Multidimensional case

- Let $X := (X^1, \ldots, X^d)$, $\Xi := (\Xi^1, \ldots, \Xi^q) : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}^q$, *square integrable* random vectors.

$$\mathbb{E}\, X = m \in \mathbb{R}^d \setminus \{0\}, \qquad \mathbb{E}(\Xi) = 0 \in \mathbb{R}^q$$

- Let $D(X) := [\mathrm{Cov}(X^i, X^j)]_{1 \leqslant i,j \leqslant d}$ and $D(\Xi)$ denote the covariance (dispersion) matrices of X and Ξ respectively.

$$D(X) \;\text{ and }\; D(\Xi) > 0$$

i.e. positive definite symmetric.

- **Problem**: Find a matrix $\Lambda \in \mathcal{M}(d,q)$ solution of the optimization problem

$$\mathrm{Var}(X - \Lambda\,\Xi) = \min\left\{\mathrm{Var}(X - L\,\Xi),\; L \in \mathcal{M}(d,q)\right\}$$

where $\mathrm{Var}(X) := \mathbb{E}\left|X - \mathbb{E}\,X\right|^2 = \min_{a \in \mathbb{R}^d} \mathbb{E}\,|X - a|^2$.

- **Solution:**

$$\Lambda = \big[D(\Xi)\big]^{-1} C(X,\Xi)$$

where

$$C(X,\Xi) = [\mathrm{Cov}(X^i,\Xi^j)]_{1\leqslant i\leqslant d, 1\leqslant j\leqslant q}.$$

- **Examples:** Traded assets $S_t = (S_t^1,\dots,S_t^d)$, $t\in[0,T]$.

– Options on various baskets

$$X^i = \left(\sum_{j=1}^{d}\theta_j^i S_T^j - K\right)_+, \; i=1,\dots,d$$

(with usually $\sum_j \theta_j^i = 1$ for every i).
Remark. Also produces an asset selection which helps for hedging.

– Portfolio of *forward start options*

$$X^{i,j} = \left(S^j_{T_{i+1}} - S^j_{T_i}\right)^+, \; i=1,\dots,r,\; j=1,\dots,d$$

Antithetic (second round: direct approach)

- If $X \stackrel{d}{=} X' \in L^2(\mathbb{P})$, $\mathrm{Var}(X - X') > 0$, then $m_X = \mathbb{E}\,X = \mathbb{E}\,X'$.
  Optimal control variate by dynamic regression optimization reads:

$$\min_{\lambda \in \mathbb{R}} \mathrm{Var}\big(\lambda X + (1-\lambda)X'\big) = \mathrm{Var}\left(\frac{X + X'}{2}\right) = \frac{\mathrm{Var}(X) + \mathrm{Cov}(X, X')}{2}.$$

- Denote $\kappa_X = \mathrm{Complexity}(X)$. Assume now $\kappa_X \simeq \kappa_{X'}$ and

$$X'' = \frac{X + X'}{2} \quad \text{with} \quad \kappa_{X''} \simeq \kappa_X + \kappa_{X'} \simeq 2\kappa_X.$$

- (Back to) Practitioner's aim ?

Minimize the complexity to approximate m_X

subject to $\mathrm{Var}(\text{estimator}) = \varepsilon^2$.

- Compare the estimators $\overline{X}_M$ and $\overline{X}''_M$.

- For $\overline{X}_M$:

$$\kappa_{\overline{X}_M} = M\kappa_X \quad \text{and} \quad \mathrm{Var}(\overline{X}_M) = \frac{\mathrm{Var}(X)}{M}.$$

Let $M = M(\varepsilon) = \frac{\mathrm{Var}(X)}{\varepsilon^2}$ be such that $\mathrm{Var}(\overline{X}_M) = \varepsilon^2$. Then

$$\kappa_{\overline{X}_M} = \frac{\kappa_X \mathrm{Var}(X)}{\varepsilon^2}$$

- Likewise, for $\overline{X}''_M$:

$$\kappa_{\overline{X''}_M} = \frac{\kappa_{X''} \mathrm{Var}(X'')}{\varepsilon^2} = 2\kappa_X \cdot \frac{\mathrm{Var}(X) + \mathrm{Cov}(X, X')}{2} \frac{1}{\varepsilon^2}$$

