Chapter 3
Variance reduction

3.1 Control Variate

Suppose we want to compute $m_X = \mathbb{E}(X)$

$$m_X = \lim_{M\to\infty} \overline{X}_M \qquad \left(\overline{X}_M = \frac{1}{M}\sum_{k=1}^{M} X_k \right)$$

CLT: $\sqrt{M}\left(\overline{X}_M - m_X\right) \xrightarrow[M\to\infty]{\text{law}} \mathcal{N}(0, \mathrm{Var}(X))$

So for all $q > 0$ (and M large enough)

$$\mathbb{P}\left(m_X \in \left[\overline{X}_M - q\sqrt{\frac{\overline{V}_M}{M}} \,;\, \overline{X}_M + q\sqrt{\frac{\overline{V}_M}{M}} \right] \right) \simeq 2\,\Phi_0(q) - 1$$

(Φ_0 is the cumulative distribution function of $\mathcal{N}(0,1)$)

Imagine we have a target / prescribed accuracy $\varepsilon > 0$: we want $\overline{X}_M$ to be in a confidence interval $[m_X - \varepsilon, m_X + \varepsilon]$ with a confidence level $\alpha = 2\,\Phi_0(q_\alpha) - 1$ (if we fix α, this prescribes the value of q_α). We need to perform a M-C simulation of size: $M \geq M^*(\varepsilon, \alpha) = q_\alpha^2 \dfrac{\mathrm{Var}(X)}{\varepsilon^2}$

(in practice, we replace $\mathrm{Var}(X)$ by $\overline{V}_M$).

This shows that, a confidence level being fixed, the size of a M-C simulation grows:

- linearly with the variance of X,
- quadratically as the inverse of the target accuracy.

3.1.1. Control Variate

Imagine we have a random variable $\Xi \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ such that:

i) $\mathbb{E}(\Xi)$ is known (or can be computed at a low cost)

ii) the r.v. $X - \Xi$ can be simulated (with same cost as X)

iii) $Var(X - \Xi) < Var(X)$

Then Ξ is called a control variate for X.

Example: using parity equations to produce control variates

$$\begin{cases} S_t = \text{risky asset} \\ S_t^0 = e^{rt} \text{, the riskless asset} \quad (S_0^0 = s_0) \end{cases}$$

We work under the risk-neutral probability:
$(e^{-rt} S_t)_{t \in [0,T]}$ is a martingale on $(\Omega, \mathcal{A}, \mathbb{P})$

Vanilla Call-Put parity ($d=1$)

$$\begin{cases} Call_0(K,T) = e^{-rT} \mathbb{E}((S_T - K)_+) \\ Put_0(K,T) = e^{-rT} \mathbb{E}((K - S_T)_+) \end{cases}$$

$(S_T - K)_+ - (K - S_T)_+ = S_T - K$

Let us set: $X := e^{-rT}(S_T - K)_+$; $X' = e^{-rT}(K - S_T)_+ + s_0 - e^{-rT}K$

We have: $Call_0(K,T) - Put_0(K,T) = s_0 - e^{-rT}K$

We have: $Call_0(K,T) = \mathbb{E}(X) = \mathbb{E}(X')$ so we set

$\Xi = X - X' = e^{-rT}K - s_0$ (so $\mathbb{E}(\Xi) = 0$)

We compute: $\overline{V}_X, \overline{V}_{X'}$. Imagine $\overline{V}_{X'} \leq \overline{V}_X$.

Then we estimate $\mathbb{E}(X - \Xi) = \mathbb{E}(X')$ by $\overline{X}'_n$

and estimate $\mathbb{E}(X)$ by $\overline{X}'_n + \mathbb{E}(\Xi)$

3.2. Rao-Blackwell method (in a simple case)

Assume that $X = g(Z_1, Z_2)$, $g \in L^2(\mathbb{R}^2, \mathcal{B}or(\mathbb{R}^2), \mathbb{P}_{Z_1,Z_2})$

with Z_1, Z_2 two independant vectors. Set $\mathcal{B} = \sigma(Z_2)$. We have:

$$\begin{aligned}
\operatorname{Var}(\mathbb{E}(X|\mathcal{B})) &= \mathbb{E}(\mathbb{E}(X|\mathcal{B})^2) - \mathbb{E}(\mathbb{E}(X|\mathcal{B}))^2 \\
&= \mathbb{E}(\mathbb{E}(X|\mathcal{B})^2) - \mathbb{E}(X)^2 \\
\text{(convexity)} \quad &\leq \mathbb{E}(\mathbb{E}(X^2|\mathcal{B})) - \mathbb{E}(X)^2 \\
&= \operatorname{Var}(X)
\end{aligned}$$

