Lecture Notes 12
Variance-Reduction Techniques

Prof. Marvin Nakayama

Outline

- Introduction
- Common Random Numbers
- Antithetic Variates
- Control Variates
- Importance Sampling

Be sure to check course webpage for reading assignment and HW

Introduction

- Suppose that we want to estimate $\alpha = \mathcal{E}[X]$.
- Standard approach: generate IID replicates $X_1, X_2, \ldots, X_n$ of X.
- Estimator $\alpha_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ of α satisfies CLT: if $0 < \mathcal{V}[X] < \infty$,

$$\frac{\sqrt{n}}{\sqrt{\mathcal{V}[X]}}\left[\alpha_n - \alpha\right] \stackrel{\mathcal{D}}{\approx} \mathcal{N}(0,1)$$

- So error in estimator α_n is

$$\alpha_n - \alpha \stackrel{\mathcal{D}}{\approx} \frac{\sqrt{\mathcal{V}[X]}}{\sqrt{n}}\mathcal{N}(0,1)$$

- 95% confidence interval (CI) for α is

$$\left[\alpha_n \pm 1.96\sqrt{\frac{\mathcal{V}[X]}{n}}\right].$$

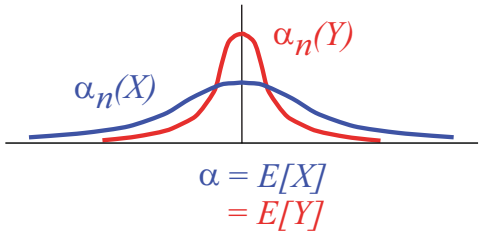
  - Typically, $\mathcal{V}[X]$ is unknown, so replace $\mathcal{V}[X]$ with S_n^2 in CI.

Introduction

- Recall: 95% CI half-width (CIHW) is $1.96\sqrt{\frac{\mathcal{V}[X]}{n}}$ from sampling X
- Can reduce CIHW by increasing sample size n.
- Alternatively, if there is another RV Y with $\mathcal{E}[Y] = \mathcal{E}[X] = \alpha$, can sample Y instead of X to estimate α.
- If $\mathcal{V}[Y] < \mathcal{V}[X]$
  - Smaller CIHW $1.96\sqrt{\frac{\mathcal{V}[Y]}{n}}$ for same sample size n
  - Need fewer outputs of Y to achieve same precision.
- Methods are called **variance-reduction techniques (VRTs)**.
  - Exploit knowledge of system to design Y with $\mathcal{V}[Y] < \mathcal{V}[X]$.

Introduction

Densities of estimators



- $\alpha_n(X)$ = estimator from sampling n outputs $X_1, X_2, \ldots, X_n$
- $\alpha_n(Y)$ = estimator from sampling n outputs $Y_1, Y_2, \ldots, Y_n$.
- But VRTs may require more computational effort per output.
  - Although $\mathcal{V}[Y] < \mathcal{V}[X]$, may take more time to generate output of Y than X.

Introduction

- Must be careful when evaluating these techniques:
  - Additional programming complexity.
  - Only worthwhile if

    reduction in variance
    outweighs
    increase in computational effort.
- For many of the VRTs we consider,
  - increase in computational effort from VRT is small
  - so the VRT is a clear win.
- How do we deal with more complicated situations?
  - Fix computer budget of c time units for each method.
  - Compare methods' CIHWs after c time units.

Example

- Suppose $\alpha = \mathcal{E}[X] = \mathcal{E}[Y]$.
  - So can estimate α by generating IID replicates of either X or Y.
  - Should we generate output X or Y ?
  - Approach: compare two approaches based on error (CIHW) for **fixed computation budget** c.
- First, consider generating replications of X, and let

$$c = \text{ computer budget (CPU time)}$$

$$X_i = i\text{th replicate of } X$$

$$\tau_x(i) = \text{ (random) computer time to generate } X_i$$

- Assume $(X_1, \tau_x(1)),\ (X_2, \tau_x(2)),\ \ldots$ are IID pairs.