- X'' should be selected iff $\kappa_{\overline{X''}_M} < \kappa_{\overline{X}_M}$ i.e.

$$\begin{aligned}
&\iff \quad \kappa_{X''} \mathrm{Var}(X'') < \kappa_X \mathrm{Var}(X) \\
&\iff \quad 2\,\kappa_X \frac{\mathrm{Var}(X) + \mathrm{Cov}(X, X')}{2} < \kappa_X \mathrm{Var}(X) \\
&\iff \quad \boxed{\mathrm{Cov}(X, X') < 0.}
\end{aligned}$$

- Hence the name antithetic for such couple of random variables.

Co-monotony

Proposition (co-monotony)

Let $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, let φ, $\psi : \mathbb{R} \to \mathbb{R}$ monotonic functions with the opposite (resp. the same) monotonicity. If $\varphi(Z)$, $\psi(Z) \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$. Then

$$\mathrm{Cov}\big(\varphi(Z), \psi(Z)\big) \leqslant 0 \quad \text{(resp. } \geqslant 0\text{)}.$$

Furthermore, equality holds if and only if $\varphi(Z) = \mathbb{E}\,\varphi(Z)$ $\mathbb{P}$-$a.s.$ or $\psi(Z) = \mathbb{E}\,\psi(Z)$ $\mathbb{P}$-$a.s.$

Applications. (a) Let $T : \mathbb{R} \to \mathbb{R}$ non-increasing s.t. $Z \stackrel{d}{=} T(Z)$. Let $X = \varphi(Z)$ and $X' = \varphi\big(T(Z)\big)$ with $\kappa_\varphi \gg \kappa_Z$.

$$X \stackrel{d}{=} X' \quad \text{and} \quad \mathrm{Cov}(X, X') \leqslant 0.$$

(b) Usual examples: $T(z) = -z$, $T(z) = L - z$. [Think to $\mathcal{N}(0,1)$ and $U([0,1])$.]

Importance sampling: abstract paradigm

- Let $(E, \mathcal{E}, \mu)$ be a σ-finite measured space, say $(\mathbb{R}^d, \lambda_d)$, $(\mathbb{N}, m_c)$, $(\mathcal{C}([0,T], \mathbb{R}^d), \mathbb{P}_W)$ (Wiener measure).
- Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ be an E-valued random variable. Assume its distribution

$$\mathbb{P}_X = f.\mu, \quad f : E \to \mathbb{R}_+, \text{ measurable.}$$

- How to speed up the computation by simulation of

$$\mathbb{E}\, h(X) = \int_E h(x)\mathbb{P}_X(dx) = \int_E h(x) f(x)\mu(dx) \quad \text{where } h \in L^1(\mathbb{P}_X)?$$

- Let g be μ-$a.s.$ positive probability density function on $(E, \mathcal{E})$.

$$\begin{aligned}\mathbb{E}\, h(X) &= \int_E h(x) f(x)\mu(dx) = \int_E \frac{h(x)f(x)}{g(x)} g(x)\mu(dx)\\ &= \mathbb{E}\left(\frac{h(Y)f(Y)}{g(Y)}\right) \quad \text{where} \quad Y \stackrel{d}{=} g.\mu.\end{aligned}$$

- In order to compute $\mathbb{E}\,h(X)$ one may implement an *MC* simulation based on Y (supposed to be simulable) *i.e.*

$$\mathbb{E}\,h(X) = \mathbb{E}\left(\frac{h(Y)f(Y)}{g(Y)}\right) = \lim_{M\to+\infty}\frac{1}{M}\sum_{k=1}^{M} h(Y_k)\frac{f(Y_k)}{g(Y_k)} \qquad a.s.$$