We have $\mathbb{E}(X) = \mathbb{E}(\underline{G(Z_2)})$

where $G(z_2) = \mathbb{E}(g(Z_1, Z_2) \mid Z_2 = z_2)$

We suppose that
- we have a closed-form expression for G
- Z_2 can be simulated with the same complexity as X

then the M-C method associated to $\mathbb{E}(G(Z_2))$ has a lower variance than the M-C method associated to $\mathbb{E}(X)$.

Example: Exchange spread options.

$X^i_T = x_i \exp\left(\left(r - \frac{\sigma_i^2}{2}\right)T + \sigma_i W^i_T\right)$ $x_i, \sigma_i > 0$ $(i = 1, 2)$

two Browian motions: W^i_T $(i = 1, 2)$ with correlation $\rho > 0$

Exchange spread option with payoff: $h_T = (X^1_T - X^2_T - K)_+$

$(W^1_T, W^2_T) = \sqrt{T}\,(\sqrt{1-\rho^2}\,Z_1 + \rho Z_2, Z_2)$ avec $(Z_1, Z_2) \sim \mathcal{N}(0, I_2)$

$$\begin{aligned}
e^{-r\cancel{T}}\,\mathbb{E}(h_T | Z_2) &= e^{-\cancel{rT}}\,\mathbb{E}((X^1_T - X^2_T - K)_+ \mid Z_2) \\
&= \operatorname{Call}_{BS}\Big(\underbrace{x_1 e^{-\frac{\rho^2\sigma_1^2 T}{2} + \sigma_1\rho\sqrt{T}Z_2}}_{\text{init. price}}, \underbrace{x_2 e^{(r - \frac{\sigma_2^2}{2})T + \sigma_2\sqrt{T}Z_2} + K}_{\text{strike}}, \underbrace{r}_{\text{interest rate}}, \underbrace{\sigma_1\sqrt{1-\rho^2}}_{\text{vol.}}, \underbrace{T}_{\text{maturity}}\Big)
\end{aligned}$$

So we have an explicit formula for $\mathbb{E}(h_T | Z_2)$.

We have: $\mathrm{Premium}_{SS}(x_1, x_2, k, \sigma_1, \sigma_2, r, T)$

$$\| \qquad \mathbb{E}(\mathbb{E}(e^{-rT} h_T | Z_2))$$

3.3 Stratified sampling

We want to compute an integral of the form $I = \mathbb{E}(g(X))$. The r.v. X has a density f (say, with respect to the Lebesgue measure on $\mathbb{R}^d$). We set: $\sigma = \mathrm{Var}(g(X))$

Suppose D is partitioned into $D_1, D_2, \dots, D_m$. We can re-write: $I = \sum_{i=1}^{m} p_i I_i$ where $\begin{cases} p_i = \mathbb{P}(X \in D_i) \\ I_i = \mathbb{E}(g(X) | X \in D_i) \end{cases}$

Let M be the global "budget" allocated to the computation of I. We split this budget into m groups by setting:

$M_i = q_i M$

to be the allocated budget to compute $\mathbb{E}(g(X) | X \in D_i)$.

This leads us to define the following (unbiased) estimator:

$$\widehat{g(X)}_M := \sum_{i=1}^{m} p_i \times \frac{1}{M_i} \sum_{k=1}^{M_i} \mathbb{E}(g(X_k^{(i)}))$$

$\hookrightarrow$ iid of law $\mathcal{L}(X | X \in D_i)$

Then, elementary computation shows that

$$\mathrm{Var}(\widehat{g(X)}_M) = \frac{1}{M} \sum_{i=1}^{m} \frac{p_i^2}{q_i} \sigma_i^2$$ where:

$$\sigma_i^2 = \mathrm{Var}(g(X) | X \in D_i)$$

We want to solve the following optimization problem:

$$\min_{(q_i) \in \mathcal{J}} \sum_{i=1}^{m} \frac{p_i^2}{q_i} \sigma_i^2 \quad \text{where } \mathcal{J} = \left\{ (q_i)_{1 \le i \le m} \in (0,1)^m \text{ s.t. } \sum_{i=1}^{m} q_i = 1 \right\}$$