Example

- Number of observations of X generated within budget c is

$$N_x(c) = \max\{n \geq 0 : \tau_x(1) + \cdots + \tau_x(n) \leq c\}$$

- Can show (using LLN for renewal processes) that with probability 1,

$$\frac{N_x(c)}{c} \to \lambda_x \qquad \text{as } c \to \infty$$

  where $\lambda_x = \frac{1}{\mathcal{E}[\tau_x]}$, so $N_x(c) \approx \lambda_x \cdot c$ for large c.
- Estimator for α based on X after expending budget c is

$$\begin{aligned} \alpha_x(c) &= \frac{1}{N_x(c)} \sum_{i=1}^{N_x(c)} X_i \\ &\approx \frac{1}{\lfloor \lambda_x \cdot c \rfloor} \sum_{i=1}^{\lfloor \lambda_x \cdot c \rfloor} X_i \end{aligned}$$

Example

- Since $\alpha_x(c) \approx \frac{1}{\lfloor \lambda_x \cdot c \rfloor} \sum_{i=1}^{\lfloor \lambda_x \cdot c \rfloor} X_i$ and $\lambda_x = \frac{1}{\mathcal{E}[\tau_x]}$, CLT implies CIHW based on sampling X with budget c is

$$\begin{aligned} \text{CIHW}_x &\approx 1.96\sqrt{\frac{\mathcal{V}(X)}{\lambda_x \cdot c}} \\ &= 1.96\sqrt{\frac{\mathcal{E}[\tau_x]\,\mathcal{V}(X)}{c}} \end{aligned}$$

- Similarly, CIHW based on sampling Y with budget c is

$$\text{CIHW}_y \approx 1.96\sqrt{\frac{\mathcal{E}[\tau_y]\,\mathcal{V}(Y)}{c}}$$

- Hence, choose estimator that minimizes

  (mean computer time per observation) $\times$ (variance per observation)

Improving Efficiency

- Reasonable measure of efficiency is

$$\frac{1}{(\text{mean computer time per observation}) \times (\text{variance per observation})}$$

- Want to increase efficiency.
  - VRTs sometimes instead called **efficiency-improvement techniques (EITs)**.
- Our definition of efficiency holds more generally.
  - Glynn, P. W. and Whitt, W. 1992. "The asymptotic efficiency of simulation estimators," *Operations Research*, Vol. 40, Iss. 3, 505–520.
- Simulation without VRTs or EITs is called **standard simulation**, **crude Monte Carlo**, or **simple random sampling**.

General Approach to Derive CI When Using VRT

- Suppose that α_n is estimator of α when using VRT.
- Design VRT estimator α_n by exploiting knowledge about system.
- To obtain CI for α, need to show α_n satisfies LLN and CLT.
  1. Show α_n can be expressed as or approximated by a sample average of IID RVs $Z_1, Z_2, \ldots, Z_n$ of Z with $\mathcal{E}[Z] = \alpha$.
  2. LLN: Since $Z_1, Z_2, \ldots, Z_n$ IID, $\frac{1}{n}\sum_{k=1}^{n} Z_k \approx \alpha$ for large n.
  3. CLT: For large n,

$$\sqrt{n}(\alpha_n - \alpha) \overset{\mathcal{D}}{\approx} \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} Z_k - \mathcal{E}[Z]\right) \overset{\mathcal{D}}{\approx} \mathcal{N}(0, \tau^2)$$

  where $\tau^2 = \mathcal{V}[Z]$, when $0 < \tau^2 < \infty$.
  4. Unfold CLT to obtain approximate 95% CI for α as $\left[\alpha_n \pm 1.96\frac{\tau_n}{\sqrt{n}}\right]$, where τ_n^2 is an estimator of $\mathcal{V}[Z]$.

Common Random Numbers (CRN)

- **Goal:** compare two or more alternate system configurations.
  - Queueing systems with different service-time distributions
  - Options with different strike prices
- **Intuition:** Get better comparison between two systems if they are both subjected to exact same random fluctuations
  - Use same streams of random numbers to simulate each system.
  - Then measured differences due to true differences in performance.

Common Random Numbers (CRN)

- More specifically, suppose under two possible system configurations, performance measures distributed as RVs X and Y.
- Want to estimate

$$\alpha = \mathcal{E}[X] - \mathcal{E}[Y]$$

- Suppose we simulated both systems to obtain IID replicates (X_1, Y_1), (X_2, Y_2), $\ldots$, (X_n, Y_n).
  - Elements within each pair (X_i, Y_i) may be correlated.
  - Generate X_i and Y_i using same stream of random numbers.