- Assume $\kappa_Y \simeq \kappa_X$ and $\kappa_{f/g} \ll \kappa_h$.
- Why or when doing that ? Clearly iff

$$\mathrm{Var}\left(\frac{h(Y)f(Y)}{g(Y)}\right) < \mathrm{Var}\big(h(X)\big).$$

or, equivalently

$$\mathbb{E}\left[\left(\frac{h(Y)f(Y)}{g(Y)}\right)^2\right] < \mathbb{E}\,h(X)^2.$$

- A priori. . . not very tractable. . .

- As

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{h(Y)f(Y)}{g(Y)}\right)^2\right] &= \int_E \left(\frac{h(x)f(x)}{g(x)}\right)^2 g(x)\,\mu(dx)\\
&= \int_E h(x)^2\frac{f(x)}{g(x)}f(x)\mu(dx)\\
&= \mathbb{E}\left[h(X)^2\frac{f}{g}(X)\right],
\end{aligned}$$

condition also reads,in terms of X,

$$\mathbb{E}\left[h(X)^2\frac{f}{g}(X)\right] < \mathbb{E}\,h(X)^2.$$

- Still not so tractable...
- But we keep the principle:

Change the r.v. to reduce the variance

Practitioner's viewpoint

- We want to price a deep out-of-the-money Call option written on a risky asset $(X_t)_{t\in[0,T]}$:

$$h_T = (X_T - K)^+ \quad \text{with } X_0 = x_0 \ll K$$

- X_T can be simulated but

most of the copies $h_T^{(m)}$of h_T will be ...0 !

- Huge loss of energy (complexity): though $\mathbb{E}\, h_T$ is small, it induces an explosion of the relative StdDev: if

$$Z \stackrel{d}{=} \mathcal{B}er\big(\{0,1\}, p\big),\ p \simeq 0,\ \mathbb{E}\, Z = p,\ \mathrm{Var}(Z) = p(1-p)$$

so that

$$\frac{\sqrt{\mathrm{Var}(Z)}}{\mathbb{E}\, Z} = \sqrt{\frac{1-p}{p}} \to +\infty \quad \text{as} \quad p \to 0.$$

- Importance sampling has been devised to defeat this rare event effect.

Toy-Example: mean translation on Gaussian r.v.

- Let $X_t = (X_t^1, \ldots, X_t^d)$, $t \in [0, T]$, be a d-dim Black-Scholes model.
- One has $e^{-rT} h_T = \varphi(Z)$, $Z \stackrel{d}{=} \mathcal{N}\big(0; I_d\big)$.

$$\text{Compute} \quad \mathbb{E}\,\varphi(Z) = \int_{\mathbb{R}^d} \varphi(z) e^{-\frac{|z|^2}{2}} \frac{dz}{(2\pi)^{\frac{d}{2}}}.$$

- Idea: mean translation: Let $\theta \in \mathbb{R}^d$. A standard change of variable

$$\begin{aligned}
\mathbb{E}\,\varphi(Z) &= \int_{\mathbb{R}^d} \varphi(z) e^{-\frac{|z|^2}{2}} \frac{dz}{(2\pi)^{\frac{d}{2}}}, \qquad z = u + \theta \\
&= \int_{\mathbb{R}^d} \varphi(u+\theta) e^{-\frac{|u+\theta|^2}{2}} \frac{du}{(2\pi)^{\frac{d}{2}}} \\
&= \int_{\mathbb{R}^d} \varphi(u+\theta) e^{-\frac{|\theta|^2}{2} - (\theta|u) - \frac{|u|^2}{2}} \frac{du}{(2\pi)^{\frac{d}{2}}} \\
&= e^{-\frac{|\theta|^2}{2}}\, \mathbb{E}\left(e^{-(\theta|Z)} \varphi(Z+\theta)\right) \\
\mathbb{E}\,\varphi(Z) &= e^{\frac{|\theta|^2}{2}}\, \mathbb{E}\Big(\varphi(Z+\theta) e^{-(\theta|Z+\theta)}\Big).
\end{aligned}$$

Connection with Importance Sampling

- Let $g_\theta(y) = \dfrac{e^{-\frac{|y-\theta|^2}{2}}}{(2\pi)^{\frac{d}{2}}}$, $y \in \mathbb{R}^d$, $\theta \in \mathbb{R}^d$ and $Z_\theta = Z + \theta$.