We have:

$$\sum_{i=1}^{m} p_i \sigma_i = \sum_{i=1}^{m} \frac{p_i \sigma_i}{\sqrt{q_i}} \sqrt{q_i} \underset{\substack{\text{Cauchy} \\ -\text{Schwarz}}}{\overset{*}{\leq}} \left(\sum_{i=1}^{m} \frac{p_i^2 \sigma_i^2}{q_i}\right)^{\frac{1}{2}} \left(\sum_{i=1}^{m} q_i\right)^{\frac{1}{2}}$$

$$= \left(\sum_{i=1}^{m} \frac{p_i^2 \sigma_i^2}{q_i}\right)^{\frac{1}{2}} \times 1$$

And we know we have equality in $*$ if and only if $\exists \lambda$ s.t. $\forall i$: $\lambda \frac{p_i \sigma_i}{\sqrt{q_i}} = \sqrt{q_i}$

i.e.: $\lambda p_i \sigma_i = q_i$

as $\sum_{i=1}^{m} q_i = 1$, this leads to: equality $\iff$ $q_i = \underbrace{\frac{p_i \sigma_i}{\sum_{j=1}^{m} p_j \sigma_j}}_{=: q_i^*}$

Remark: The σ_i may not be known explicitely. This makes the implementation less straightforward.

With $q_i = q_i^*$, we get the minimal variance:

$$\sum_{i=1}^{m} \frac{p_i^2}{q_i^*} \sigma_i^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} p_j \sigma_j\right) p_i \sigma_i = \left(\sum_{i=1}^{m} p_i \sigma_i\right)^2$$

We compute:

$$\mathrm{Var}(g(X)) = \mathbb{E}\left(\left(\sum_{i=1}^{m} g(X) \mathbb{1}_{D_i}(X)\right)^2\right) - \mathbb{E}(g(X))^2$$

$$= \mathbb{E}\left(\left(\sum_{i=1}^{m} g(X) \mathbb{1}_{D_i}(X)\right)^2\right) - \left(\sum_{i=1}^{m} p_i \mathbb{E}(g(X) | X \in D_i)\right)^2$$

$$= \sum_{i=1}^{m} \mathbb{E}(g(X)^2 \mathbb{1}_{D_i}(X)) - \left(\sum_{i=1}^{m} p_i \mathbb{E}(g(X) | X \in D_i)\right)^2$$

$$= \sum_{i=1}^{m} \left\{ p_i \mathbb{E}(g(X)^2 | X \in D_i) - p_i \mathbb{E}(g(X) | X \in D_i)^2 \right\}$$

$$+ \underbrace{\sum_{i=1}^{m} p_i \mathbb{E}(g(X) | X \in D_i)^2 - \left(\sum_{i=1}^{m} p_i \mathbb{E}(g(X) | X \in D_i)\right)^2}_{\geq 0 \text{ by Minkowski's Inequality}}$$

So $\text{Var}(g(X)) \geq \sum_{i=1}^{m} \{ p_i \mathbb{E}(g(X)^2 | X \in D_i)$
$- p_i \mathbb{E}(g(X) | X \in D_i)^2 \}$

$\geq \left(\sum_{i=1}^{m} p_i (\mathbb{E}(g(X)^2 | X \in D_i) - \mathbb{E}(g(X) | X \in D_i)^2)^{1/2} \right)^2$

again Minkowski's Inequality

remember (if $\sum_{i=1} \lambda_i = 1$, $\lambda_i \geq 0$)

$$\boxed{\sum_{i=1}^{m} \lambda_i x_i^2 \geq \left(\sum_{i=1}^{m} \lambda_i x_i \right)^2}$$

(It is also a convexity inequality)

So $\text{Var}(g(X)) \geq \left(\sum_{i=1}^{m} p_i \sigma_i \right)^2$.

3.4 Importance sampling

We want to compute $\mathbb{E}(h(X))$ where X has a density $f : (E, \mathcal{E}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ with respect to some reference measure μ.