Common Random Numbers

- Estimator of $\alpha = \mathcal{E}[X] - \mathcal{E}[Y]$ is

$$\alpha_n = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i),$$

$$\text{so} \quad \mathcal{V}[\alpha_n] = \frac{1}{n}\mathcal{V}[X - Y] = \frac{1}{n}\Big(\mathcal{V}[X] + \mathcal{V}[Y] - 2\,\text{Cov}[X, Y]\Big).$$

- If X_i's simulated independently of Y_i's (standard approach), then

$$\text{Cov}[X, Y] = 0$$

- If we could somehow arrange things so that

$$\text{Cov}[X, Y] > 0,$$

then variance of α_n would be reduced.

Common Random Numbers

- Suppose
  - generating an output X of **first system** requires m IID unif[0, 1] $U_1, \ldots, U_m$, so
$$X = h_x(U_1, \ldots, U_m)$$
  - generating an output Y of **second system** requires m IID unif[0, 1] $U'_1, \ldots, U'_m$, so
$$Y = h_y(U'_1, \ldots, U'_m)$$
- Without CRN, we **independently** generate $(U_1, \ldots, U_m)$ and $(U'_1, \ldots, U'_m)$, and compute
$$X - Y = h_x(U_1, \ldots, U_m) - h_y(U'_1, \ldots, U'_m)$$
- Using CRN, use **same** $(U_1, \ldots, U_m)$ to compute
$$X - Y = h_x(U_1, \ldots, U_m) - h_y(U_1, \ldots, U_m)$$

Common Random Numbers

- Recall: with CRN, we let

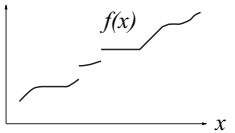
$$\begin{aligned} X &= h_x(U_1, \ldots, U_m), \\ Y &= h_y(U_1, \ldots, U_m), \\ \mathcal{V}[X - Y] &= \mathcal{V}[X] + \mathcal{V}[Y] - 2\text{Cov}[X, Y]. \end{aligned}$$

- For variance reduction (relative to indep simulations), need

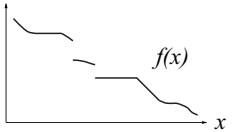
$$0 < \text{Cov}(X, Y) = \text{Cov}[\, h_x(U_1, \ldots, U_m),\ h_y(U_1, \ldots, U_m)\,]$$

- When can we guarantee $\text{Cov}(X, Y) > 0$?

CRN: Monotone Functions



- Let f be a function.
- **Defn:** f is **monotonically increasing** if $x < y$ implies $f(x) \le f(y)$.
- **Defn:** f is **monotonically decreasing** if $x < y$ implies $f(x) \ge f(y)$.



- **Defn:** f is **monotone** if it is either monotonically increasing or monotonically decreasing.
- **Thm:** $\text{Cov}(X, Y) > 0$ if h_x and h_y are both **monotone in same direction** in each argument
  - for each $i = 1, 2, \ldots, m$,
$$h_x(u_1, \ldots, u_i, \ldots, u_m) \quad \text{and} \quad h_y(u_1, \ldots, u_i, \ldots, u_m)$$
both m-increasing in u_i or both m-decreasing in u_i, other u_j fixed.
- May still have $\text{Cov}(X, Y) > 0$ if not monotone, but no guarantee.

Common Random Numbers: Point Estimator

- General approach for applying CRN to estimate $\alpha = \mathcal{E}[X] - \mathcal{E}[Y]$:
  - Use CRN to generate (X_1, Y_1), (X_2, Y_2), $\ldots$, (X_n, Y_n) as IID pairs of (X, Y).
  - For kth pair, generate X_k and Y_k using **same sequence** of uniform random numbers.
  - CRN estimator of α is
$$\alpha_n = \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} Y_k$$
- How do we construct CI for α when using CRN?
  - Need to show α_n satisfies LLN and CLT.
  - **Key Idea:** express α_n as sample average of IID RVs.

Common Random Numbers: LLN and CLT

- **Goal:** show CRN estimator α_n of $\alpha = \mathcal{E}[X] - \mathcal{E}[Y]$ satisfies LLN and CLT
- **Approach:** express α_n as sample average of IID RVs.
  - Let $D_k = X_k - Y_k$ for each $k = 1, \ldots, n$.
  - Since $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are IID pairs, $D_1, D_2, \ldots, D_n$ are IID, and
$$\alpha_n = \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} Y_k = \frac{1}{n}\sum_{k=1}^{n} D_k$$
  - $\mathcal{E}[D_k] = \mathcal{E}[X_k] - \mathcal{E}[Y_k] = \alpha$.
  - **LLN**: If $\mathcal{E}[|D_k|] < \infty$, then $\alpha_n \approx \alpha$ for large n.
  - **CLT**: If $0 < \mathcal{V}[D_k] < \infty$, then for large n,
$$\frac{\sqrt{n}}{\sqrt{\mathcal{V}[D_k]}}[\alpha_n - \alpha] \overset{\mathcal{D}}{\approx} \mathcal{N}(0, 1).$$

Common Random Numbers: Generic Algorithm

Algorithm: to construct CI for $\alpha = \mathcal{E}[X] - \mathcal{E}[Y]$ when using CRN.

