Solutions — Monte Carlo Methods
MIDO Master 1 (2025–2026)

Prepared by ChatGPT

Final Exam — 06/01/2026

PART I: Advanced Monte Carlo methods

Question 1 (Antithetic variates)

Let $U \sim \mathrm{Unif}(0,1)$ and $I = \mathbb{E}[g(U)]$. Define

$$\hat{I}_N^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} g(U_i), \qquad \hat{I}_N^{\mathrm{ant}} = \frac{1}{2N}\sum_{i=1}^{N}\big(g(U_i) + g(1-U_i)\big),$$

with U_i i.i.d. $\mathrm{Unif}(0,1)$.

(a) **Sufficient condition for variance reduction.**

Write for a single pair

$$Y_i = \frac{g(U_i) + g(1-U_i)}{2}.$$

Then $\mathbb{E}[Y_i] = \mathbb{E}[g(U)] = I$. We compare $\mathrm{Var}(Y_i)$ and $\mathrm{Var}(g(U))$. Using covariance,

$$\mathrm{Var}(Y_i) = \frac{1}{4}\big(\mathrm{Var}(g(U)) + \mathrm{Var}(g(1-U)) + 2\,\mathrm{Cov}(g(U), g(1-U))\big).$$

Since $g(1-U)$ has same law as $g(U)$, $\mathrm{Var}(g(1-U)) = \mathrm{Var}(g(U))$. Therefore

$$\mathrm{Var}(Y_i) = \frac{1}{2}\mathrm{Var}(g(U)) + \frac{1}{2}\mathrm{Cov}(g(U), g(1-U)).$$

Thus a sufficient (and easily interpretable) condition for $\mathrm{Var}(\hat{I}_N^{\mathrm{ant}}) \le \mathrm{Var}(\hat{I}_N^{\mathrm{MC}})$ is

$$\mathrm{Cov}(g(U), g(1-U)) \le \frac{1}{2}\mathrm{Var}(g(U)) \quad \Longrightarrow \quad \mathrm{Var}(Y_i) \le \mathrm{Var}(g(U)).$$

(ChatGPT did not spot that this condition always holds!)
In particular, a simple sufficient condition is $\mathrm{Cov}(g(U), g(1-U)) \le 0$. Equivalently, if g is monotone (either nondecreasing or nonincreasing) on $[0,1]$, then $g(U)$ and $g(1-U)$ are negatively correlated and antithetic sampling reduces variance.

Justification: If g is monotone nondecreasing then $g(U)$ and $g(1-U)$ are negatively correlated because when U is large $1-U$ is small, so their product covariance is negative; this yields variance reduction.

(b) **Apply to $g_1(x) = e^x$, $g_2(x) = (x-0.5)^2$, $g_3(x) = \sin(4\pi x)$.**

- $g_1(x) = e^x$ is strictly increasing on $[0, 1]$. Hence $g_1(U)$ and $g_1(1 - U)$ are negatively correlated $\Rightarrow$ antithetic reduces variance.
- $g_2(x) = (x - 0.5)^2$ is symmetric about $x = 0.5$ and convex. Note $g_2(1 - x) = g_2(x)$ so $g_2(U) = g_2(1 - U)$ almost surely. Then

$$Y_i = \frac{g_2(U_i) + g_2(1 - U_i)}{2} = g_2(U_i),$$

so antithetic sampling gives exactly the same estimator as standard Monte Carlo; there is *no variance reduction* (variance is identical).
- $g_3(x) = \sin(4\pi x)$ has period $1/2$ and is an odd-symmetric-like oscillatory function on $[0, 1]$: $\sin(4\pi(1 - x)) = \sin(4\pi - 4\pi x) = -\sin(4\pi x)$. Thus $g_3(1 - x) = -g_3(x)$. Hence

$$Y_i = \frac{g_3(U_i) + g_3(1 - U_i)}{2} = \frac{g_3(U_i) - g_3(U_i)}{2} = 0,$$

so antithetic estimator has zero variance (it gives the exact mean 0 for each pair). Therefore antithetic sampling *greatly* reduces variance (to 0 for paired estimator).

Question 2 (Control variates)

Let $X \sim \text{Unif}(0, 1)$, target $\mu = \mathbb{E}[X^3]$, and control $h_0(X) = X^2$ with known mean $\mathbb{E}[h_0(X)] = 1/3$.