We have (first approximation): (1)

$$\mathbb{E}(h(X)) = \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} h(X_k)$$

X_k → ind. copies of X

For any probability density function g, we have:

$$\mathbb{E}(h(X)) = \int h(x) f(x) \mu(dx)$$

$$= \int \frac{h(x) f(x)}{g(x)} \cdot g(x) \mu(dx)$$

$$= \mathbb{E}\left(\frac{h(Y) f(Y)}{g(Y)} \right) \text{ with } Y \text{ of density } g$$

$$= \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \frac{h(Y_k) f(Y_k)}{g(Y_k)} \quad (2)$$

Y_k → iid copies of Y

In practice

* We need to be able to compute $\frac{h(x) f(x)}{g(x)}$ for all x.

* We need to be able to make a simulation of Y at a reasonable cost (same as for X).
* Comparing the variances of methods ① and ② amounts to comparing the squared quadratic norms (since they have same expectation). So the new method is interesting if

$$\underbrace{\mathbb{E}\left(\left(h(Y)\frac{f(Y)}{g(Y)}\right)^2\right)} < \mathbb{E}(h(X)^2)$$

$$= \int \frac{h(y)^2 f(y)^2}{g(y)^2} \times g(y)\,\mu(dy)$$

$$= \int \frac{h(y)^2 f(y)}{g(y)} \times f(y)\,\mu(dy)$$

$$= \mathbb{E}\left(\frac{h(X)^2 f(X)}{g(X)}\right)$$

We have: $\mathbb{E}(h(X))^2 = \left(\int \frac{h(x)f(x)}{\sqrt{g(x)}}\sqrt{g(x)}\,\mu(dx)\right)^2$

$$\overset{(C-S)}{\leq} \int \frac{(h(x)f(x))^2}{g(x)^2}\, g(x)\,\mu(dx) \quad \times \int g(x)\mu(dx)$$

$$= \mathbb{E}\left(\left(\frac{h(Y)f(Y)}{g(Y)}\right)^2\right)$$

with equality if and only if

$$g(x) \propto \frac{(h(x)f(x))^2}{g(x)}$$ "proportional to"

as g is a density, this means $g(x) = \frac{h(x)f(x)}{\int f(u)h(u)\mu(du)}$

$$= \frac{h(x)f(x)}{\mathbb{E}(h(X))}$$

This cannot be attained because we do not know $\mathbb{E}(h(X))$.

3.4.1 Parametric importance sampling

Example : We have a Black-Scholes model with $X_T^x = x e^{\mu T + \sigma W_T}$

$$= x e^{\mu T + \sigma\sqrt{T} Z} \quad (Z \sim \mathcal{N}(0,1)) \rightarrow \text{density } f$$

$(x > 0,\ \sigma > 0,\ \mu = r - \frac{\sigma^2}{2})$. Then the premium of an option with payoff $h : (0, +\infty) \to [0, +\infty)$ reads :

$$\cancel{e^{-rT}}\,\mathbb{E}(h(X_T^x)) = \mathbb{E}(\varphi(Z)) = \int_{\mathbb{R}} \varphi(z) e^{-z^2/2} \frac{dz}{\sqrt{2\pi}}$$

$$\left(\varphi(z) := \cancel{e^{-rT}} h(x e^{\mu T + \sigma\sqrt{T} z})\right)$$

We introduce a variable $Y_\theta \sim \mathcal{N}(\theta, 1)$, density g_θ.

We can rewrite :

$$\mathbb{E}(\varphi(Z)) = \mathbb{E}\left(\frac{\varphi(Y_\theta) f(Y_\theta)}{g_\theta(Y_\theta)}\right) \quad (f \text{ density of } Z)$$

We want to find θ minimizing :

$$\mathbb{E}\left(\frac{\varphi(Z)^2 f(Z)}{g_\theta(Z)}\right)$$

$$\hookrightarrow = \int \frac{\varphi(z)^2 e^{-z^2/2}}{e^{-(z-\theta)^2/2}} dz$$

3.4.2 Heuristic approach

* In the above example, suppose :

$$\varphi(z) = \cancel{e^{-rT}}(x e^{\mu T + \sigma\sqrt{T} z} - K)_+$$

with $x \ll K$ (deep-out-of-the-money option)

For most simulations of Z : $\varphi(Z(\omega)) = 0$

The idea is to "re-center the simulation" of X_T^x around K by replacing Z by $Z+\theta$ where θ satisfies $\mathbb{E}\left(x \exp\left(rT + \sigma\sqrt{T}(Z+\theta)\right)\right) = K$

as $\mathbb{E}(X_T^x) = x e^{rT}$, this leads to

$$\theta = -\frac{\log(x/K) + rT}{\sigma\sqrt{T}}$$

* As the ideal density is $g^*(x) = \frac{h(x) f(x)}{\mathbb{E}(h(X))}$, we can look for a density $g \propto h \cdot f$ from which we are able to simulate a random variable.