- Then

$$\mathbb{E}\,\varphi(Z) = \mathbb{E}\,\varphi(Z_0) = \mathbb{E}\left(\varphi(Z_\theta)\frac{g_0(Z_\theta)}{g_\theta(Z_\theta)}\right) = e^{\frac{|\theta|^2}{2}}\,\mathbb{E}\Big(\varphi(Z+\theta)e^{-(\theta|Z+\theta)}\Big)$$

since $\dfrac{g_0}{g_\theta}(z) = e^{-(\theta|z)+\frac{|\theta|^2}{2}}$.

- Note the there is no need to know 2π !

Choosing a good θ

- Starting from

$$\mathbb{E}\,\varphi(Z) = \mathbb{E}\left[e^{-\frac{|\theta|^2}{2}}\,e^{-(\theta|Z)}\varphi(Z+\theta)\right],$$

the question is: finding a good θ *i.e.* such that

$$V(\theta) := \mathbb{E}\left[e^{-\frac{|\theta|^2}{2}}\,e^{-(\theta|Z)}\varphi(Z+\theta)\right]^2 \ll \mathbb{E}\,\varphi(Z)^2.$$

- Heuristic approach (guided by the intuition on rare events):

Move Z where things should happen

  - Black-Scholes model 1D: $\mu = r - \frac{\sigma^2}{2}$ and

$$X_t^{x_0} = x_0 \exp\left(\mu T + \sigma W_T\right) = x_0 \exp\left(\mu T + \sigma\sqrt{T}Z\right), \stackrel{d}{=} \mathcal{N}(0;1).$$

  - (Deep) Out-of-the-money Call:

$$\varphi(z) = e^{-rT}\left(x_0 \exp\left(\mu T + \sigma\sqrt{T}z\right) - K\right)^+,\ z \in \mathbb{R}\ \text{ with }\ x \ll K.$$

- Idea 1: Move $Z = Z_0$ to $Z_\theta = Z + \theta$ such that

$$\mathbb{E}\left[x_0 \exp\left(\mu T + \sigma\sqrt{T} Z_\theta\right)\right] \simeq K$$

i.e.

$$\theta_1 = -\frac{\log(x_0/K) + rT}{\sigma\sqrt{T}}.$$

- Idea 2: Move $Z = Z_0$ to $Z_\theta = Z + \theta$ such that

$$\mathbb{P}\Big(x_0 \exp\left(\mu T + \sigma\sqrt{T}(Z+\theta)\right) < K\Big) = \frac{1}{2}$$

which yields

$$\theta_2 := -\frac{\log(x_0/K) + \mu\, T}{\sigma\sqrt{T}}.$$

Numerical experiments I

- $r = 0$, $\sigma = 0.2$, $x_0 = 70$, $K = 100$. $\mathrm{Call}_{BS} \simeq 0.248$
- Crude MC ($\theta = 0$):

$$\frac{\mathrm{StdDev}}{\mathrm{Call}_{BS}} \simeq \sqrt{\frac{3.9311}{0.248^2} - 1} \approx 7.9320$$

- Idea 1: $\theta_1 \approx 1.7834$

$$\frac{\mathrm{StdDev}}{\mathrm{Call}_{BS}} \simeq \sqrt{\frac{0.1517}{0.248^2} - 1} \approx 1.2110$$

- Idea 2: $\theta_2 \approx 1.8834$

$$\frac{\mathrm{StdDev}}{\mathrm{Call}_{BS}} \simeq \sqrt{\frac{0.1410}{0.248^2} - 1} \approx 1.1369$$

Numerical experiments II

- Optimization in θ?