1. For each independent replication $k = 1, 2, \ldots, n$,
   - Generate output pair (X_k, Y_k) using CRN.
     - Since X_k and Y_k generated with CRN, they are dependent.
   - Compute $D_k = X_k - Y_k$.
2. Compute sample mean and sample variance of IID $D_1, \ldots, D_n$:
$$\alpha_n = \frac{1}{n}\sum_{k=1}^{n} D_k, \qquad S_D^2(n) = \frac{1}{n-1}\sum_{k=1}^{n} (D_k - \alpha_n)^2,$$
to get 95% CI for α as
$$\left[\, \alpha_n \pm \frac{1.96\, S_D(n)}{\sqrt{n}} \,\right].$$

Common Random Numbers

- **Caveat:** If CRN is not done correctly,
  - can have Cov$[X, Y] < 0$
  - so variance of α_n will be **larger** than if X and Y indep.
- **In Practice:** To simulate real systems using this idea,
  - Try to synchronize random numbers between the two systems as much as possible
  - Use several streams of pseudorandom Unif[0, 1] numbers
    - ▷ one for each type of event in the system.
  - Arena
    - ▷ 1.8×10^{19} separate streams, each of length 1.7×10^{38}.
    - ▷ EXPO(4.3, 6) uses 6th stream to generate expo with mean 4.3
    - ▷ Stream 10 is default; EXPO(4.3) and EXPO(4.3, 10) are same.

Example: CRN for Comparing 2 G/G/1 Queues

- Compare expected average waiting time of first m customers for two G/G/1 queues.
  - $W_{i,j}$ = waiting time of jth customer in system i $(i = 1, 2)$.
  - $X = \frac{1}{m}\sum_{j=1}^{m} W_{1,j}$ and $Y = \frac{1}{m}\sum_{j=1}^{m} W_{2,j}$.
  - Want to estimate $\alpha = \mathcal{E}[X] - \mathcal{E}[Y]$.
- In both systems, interarrival times are IID with CDF G.
- In system i $(i = 1, 2)$, successive service times are IID with CDF H_i.
- To apply CRN, use **two streams** of Unif[0, 1] random numbers: $(U_j : j \geq 0)$ for interarrivals, and $(V_j : j \geq 0)$ for service times.
- For **both** systems, generate jth **interarrival** time using **inversion**:
$$A_j = G^{-1}(U_j).$$
  - Thus, jobs arrive at same times in both systems.
  - $G^{-1}(\cdot)$ is monotonically increasing function.

Example: CRN for Comparing 2 G/G/1 Queues

- For jth **service** time $S_{i,j}$ of system $i = 1, 2$, use **inversion** with **same** uniform V_j: $S_{1,j} = H_1^{-1}(V_j)$, $S_{2,j} = H_2^{-1}(V_j)$.
- $H_1^{-1}(\cdot)$ and $H_2^{-1}(\cdot)$ are both m-increasing functions.
- Lindley's equation for $(j+1)$th waiting time in system $i = 1, 2$,
$$W_{i,j+1} = [W_{i,j} - A_j + S_{i,j}]^+ = [W_{i,j} - G^{-1}(U_j) + H_i^{-1}(V_j)]^+ \equiv h_{i,j+1}(U_1, \ldots, U_j, V_1, \ldots, V_j)$$
- Can show $h_{1,j+1}$ and $h_{2,j+1}$ are monotone in same direction in each argument, so $\mathrm{Cov}[W_{1,j+1}, W_{2,j+1}] > 0$ and $\mathrm{Cov}[X, Y] > 0$.
- Instead, if we used only a **single stream** of random numbers, things may soon get **out of sync**.
  - The same pseudorandom number might be used to generate interarrival in one system and service time in other system.