3.4.3 Computing the Value-At-Risk

Let X be a real-valued r.v. (representing a loss). We suppose X has a continuous distribution ($F(x) := \mathbb{P}(X \leq x)$ is continuous). For a given confidence level $\alpha \in (0,1)$ (close to 1), the value-at-risk at level α ($VaR_{\alpha,X}$) is any real number satisfying:

$\mathbb{P}(X \leq VaR_{\alpha,X}) = \alpha \in (0,1)$ $\boxed{VaR}$

Naive approach: estimate the empirical distribution function at some points ξ (on a grid $\Gamma := \{\xi_i, i \in I\}$), namely

$$\widehat{F(\xi)}_M := \frac{1}{M}\sum_{k=1}^{M} \mathbb{1}_{X_k \leq \xi} \qquad (\xi \in \Gamma)$$

↳ iid sequence of X-distributed r.v.

We need to simulate extreme values of X since α is close to 1 (so the variance will not be good).

Assume, for example, that:

$X = \varphi(Z)$, $Z \sim \mathcal{N}(0,1)$

Then $\mathbb{P}(X \leq \xi) = \mathbb{E}(\mathbb{1}_{\varphi(Z) \leq \xi})$

$$= \int \mathbb{1}_{\varphi(z) \leq \xi} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

(change of variable) $$= \int \mathbb{1}_{\varphi(u+\theta) \leq \xi} \frac{e^{-(u+\theta)^2/2}}{\sqrt{2\pi}} du$$

$$= e^{-\theta^2/2} \mathbb{E}(\mathbb{1}_{\varphi(Z+\theta) \leq \xi} e^{-\theta Z})$$

$$= e^{\theta^2/2} \mathbb{E}(\mathbb{1}_{\varphi(Z+\theta) \leq \xi} e^{-\theta(Z+\theta)})$$

So Equation $\boxed{VaR}$ becomes:

$$\mathbb{E}(\mathbb{1}_{\varphi(Z+\theta) \leq VaR_{\alpha,X}} e^{-\theta Z}) = e^{\theta^2/2} \alpha$$

We now want θ to minimize the variance in the neighborhood of $VaR_{\alpha,X}$.

Idea: if we can approximate correctly

$x \mapsto \mathbb{E}(\mathbb{1}_{\varphi(Z+\theta) \leq x} e^{-\theta Z})$

in the neighborhood of $VaR_{\alpha,X}$ then it will be easier to find $VaR_{\alpha,X}$

3.5. Antithetic variables

i) Symmetric random variable $Z \overset{law}{=} -Z$

ii) $[0,L]$-valued random variable such that $Z \overset{law}{=} L - Z$

Example: We want to compute $\mathbb{E}(\varphi(Z))$, $Z \sim \mathcal{N}(0,1)$

$$\mathbb{E}\left(\frac{\varphi(Z) + \varphi(-Z)}{2}\right) = \mathbb{E}(\varphi(Z))$$

$$\mathbb{E}\left(\left(\frac{\varphi(z)+\varphi(-z)}{2}\right)^2\right) = \mathbb{E}\left(\frac{\varphi(z)^2+\varphi(-z)^2+2\varphi(z)\varphi(-z)}{4}\right)$$

$$= \frac{\mathbb{E}(\varphi(z)^2)}{2} + \frac{\mathbb{E}(\varphi(z)\varphi(-z))}{2}$$

$$\leq \frac{\mathbb{E}(\varphi(z)^2)}{2} + \frac{\mathbb{E}(\varphi(z)^2)^{1/2}\,\mathbb{E}(\varphi(-z)^2)^{1/2}}{2}$$

$$= \mathbb{E}(\varphi(z)^2)$$

So the variance is reduced

25