwise perspective.

Introducing robustness into DH/DT with the MMD

The methodology works similarly in the (automated) derivation of robust hedging strategies. The original (discretized) setup is to find a strategy ϕ that maximizes

$$\mathbf{U}^{\mathbb{Q}}\left(\sum_{n=1}^{N}\phi_{t_{n-1}}^{\intercal}\Delta\mathbf{S}_{t_n}^{\mathbb{P}}-\mathbf{C}_T\right),$$

where $\mathbf{U}^{\mathbb{P}}\equiv\mathbb{E}^{\mathbb{P}}[U(\cdot)]$ for some utility function $U:\mathbb{R}\to\mathbb{R}$ (or in fact $\mathbf{R}[U(\cdot)]$, for some convex risk measure $\mathbf{R}$) qnd $\mathbf{C}_T$ denotes the payoff of the contingent claim.

Introducing robustness into DH/DT with the MMD

The methodology works similarly in the (automated) derivation of robust hedging strategies. The original (discretized) setup is to find a strategy ϕ that maximizes

$$\mathbf{U}^{\mathbb{Q}}\Big(\sum_{n=1}^{N}\phi_{t_{n-1}}^{\intercal}\Delta\mathbf{S}_{t_n}^{\mathbb{P}} - \mathbf{C}_T\Big),$$

where $\mathbf{U}^{\mathbb{P}} \equiv \mathbb{E}^{\mathbb{P}}[U(\cdot)]$ for some utility function $U : \mathbb{R} \to \mathbb{R}$ (or in fact $\mathbf{R}[U(\cdot)]$, for some convex risk measure $\mathbf{R}$) qnd $\mathbf{C}_T$ denotes the payoff of the contingent claim. We again consider its robustification as an optimization problem with smooth ambiguity aversion, where $\alpha : \mathcal{P} \to \mathbb{R}$ represents the belief that the agent has in any $\mathbb{P} \in \mathcal{P}$.

$$\inf_{\mathbb{Q}\in\mathcal{Q}}\mathbb{U}^{\mathbb{Q}}\Big(\sum_{n=1}^{N}\phi_{t_{n-1}}^{\intercal}\Delta\mathbf{S}_{t_n}^{\mathbb{Q}} - \mathbf{C}_T\Big) + \alpha(\mathbb{Q}). \tag{1}$$

The penalty can be represented as $\alpha(\mathbb{P}) := \gamma\, d(\mathbb{P},\mathbb{Q})$, where $d : \mathcal{P}^2 \to \mathbb{R}_+$ with $\mathbb{P}$ being the reference model and $\gamma \in \mathbb{R}_+$ the sensitivity (aversion) to model ambiguity. It can be chosen the Sig-MMD as a model agnostic distance on pathspace.

Robust Hedging GANs

- In classical optimisation, the standard optimal strategy is obtained when the agent maximises its (expected) utility based on the assumption that its model description is correct.
- However, if the agent is uncertain about which of the possible models is realised one approach (worst-case approach) is to assume that mother nature picks the worst-case model among a set of plausible models, whereafter the agent maximises the (expected) utility given this worst-case model.

Robust Hedging GANs

- In classical optimisation, the standard optimal strategy is obtained when the agent maximises its (expected) utility based on the assumption that its model description is correct.
- However, if the agent is uncertain about which of the possible models is realised one approach (worst-case approach) is to assume that mother nature picks the worst-case model among a set of plausible models, whereafter the agent maximises the (expected) utility given this worst-case model.
- **This lends itself well to a GAN (Generative Adversarial Network) setting:**

  Though the roles of Generator, Discriminator are slightly modified, compared to the standard GAN setting: **Generator = Mother nature** (adversarial) and **Discriminator = hedger** (robustified).

Structure

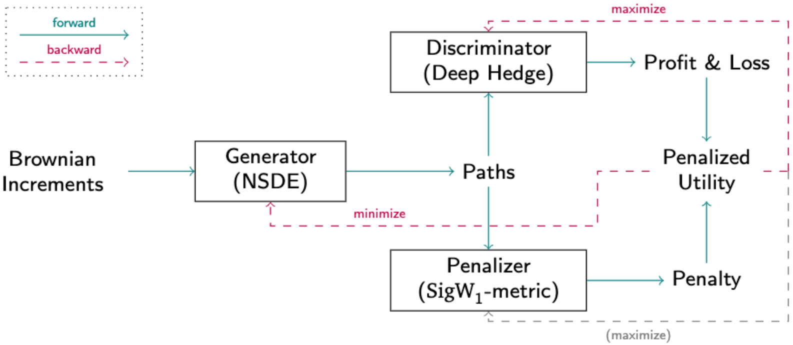


Figure: Structure of the Robust Hedging GAN with backpropagation of the components.