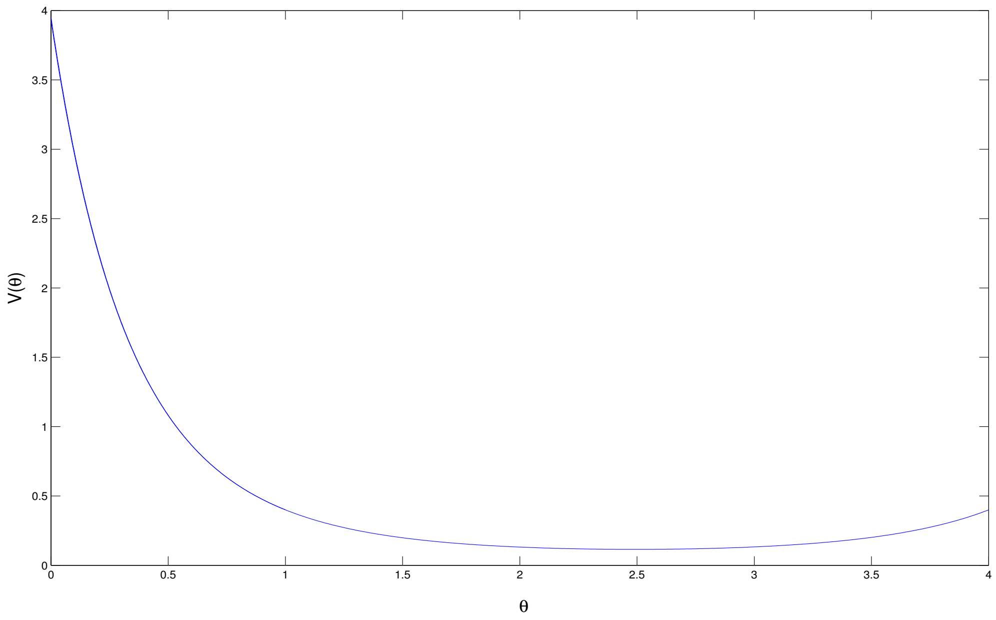

Figure: Out-of-the-money B-S Call: $\theta \mapsto \mathbb{E}\left[e^{-\frac{|\theta|^2}{2}} e^{-(\theta|Z)}\varphi(Z+\theta)\right]^2$

- $\theta_{\text{min}} = 2.5$

$$\frac{Std}{\mathrm{Call}_{BS}} \simeq \sqrt{\frac{0.1149}{0.248^2} - 1} \approx 0.9318$$

One step beyond ?

- Heuristics is not so bad... in 1-dimension.
- In higher dimensions (basket, index, etc), intuition is no longer clear.
- Try to take advantage of the optimization problem
  - Glasserman et al. (1999): Laplace's method formula + Large deviation techniques lead to a deterministic optimization problem. But needs regularity on φ.
  - Pre-processing optimization ? Solving

$$\min_{\theta\in\mathbb{R}^d}\left[V(\theta) := \mathbb{E}\left(e^{-|\theta|^2}\,e^{-2(\theta|Z)}\varphi(Z+\theta)^2\right)\right].$$

  - Numerical experiments suggest that V is convex. In fact, the reverse change of variable $z' = z-\theta$ yields

$$V(\theta) = e^{\frac{|\theta|^2}{2}}\mathbb{E}\left(\varphi^2(Z)e^{-(Z|\theta)}\right).$$

- As soon as $\varphi \not\equiv 0$, the formula

$$V(\theta) = e^{\frac{|\theta|^2}{2}} \mathbb{E}\left(\varphi^2(Z) e^{-(\theta|Z)}\right).$$

implies

  - V is strictly convex and $\lim\limits_{|\theta|\to+\infty} V(\theta) = +\infty$ by Fatou's Lemma.
  - Hence
$$\mathrm{argmin}_{\theta\in\mathbb{R}^d} V(\theta) = \{\theta_*\}.$$
  - θ is no longer inside $\varphi\ldots$