(a) **Control variate estimator and unbiasedness.**

For sample $X_1, \ldots, X_N$ i.i.d., define the estimator

$$\hat{\mu}_\beta = \frac{1}{N} \sum_{i=1}^{N} \big(X_i^3 - \beta(X_i^2 - 1/3)\big).$$

Then $\mathbb{E}[\hat{\mu}_\beta] = \mathbb{E}[X^3] - \beta(\mathbb{E}[X^2] - 1/3) = \mu - \beta(1/3 - 1/3) = \mu$, so it is unbiased for any $\beta \in \mathbb{R}$.

(b) **Optimal coefficient $\beta^\star$.**

Minimize $\text{Var}(X^3 - \beta(X^2 - 1/3))$ w.r.t. β. Optimal $\beta^\star$ is given by

$$\beta^\star = \frac{\text{Cov}(X^3, X^2)}{\text{Var}(X^2)}.$$

Compute the moments: for $X \sim \text{Unif}(0, 1)$,

$$\mathbb{E}[X^2] = \frac{1}{3}, \quad \mathbb{E}[X^3] = \frac{1}{4}, \quad \mathbb{E}[X^4] = \frac{1}{5}.$$

Then

$$\text{Cov}(X^3, X^2) = \mathbb{E}[X^5] - \mathbb{E}[X^3]\mathbb{E}[X^2], \quad \mathbb{E}[X^5] = \frac{1}{6}.$$

So

$$\text{Cov}(X^3, X^2) = \frac{1}{6} - \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{6} - \frac{1}{12} = \frac{1}{12}.$$

And

$$\text{Var}(X^2) = \mathbb{E}[X^4] - (\mathbb{E}[X^2])^2 = \frac{1}{5} - \frac{1}{9} = \frac{9 - 5}{45} = \frac{4}{45}.$$

Hence

$$\beta^\star = \frac{\frac{1}{12}}{\frac{4}{45}} = \frac{1}{12} \cdot \frac{45}{4} = \frac{45}{48} = \frac{15}{16} = 0.9375.$$

Question 3 (Stratified sampling for $g(x) = e^x$)

We stratify $[0, 1)$ into K equal-length strata $D_k = [(k-1)/K, k/K)$, $k = 1, \dots, K$. In stratum D_k, X conditional is uniform on D_k.

(a) **Intra-stratum mean and variance.**

For $X \sim \text{Unif}([(k-1)/K, k/K))$,

$$\mu_k = \mathbb{E}[e^X \mid X \in D_k] = \int_{(k-1)/K}^{k/K} e^x \frac{K\,\mathrm{d}x}{1} = K\big(e^{k/K} - e^{(k-1)/K}\big).$$

The second moment in stratum:

$$\mathbb{E}[e^{2X} \mid X \in D_k] = K \int_{(k-1)/K}^{k/K} e^{2x}\mathrm{d}x = \frac{K}{2}\big(e^{2k/K} - e^{2(k-1)/K}\big).$$

Thus the intra-stratum variance is

$$\sigma_k^2 = \text{Var}(e^X \mid X \in D_k) = \mathbb{E}[e^{2X}|D_k] - \mu_k^2 = \frac{K}{2}\big(e^{2k/K} - e^{2(k-1)/K}\big) - \Big(K\big(e^{k/K} - e^{(k-1)/K}\big)\Big)^2,$$

which depends on k (so variances differ across strata).

(b) **Variance of stratified estimator.**

With $Y_{k,j} = e^{X_{k,j}}$, $Y_{\bar{k}} = \frac{1}{N_k}\sum_{j=1}^{N_k} Y_{k,j}$ and the stratified estimator

$$I_{\text{str}} = \frac{1}{K}\sum_{k=1}^{K} Y_{\bar{k}},$$

assuming independent simulation across and within strata,

$$\text{Var}(I_{\text{str}}) = \frac{1}{K^2}\sum_{k=1}^{K} \frac{\sigma_k^2}{N_k}.$$

(c) **Optimal allocation N_k (N total).**

Minimize $\text{Var}(I_{\text{str}})$ under constraint $\sum_{k=1}^{K} N_k = N$. Standard result (Neyman allocation) yields

$$N_k^\star \propto \sigma_k \quad \Longrightarrow \quad N_k^\star = N\frac{\sigma_k}{\sum_{j=1}^{K}\sigma_j}.$$

(d) **R function MC(N).**

R pseudocode (to be placed in R file):

```
MC <- function(N){
  X <- runif(N)
  Y <- exp(X)
  est <- mean(Y)
  var_est <- var(Y)/N
  return(list(est=est, var=var_est))
}
```