Image: Yannick Limmer

Experiments

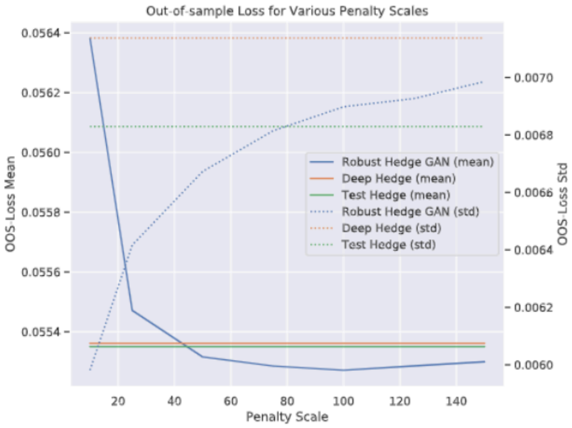


Image: Yannick Limmer

Our approach to introducing robustness into DH/DT with the MMD

Our approach to robustification: In the context of deriving (robust) trading/hedging strategies, the MMD is helpful to introduce a smoother version of model ambiguity

Our approach to introducing robustness into DH/DT with the MMD

Our approach to robustification: In the context of deriving (robust) trading/hedging strategies, the MMD is helpful to introduce a smoother version of model ambiguity

- Taking $d(\cdot,\cdot)$ as the Sig-MMD allows to take a fully pathwise perspective.
- Permits a fully model agnostic approach.
- This problem can quickly become a challenge for reasonable general underlying dynamics X. In special cases benchmark solutions and asymptotic results for generalisations exist.
- A combination of a pathwise perspective and an adversarial (GAN) approach is helpful in obtaining (candidte) optimizers, which were out of the reach of our traditional techniques.

Our approach to introducing robustness into DH/DT with the MMD

Our approach to robustification: In the context of deriving (robust) trading/hedging strategies, the MMD is helpful to introduce a smoother version of model ambiguity

- Taking $d(\cdot, \cdot)$ as the Sig-MMD allows to take a fully pathwise perspective.
- Permits a fully model agnostic approach.
- This problem can quickly become a challenge for reasonable general underlying dynamics X. In special cases benchmark solutions and asymptotic results for generalisations exist.
- A combination of a pathwise perspective and an adversarial (GAN) approach is helpful in obtaining (candidte) optimizers, which were out of the reach of our traditional techniques.

Thank you for your attention!

Experiments

Strategy	γ	Mean	Standard Deviation
Deep Hedge		0.055362	0.007137
robust hedging GAN	10	0.056383	0.005981
	25	0.055471	0.006417
	50	0.055316	0.006673
	75	0.055285	0.006815
	100	0.055271	0.006898
	125	0.055286	0.006927
	150	0.055300	0.006986
Hedge on Test Set		0.055350	0.006839

Now really thank you for your attention!

Mathematical Appendix and Background (Rough) Paths-wise Approach

Levin, Lyons, & Ni. (2013) proposed the signature of a path as a basis of functions for a functional on path space.

Definition (Signature of a path)

Let $X : [0, T] \to \mathbb{R}^d$ be a continuous path of bounded variation. The signature of X is then defined by the sequence of iterated integrals given by

$$\mathbb{X}_T^{<\infty} := (1, \mathbb{X}_t^1, \ldots, \mathbb{X}_T^n, \ldots), \quad \text{where}$$

$$\mathbb{X}_T^n := \int_{0<u_1<\ldots<u_k<T} dX_{u_1} \otimes \ldots \otimes dX_{u_k} \in (\mathbb{R}^d)^{\otimes n}$$

with $\otimes$ the tensor product. Similarly, given $N \in \mathbb{N}$, the truncated signature of order N is defined by

$$\mathbb{X}_T^{\leq N} := (1, \mathbb{X}_T^1, \ldots, \mathbb{X}_T^N).$$

The path X has b.v. (discrete data) $\Rightarrow$ the integrals can be defined i.s.o. Riemann-Stieltjes.

Mathematical Appendix and Background

We work with the log-signatures (**Liao, Lyons, Ni, Yang (2019)**)

Definition (Log-signature)

Let $X : [0, T] \to \mathbb{R}^d$ be a path such that its signature $\mathbb{X}^{<\infty}_{0,T}$ is well-defined. The log-signature is then defined by

$$\log \mathbb{X}^{<\infty}_{T} := -\mathbb{X}^{<\infty}_{T} + \frac{1}{2}(\mathbb{X}^{<\infty}_{T})^{\otimes 2} - \frac{1}{3}(\mathbb{X}^{<\infty}_{T})^{\otimes 3} + \ldots + (-1)^n \frac{1}{n}(\mathbb{X}^{<\infty}_{T})^{\otimes n} + \ldots,$$

which can be shown to be well-defined.

- There is a one-to-one map between signatures and log-signatures.
- Log-signatures have all positive properties listed above.
- They allow for lower dimensional representation and are better suited to VAE.