- V is differentiable and

$$\nabla V(\theta) = e^{\frac{1}{2}|\theta|^2}\mathbb{E}\left((\theta - Z)\varphi(Z)^2 e^{-(\theta|Z)}\right)$$

- Then $\{\nabla V = 0\} = \{\theta_*\}$.
- Iterative zero search ? Unfortunately no growth control, because of $e^{-(\theta|Z)}$ inside the $\mathbb{E}$.
- Third change of variable: $z = u - \theta$ yields

$$\nabla V(\theta) = e^{|\theta|^2}\mathbb{E}\left(\varphi^2(Z-\theta)(2\,\theta - Z)\right).$$

- $\nabla V(\theta) = 0$ iff $\mathbb{E}\left[\varphi^2(Z-\theta)(2\,\theta - Z)\right] = 0$.

A gradient descent?

- Self-growth control:

$$\nabla V(\theta) = 0 \quad \text{iff} \quad g(\theta) := \mathbb{E}\left[\frac{\varphi^2(Z-\theta)}{1+\varphi(-\theta)^2}(2\,\theta - Z)\right] = 0.$$

- Gradient descent (sort of)?
  - g has linear growth (…under natural assumptions on φ).
  - V is strictly convex implies

$$(g(\theta)\,|\,\theta-\theta_*) = (1+\varphi(-\theta)^2)e^{-\frac{|\theta|^2}{2}}(\nabla V(\theta)|\theta-\theta_*) > 0 \text{ if } \theta \neq \theta_*.$$

  Then

$$\theta_{n+1} = \theta_n - \gamma_{n+1}\, g(\theta_n) \to \theta_*.$$

  with a step sequence $(\gamma_n)_{n\geqslant 1}$ satisfying $\sum_n \gamma_n = +\infty$ and $\sum_n \gamma_n^2 < +\infty$.

- Unfortunately…Not implementable !!

A stochastic gradient descent?

- Get rid of the expectation: .
- Simulate i.i.d. copies $(Z_n)_{n\geqslant 1}$ of $\mathcal{N}(0; I_d)$ (by Box-Muller's or Marsaglia's polar method).
- Implement: (Lemaire-P., AAP, 2010). Initialize $\theta_0 \in \mathbb{R}^d$ and

$$\theta_{n+1} = \theta_n - \gamma_{n+1} \frac{\varphi^2(Z - \theta_n)}{1 + \varphi(-\theta_n)^2}(2\,\theta_n - Z_{n+1})$$

  still with $\sum_n \gamma_n = +\infty$ and $\sum_n \gamma_n^2 < +\infty$.
- Finally

$$\theta_n \to \theta_* \quad \text{as} \quad n \to +\infty.$$

  (with a Central Limit Theorem at rate $\sqrt{\gamma_n}$ under appropriate conditions so that $\gamma_n = \frac{c}{n}$ seems a good family of steps).
- But the story is not so simple as for the rate...

Batch versus adaptive...

Keep in mind that

$$\mathbb{E}\,\varphi(Z) = \mathbb{E}\left[e^{-\frac{|\theta|^2}{2}}\,e^{-(\theta|Z)}\varphi(Z+\theta)\right].$$

Like for dynamic control variate by regression, two possible approaches:

- Batch estimator: pre-processing of "size" M yields $\theta = \theta_{M'}$, followed by crude MC of size M, $M' \ll M$.
- Adaptive estimator (Arouna-Lapeyre):

$$\frac{1}{M}\sum_{k=1}^{M} e^{-\frac{|\theta_{k-1}|^2}{2}}\,e^{-(\theta_{k-1}|Z_k)}\varphi(Z_k+\theta_{k-1})$$

(Note the shift of -1 between Z_k and θ_{k-1}).