(e) **R function Stratified_prop(N,K).**

Proportional allocation (i.e. $N_k = \lfloor N/K \rfloor$ or exactly $N_k = \lfloor N/K \rfloor$ with remainder). R pseudocode:

```
Stratified_prop <- function(N, K){
  base <- floor(N/K)
  rem <- N - base*K
  Ns <- rep(base, K)
  if(rem>0) Ns[1:rem] <- Ns[1:rem]+1
  estimates <- numeric(K)
  vars <- numeric(K)
  for(k in 1:K){
    a <- (k-1)/K
    b <- k/K
    Xk <- runif(Ns[k], min=a, max=b)
    Yk <- exp(Xk)
    estimates[k] <- mean(Yk)
    vars[k] <- var(Yk)/Ns[k]
  }
  est <- mean(estimates)
  var_est <- sum(vars)/K^2
  return(list(est=est, var=var_est))
}
```

PART II: Multilevel Monte Carlo (MLMC)

Question 1 (Standard Monte Carlo)

Let $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ with i.i.d. samples.

(a) $\mathrm{Var}(\hat{\mu}_N) = \mathrm{Var}(X)/N$. Hence to achieve $\mathrm{MSE}(\hat{\mu}_N) = \mathbb{E}[(\hat{\mu}_N - \mu)^2] = \mathrm{Var}(\hat{\mu}_N) = O(\varepsilon^2)$ we require $N = O(\varepsilon^{-2})$.

(b) If cost per sample is C, total cost is $C_{\text{tot}} = N \cdot C = O(C\varepsilon^{-2})$.

Question 2 (Quadrature hierarchy)

Define for integer $p \geq 1$,

$$Q_\ell = \frac{1}{2^\ell}\sum_{k=1}^{2^\ell}\Big(\frac{k}{2^\ell}\Big)^p.$$

(a) $I = \int_0^1 x^p \mathrm{d}x = \frac{1}{p+1}$.

(b) Using Riemann sum error for smooth integrands, the uniform grid Riemann sum error scales as $O(2^{-\ell})$. Hence $Q_\ell - I = O(2^{-\ell})$.

(c) Using Faulhaber's formula

$$\sum_{k=1}^{N} k^p = \frac{N^{p+1}}{p+1} + \frac{N^p}{2} + O(N^{p-1}),$$

set $N = 2^\ell$ and compute difference

$$Q_\ell - Q_{\ell-1} = O(2^{-\ell}),$$

so (Q_ℓ) is Cauchy and converges (to I).

Question 3 (Telescoping identity)

For integrable random variables Y_ℓ,

$$\mathbb{E}[Y_L] = \mathbb{E}[Y_0] + \sum_{\ell=1}^{L} \mathbb{E}[Y_\ell - Y_{\ell-1}],$$

which is immediate by telescoping the finite sum.

Question 4 (Monte Carlo version of telescoping sum)

Let Z_ℓ be r.v. with $\mathbb{E}[Z_\ell] = Q_\ell$, and define

$$\hat{Y}_L = \frac{1}{N_0} \sum_{i=1}^{N_0} Z_0^{(i)} + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \big(Z_\ell^{(i)} - Z_{\ell-1}^{(i)}\big).$$

(a) By linearity of expectation and $\mathbb{E}[Z_\ell] = Q_\ell$, $\mathbb{E}[\hat{Y}_L] = Q_0 + \sum_{\ell=1}^{L}(Q_\ell - Q_{\ell-1}) = Q_L$.

(b) If $N_\ell \to \infty$ for all ℓ, each term converges in L^2 to its expectation (variance $\to 0$), hence $\hat{Y}_L \to Q_L$ in L^2, thus also in L^1.

(c) Choice of (N_ℓ) affects variance since $\mathrm{Var}(\hat{Y}_L) = \sum_{\ell=0}^{L} \mathrm{Var}(D_\ell)/N_\ell$ plus covariances if dependent; allocating more samples to levels with larger variance reduces total variance.

(d) If coupling uses same $Z_\ell^{(i)}$ across adjacent differences, cross-level covariances appear: with $D_\ell^{(i)} = Z_\ell^{(i)} - Z_{\ell-1}^{(i)}$,

$$\mathrm{Var}(\hat{Y}_L) = \sum_{\ell=0}^{L} \frac{\mathrm{Var}(D_\ell)}{N_\ell} + 2 \sum_{\ell<m} \frac{\mathrm{Cov}(\sum_i D_\ell^{(i)}/N_\ell, \sum_j D_m^{(j)}/N_m)}{\cdot}$$

If the same indices are used, cross terms for adjacent levels may not vanish.