Algorithm: CRN for Comparing 2 G/G/1 Queues

Goal: to construct CI for $\alpha = \mathcal{E}[X] - \mathcal{E}[Y]$ when using CRN

1. On each independent replication $k = 1, 2, \ldots, n$, do
   - Generate $U_{k,1}, U_{k,2}, \ldots, U_{k,m-1}$ and $V_{k,1}, V_{k,2}, \ldots, V_{k,m-1}$ as independent streams of IID unif[0, 1] random numbers.
   - Compute $D_k = X_k - Y_k$, where
$$X_k = \frac{1}{m}\sum_{j=1}^{m} W_{k,1,j}, \qquad Y_k = \frac{1}{m}\sum_{j=1}^{m} W_{k,2,j},$$
   with $W_{k,i,1} = 0$ for $i = 1, 2$,
$$W_{k,i,j} = (W_{k,i,j-1} + S_{k,i,j-1} - A_{k,i,j-1})^+, \quad j = 2, \ldots, m,$$
$$A_{k,1,j} = A_{k,2,j} = G^{-1}(U_{k,j}), \quad S_{k,i,j} = H_i^{-1}(V_{k,j}), \quad i = 1, 2.$$
2. Compute 95% CI for α as $[\bar{D}_n \pm 1.96 S_n/\sqrt{n}]$, where $\bar{D}_n = \frac{1}{n}\sum_{k=1}^{n} D_k$ and $S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(D_k - \bar{D}_n)^2$.

Example: CRN for Sensitivity of Euro Knockout Call Option

- Expiration T, which is m days
  - ΔT = one day; $T = m\,\Delta T$; $t_i = i\Delta T$ is end of ith day.
- K = strike price
- A = lower barrier, B = upper barrier
- S_t = price of underlying asset at time t; given S_0.
- σ = volatility
- r = risk-free rate
- **Payoff:** $I\{A < S_{t_i} < B,\ \forall\, i \leq m\}\ (S_T - K)^+$
  - If $A < S_{t_i} < B, \forall\, i = 1, 2, \ldots, m$, then payoff for option is $(S_T - K)^+ \equiv \max(S_T - K, 0)$
  - Otherwise, payoff is 0.
- $p(S_0)$ = fair price of knockout option when initial asset price is S_0
- **Goal:** to estimate $\Delta = \partial p(S_0)/\partial S_0$, which is useful for hedging.

Example: CRN for Sensitivity of Knockout Option

- Price of underlying asset follows geometric Brownian motion (GBM):

$$S_{t_i} \;=\; S_{t_{i-1}}\, e^{X_i}, \qquad i = 1, 2, \ldots, m$$

  - $X_1, \ldots, X_m \;\sim\; \mathcal{N}(\gamma, \psi^2)$ IID
  - $\gamma \;=\; \left(r - \frac{\sigma^2}{2}\right) \Delta T$ (mean under risk-neutral valuation)
  - $\psi^2 \;=\; \sigma^2 \Delta T$ (variance)
  - $S_{t_i} \;=\; S_{t_{i-1}}\, e^{X_i} \;=\; S_{t_{i-2}}\, e^{X_{i-1}}\, e^{X_i} \;=\; \cdots \;=\; S_0 \prod_{j=1}^{i} e^{X_j}$

- Discounted payoff of knockout option :

$$\begin{aligned} & e^{-rT}\, I\{A < S_{t_i} < B,\ \forall\, i \le m\}\, (S_T - K)^+ \\ &= \; e^{-rT}\, I\left\{A < S_0 \prod_{j=1}^{i} e^{X_j} < B,\ \forall\, i \le m\right\} \left(S_0 \prod_{j=1}^{m} e^{X_j} - K\right)^+ \\ &\equiv \; h(S_0, X_1, \ldots, X_m) \end{aligned}$$

Example: CRN for Sensitivity of Knockout Option

- **Recall:** Discounted payoff of knockout option

$$\begin{aligned} & h(S_0, X_1, \ldots, X_m) \\ &= \; e^{-rT}\, I\left\{A < S_0 \prod_{j=1}^{i} e^{X_j} < B,\ \forall\, i \le m\right\} \left(S_0 \prod_{j=1}^{m} e^{X_j} - K\right)^+ \end{aligned}$$

- Fair price of option using risk-neutral valuation:

$$p(S_0) \;=\; \mathcal{E}[\, h(S_0, X_1, \ldots, X_m)\,]$$

- **Goal:** to estimate the **delta** of the option:

$$\Delta \;=\; \frac{\partial}{\partial S_0} p(S_0)$$

  - Δ is useful for hedging.