Stratified sampling

- Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$.
- Let $(A_i)_{i\in I}$ be a finite $\mathcal{E}$-measurable partition of the state space E (the *strata*)
- Assume that the weights of the strata

$$p_i = \mathbb{P}(X \in A_i) > 0,\ i \in I,\ \text{are known}$$

- Assume that

$$\mathcal{L}(X \,|\, X \in A_i) \stackrel{d}{=} \varphi_i(U),\ U \stackrel{d}{=} [0,1]^{r_i},\ r_i \in \mathbb{N} \cup \{\infty\}, i \in I.$$

where φ_i is explicit. (Note that $r_i = r \in \mathbb{N}^*$ in what follows).

- Let $F : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}or(\mathbb{R}))$ such that $\mathbb{E}\, F^2(X) < +\infty$
- Idea 1:Then, by elementary conditioning,

$$\begin{aligned} \mathbb{E}\, F(X) &= \sum_{i\in I} \mathbb{E}\Big(\mathbf{1}_{\{X\in A_i\}} F(X)\Big) = \sum_{i\in I} p_i\, \mathbb{E}\big(F(X)\,|\, X \in A_i\big) \\ &= \sum_{i\in I} p_i\, \mathbb{E}\big(F(\varphi_i(U_i))\big). \end{aligned}$$

- Idea 2: Since all $F(\varphi_i(U_i))$ have not the same variance, why not adapting the size of the devoted MC simulation on each stratum?
- Let M be the global budget of the simulation. Set $M_i = q_i M$, $i \in I$.
- Define the following (unbiased) estimator

$$\widehat{F(X)}^{I}_{M} := \sum_{i\in I} p_i \underbrace{\frac{1}{M_i}\sum_{k=1}^{M_i} F(\varphi_i(U_i^k)),}_{\text{stratum } i} \quad (U_i^k)_{i,k} \text{ i.i.d.}$$

- Then $$\mathrm{Var}\left(\widehat{F(X)}^{I}_{M}\right) = \frac{1}{M}\sum_{i\in I}\frac{p_i^2}{q_i}\sigma^2_{F,i}.$$

- where, for every $i \in I$,

$$\sigma^2_{F,i} \;=\; \mathrm{Var}\Big(F(\varphi_i(U))\Big) = \mathrm{Var}\big(F(X)\,|\,X \in A_i\big).$$

- Optimizing the sample path allocation to each stratum amounts to solving

$$\min_{(q_i)\in\mathcal{S}_I} \sum_{i\in I} \frac{p_i^2}{q_i}\sigma^2_{F,i}$$

where $\mathcal{S}_I := \Big\{(q_i)_{i\in I} \in \mathbb{R}_+^I \,:\, \sum_{i\in I} q_i = 1\Big\}$.

A sub-optimal choice: $q_i = p_i,\ i \in I$

- Why ? the weights p_i are known ! And

$$\begin{aligned}\sum_{i\in I}\frac{p_i^2}{q_i}\sigma^2_{F,i} &= \sum_{i\in I} p_i\,\sigma^2_{F,i}\\ &= \sum_{i\in I}\mathbb{E}\Big(\mathbf{1}_{\{X\in A_i\}}\big(F(X)-\mathbb{E}(F(X)|X\in A_i)\big)^2\Big)\\ \sum_{i\in I}\frac{p_i^2}{q_i}\sigma^2_{F,i} &= \big\|F(X)-\mathbb{E}\left(F(X)\middle|\,\mathcal{A}_I\right)\big\|_2^2,\end{aligned}$$

$\mathcal{A}_I = \sigma(\{X\in A_i\}, i\in I))$ is the σ-field spanned by $\{X\in A_i\}, i\in I$.

- Consequently, the constant $\mathbb{E}(F(X))$ being $\mathcal{A}_I$-measurable,

$$\sum_{i\in I}\frac{p_i^2}{q_i}\sigma^2_{F,i} < \big\|F(X)-\mathbb{E}(F(X))\big\|_2^2 = \mathrm{Var}(F(X))$$

with equality if and only if $\mathbb{E}\big(F(X)|\mathcal{A}_I\big) \equiv \mathbb{E}\,F(X)$ $\mathbb{P}$-$a.s.$

Optimal choice.