(e) If level terms are independent, cross-level covariance terms vanish and

$$\mathrm{Var}(\hat{Y}_L) = \sum_{\ell=0}^{L} \frac{\mathrm{Var}(D_\ell)}{N_\ell}.$$

This is smaller or simpler to analyze than the coupled case but coupling typically reduces variances of differences and is therefore preferred.

Question 5 (Variance and cost analysis)

Let $D_\ell^{(i)} = Z_\ell^{(i)} - Z_{\ell-1}^{(i)}$, $v_\ell = \mathrm{Var}(D_\ell)$, cost per sample c_ℓ, covariance $\gamma_{\ell,m} = \mathrm{Cov}(D_\ell, D_m)$ and assume $\gamma_{\ell,m} = 0$ for $|\ell - m| > 1$.

(a) Aggregating variances and covariances (with possible non-equal N_ℓ), one obtains

$$\mathrm{Var}(\hat{Y}_L) = \sum_{\ell=0}^{L} \frac{v_\ell}{N_\ell} + 2 \sum_{\ell=0}^{L-1} \frac{\gamma_{\ell,\ell+1}}{\max(N_\ell, N_{\ell+1})},$$

where the $\max(N_\ell, N_{\ell+1})$ appears when samples for adjacent levels are shared or partially shared; this matches the exam statement. *(But since (N_ℓ) is non-increasing, this amounts to using N_ℓ)*

(b) In the independent-level case ($\gamma_{\ell,m} = 0$ for all $\ell \neq m$), minimize total cost $C = \sum_{\ell=0}^{L} N_\ell c_\ell$ subject to $\sum_{\ell=0}^{L} v_\ell / N_\ell \leq V_0$. Lagrangian:

$$\mathcal{L} = \sum_{\ell=0}^{L} N_\ell c_\ell + \lambda\Big(\sum_{\ell=0}^{L} \frac{v_\ell}{N_\ell} - V_0\Big).$$

Stationarity: $c_\ell - \lambda v_\ell N_\ell^{-2} = 0 \Rightarrow N_\ell = \sqrt{\lambda}\sqrt{v_\ell / c_\ell}$. Using constraint,

$$\sum_{\ell=0}^{L} \frac{v_\ell}{N_\ell} = \frac{1}{\sqrt{\lambda}} \sum_{\ell=0}^{L} \sqrt{v_\ell c_\ell} = V_0.$$

Hence $\sqrt{\lambda} = \frac{1}{V_0} \sum_{\ell=0}^{L} \sqrt{v_\ell c_\ell}$ and

$$N_\ell^\star = \sqrt{v_\ell / c_\ell} \cdot \frac{1}{V_0} \sum_{m=0}^{L} \sqrt{v_m c_m}.$$

Total minimal cost:

$$C_{\min} = \sum_{\ell=0}^{L} N_\ell^\star c_\ell = \frac{1}{V_0}\Big(\sum_{\ell=0}^{L} \sqrt{v_\ell c_\ell}\Big)^2.$$

Question 6 (Optimal MLMC complexity)

Assume $v_\ell \asymp 2^{-2\ell}$ and $c_\ell \asymp 2^{\ell}$.

(a) Using $C_{\min}$ with $V_0 \asymp \varepsilon^2$,

$$\sqrt{v_\ell c_\ell} \asymp \sqrt{2^{-2\ell} \cdot 2^{\ell}} = 2^{-\ell/2}.$$

Then

$$\sum_{\ell=0}^{L} \sqrt{v_\ell c_\ell} \asymp \sum_{\ell=0}^{L} 2^{-\ell/2} \asymp O(1)$$

(as $L \to \infty$ this is a convergent geometric series). Hence

$$C_{\min} \asymp \frac{1}{V_0} \cdot O(1) = O(\varepsilon^{-2}).$$

Thus MLMC cost scales as ε^{-2}.

(b) For standard Monte Carlo at finest level L with bias and variance both $O(\varepsilon^2)$, choose L such that bias $|E[Z_L] - I| \asymp 2^{-L} \asymp \varepsilon$. Thus $2^L \asymp \varepsilon^{-1}$ and cost per sample $c_L \asymp 2^L \asymp \varepsilon^{-1}$. Number of samples $N \asymp \varepsilon^{-2}$ to control variance, therefore total cost

$$C_{\mathrm{SL}} \asymp N \cdot c_L \asymp \varepsilon^{-2} \cdot \varepsilon^{-1} = \varepsilon^{-3}.$$

Hence standard Monte Carlo at fine resolution is ε^{-3} while MLMC attains ε^{-2}: MLMC is asymptotically far cheaper.

Question 7 (Rigorous MLMC error bound)

Assume bias $|E[Y_L] - I| \le B2^{-L}$.