- Simply solve the above minimization problem !!
- We start from

$$\Big(\sum_{i\in I} p_i\sigma_{F,i}\Big)^2 = \Big(\sum_{i\in I}\frac{p_i\sigma_{F,i}}{\sqrt{q_i}}\sqrt{q_i}\Big)^2 \overset{\text{Schwarz Ineq.}}{\leqslant} \Big(\sum_{i\in I}\frac{p_i^2\sigma_{F,i}^2}{q_i}\Big)\underbrace{\Big(\sum_{i\in I} q_i\Big)}_{=1}$$

$$= \sum_{i\in I}\frac{p_i^2\sigma_{F,i}^2}{q_i}.$$

- With equality iff

$$q^*_{F,i} = \frac{p_i\sigma_{F,i}}{\sum_j p_j\sigma_{F,j}},\quad i\in I,$$

and a resulting minimal variance $\Big(\sum_{i\in I} p_i\sigma_{F,i}\Big)^2.$

- Note by the way

$$\begin{aligned}\sigma_{F,i} &= \Big[\mathbb{E}\Big(|X-\mathbb{E}(X|\{X\in A_i\})|^2 \,\big|\, \{X\in A_i\}\Big)\Big]^{\frac12}\\ &\geqslant \mathbb{E}\Big(\big|X-\mathbb{E}(X|\{X\in A_i\})\big|\,\big|\,\{X\in A_i\}\Big)\\ &= \frac{\mathbb{E}\Big(\big|F(X)-\mathbb{E}(F(X)|X\in A_i)\big|\mathbf{1}_{\{X\in A_i\}}\Big)}{p_i}\end{aligned}$$

so that, owing to Minkowski's Inequality,

$$\Big(\sum_{i\in I} p_i\sigma_{F,i}\Big)^2 \geqslant \big\|F(X)-\mathbb{E}\big(F(X)\,|\,\sigma(\{X\in A_i\}, i\in I)\big)\big\|_1^2.$$

- ...which show the magnitude of the possible gain, compared to the choice $q_i = p_i$;
- The gain lies in between

$$\big\|F(X)-\mathbb{E}\big(F(X)|\mathcal{A}_I\big)\big\|_1^2 \text{ and } \big\|F(X)-\mathbb{E}\big(F(X)\,|\,\mathcal{A}_I\big)\big\|_2^2.$$

- Unfortunately, the local inertia $\sigma^2_{F,i}$ are unknown...

Example: Computation of $\mathbb{E}\, F(X)$, $X \stackrel{d}{=} \mathcal{N}(0; I_d)$

- Strata: $A_i := \big\{x \in \mathbb{R}^d \text{ s.t. } x_1 = (e_1|x) \in [y_{i-1}, y_i]\big\},\ i = 1, \ldots, N$ with

$$p_i = \mathbb{P}(X \in A_i) = \mathbb{P}(Z \in [y_{i-1}, y_i]) = \Phi_0(y_i) - \Phi_0(y_{i-1}) = \frac{1}{N}.$$

- Then

$$\mathcal{L}\big(X \,|\, X_1 \in [a, b]\big) = \mathcal{L}(Z \,|\, Z \in [a, b]) \otimes \mathcal{N}(0, I_{d-1})$$

where the intra-strata simulation can be performed by inverting the conditional c.d.f.:

$$\mathcal{L}(Z \,|\, Z \in [a, b]) \stackrel{d}{=} \Phi_0^{-1}\Big((\Phi_0(b) - \Phi_0(a))U + \Phi_0(a)\Big),\ U \stackrel{d}{=} \mathcal{U}([0, 1]).$$

- Requires an accurate tabulation of function Φ_0.
- One can refine by considering hyper-rectangles...