(a) Choose $L = \lceil \log_2(B/\varepsilon) \rceil$. Then

$$|E[Y_L] - I| \le B2^{-L} \le B2^{-\log_2(B/\varepsilon)} = \varepsilon.$$

(b) Using independence across levels and $v_\ell \asymp 2^{-2\ell}$,

$$\mathrm{Var}(\hat{Y}_L) = \sum_{\ell=0}^{L} \frac{v_\ell}{N_\ell} \le C \sum_{\ell=0}^{L} \frac{2^{-2\ell}}{N_\ell}.$$

(c) Choose $N_\ell \asymp 2^{-\ell}\varepsilon^{-2}$ (note: N_ℓ decreasing in ℓ). Then

$$\frac{2^{-2\ell}}{N_\ell} \asymp \frac{2^{-2\ell}}{2^{-\ell}\varepsilon^{-2}} = 2^{-\ell}\varepsilon^2.$$

Summing gives $\mathrm{Var}(\hat{Y}_L) \asymp \varepsilon^2 \sum_{\ell=0}^{L} 2^{-\ell} = O(\varepsilon^2)$.

(d) Combining bias squared $O(\varepsilon^2)$ and variance $O(\varepsilon^2)$ yields MSE $O(\varepsilon^2)$.

Question 8 (Debiasing single-term estimator)

Let $(Y_\ell)_{\ell \ge 1}$ be biased approximations with $\mathbb{E}[Y_\ell] \to \mu$. We further define $Y_0 = 0$.

(a) Formal telescoping:

$$\mu = \lim_{L\to\infty} \mathbb{E}[Y_L] = \sum_{\ell=1}^{\infty} \mathbb{E}[Y_\ell - Y_{\ell-1}],$$

provided the infinite sum converges absolutely (or at least conditionally with integrability).

(b) Suppose L is random with $\Pr(L = \ell) = p_\ell > 0$ for $\ell \ge 1$ and p_ℓ chosen so $\sum p_\ell = 1$. Consider the single-term unbiased estimator

$$\hat{Y} = \frac{Y_L - Y_{L-1}}{p_L}.$$

Then

$$\mathbb{E}[\hat{Y}] = \sum_{\ell \ge 1} p_\ell \frac{\mathbb{E}[Y_\ell - Y_{\ell-1}]}{p_\ell} = \sum_{\ell \ge 1} \mathbb{E}[Y_\ell - Y_{\ell-1}] = \mu$$

yields an unbiased estimator of μ. Unbiasedness requires the series of expectations to converge and $\sum p_\ell = 1$ with $p_\ell > 0$ for terms used.

(c) Suppose $\mathrm{Var}(Y_\ell - Y_{\ell-1}) \asymp 2^{-2\ell}$ and choose $p_\ell \propto 2^{-\alpha\ell}$. Then

$$\mathrm{Var}\Big(\frac{Y_\ell - Y_{\ell-1}}{p_\ell}\Big) \asymp \frac{2^{-2\ell}}{p_\ell^2} \propto 2^{-2\ell} \cdot 2^{2\alpha\ell} = 2^{(2\alpha-2)\ell}.$$

For $\mathrm{Var}(\hat{Y}) = \sum_\ell p_\ell \,\mathrm{Var}\big((Y_\ell - Y_{\ell-1})/p_\ell\big)$ *(not exactly, ChatGPT!)* to be finite (roughly), we need the terms to decay: require $2\alpha - 2 < 0 \Rightarrow \alpha < 1$. More carefully, asymptotics show that we need $\alpha < 1$ to have finite variance.

(d) Cost: if cost to compute Y_ℓ is $O(2^\ell)$ and $p_\ell \propto 2^{-\alpha\ell}$ then expected cost

$$\mathbb{E}[\text{cost}] \propto \sum_{\ell \geq 1} p_\ell 2^\ell \propto \sum_{\ell \geq 1} 2^{(1-\alpha)\ell},$$

which is finite only if $\alpha > 1$. Thus there is a trade-off:

- For finite variance require $\alpha < 1$.
- For finite expected cost require $\alpha > 1$.

Therefore no single α makes both variance and expected cost finite in this toy scaling; in practice one tunes the tail p_ℓ to balance variance and cost (e.g., choose slightly above 1 and add variance control by coupling or truncation). Compared to MLMC: MLMC attains optimal ε^{-2} cost under favorable scalings, while naive debiasing with single-term randomization can have worse cost/variance trade-offs unless p_ℓ is chosen very carefully; unbiased debiasing can be competitive but typically requires careful design (e.g., stratified randomization, coupled differences).