Example: CRN for Sensitivity of Knockout Option

$$\Delta \;=\; \frac{\partial}{\partial S_0} p(S_0)$$

Sensitivity Δ can be defined 3 different ways

- Forward difference

$$\Delta \;=\; \lim_{\epsilon \to 0} \frac{p(S_0 + \epsilon) - p(S_0)}{\epsilon}$$

- Backward difference

$$\Delta \;=\; \lim_{\epsilon \to 0} \frac{p(S_0) - p(S_0 - \epsilon)}{\epsilon}$$

- Central difference

$$\Delta \;=\; \lim_{\epsilon \to 0} \frac{p(S_0 + \epsilon) - p(S_0 - \epsilon)}{2\epsilon}$$

- All 3 definitions equivalent since p is a "smooth" function of S_0

Example: CRN for Sensitivity of Knockout Option

- We'll develop an estimator of Δ based on the central difference:

$$\Delta \;=\; \lim_{\epsilon \to 0} \frac{p(S_0 + \epsilon) - p(S_0 - \epsilon)}{2\epsilon}.$$

- Fix small $\epsilon > 0$, and approximate Δ using **finite difference**

$$\Delta_\epsilon \;=\; \frac{p(S_0 + \epsilon) - p(S_0 - \epsilon)}{2\epsilon}$$

  - Estimate $p(S_0 + \epsilon)$ and $p(S_0 - \epsilon)$ using simulation.
  - Since ϵ is fixed, typically have $\Delta_\epsilon \neq \Delta$.
  - So even if $\widehat{p}(S_0 + \epsilon)$ and $\widehat{p}(S_0 - \epsilon)$ are unbiased estimators of $p(S_0 + \epsilon)$ and $p(S_0 - \epsilon)$,

$$\widehat{\Delta}_\epsilon \;=\; \frac{\widehat{p}(S_0 + \epsilon) - \widehat{p}(S_0 - \epsilon)}{2\epsilon}$$

    is a **biased** estimator of Δ; i.e., $\mathcal{E}[\widehat{\Delta}_\epsilon] \neq \Delta$ for each $\epsilon > 0$

Example: CRN for Sensitivity of Knockout Option

- Recall: use simulation to estimate finite difference
$$\Delta_\epsilon \;=\; \frac{p(S_0+\epsilon)-p(S_0-\epsilon)}{2\epsilon}$$
for fixed small $\epsilon > 0$.
- Without CRN
  - Estimate $p(S_0+\epsilon)$ and $p(S_0-\epsilon)$ using independent simulations.
- With CRN
  - Estimate $p(S_0+\epsilon)$ and $p(S_0-\epsilon)$ using **same** simulations.
    - ▷ Generate $X_1,\ldots,X_m \sim \mathcal{N}(\gamma,\psi^2)$ IID
    - ▷ Compute discounted payoffs
$$h(S_0+\epsilon,X_1,\ldots,X_m),$$
$$h(S_0-\epsilon,X_1,\ldots,X_m),$$
both using **same** $(X_1,\ldots,X_m)$.

Example: CRN for Sensitivity of Knockout Option

- Recall: use simulation to estimate central finite difference (fixed ϵ)
$$\Delta_\epsilon \;=\; \frac{p(S_0+\epsilon)-p(S_0-\epsilon)}{2\epsilon}$$
- Suppose that we generated n IID replicates
$(X_{1,1},X_{1,2},\ldots,X_{1,m}),\ (X_{2,1},X_{2,2},\ldots,X_{2,m}),\ \ldots,\ (X_{n,1},X_{n,2},\ldots,X_{n,m})$,
where the $X_{i,j} \sim \mathcal{N}(\gamma,\psi^2)$ IID
- Estimators of $p(S_0\pm\epsilon)$ are
$$\widehat{p}_n(S_0\pm\epsilon) \;=\; \frac{1}{n}\sum_{k=1}^{n} h(S_0\pm\epsilon,X_{k,1},\ldots,X_{k,m})$$
- Estimator of central finite difference Δ_ϵ is
$$\widehat{\Delta}_\epsilon(n) \;=\; \frac{\widehat{p}(S_0+\epsilon)\;-\;\widehat{p}(S_0-\epsilon)}{2\epsilon} \;=\; \frac{1}{n}\sum_{k=1}^{n} D_k$$
where
$$D_k \;=\; \frac{1}{2\epsilon}\left[\,h(S_0+\epsilon,X_{k,1},\ldots,X_{k,m}) \;-\; h(S_0-\epsilon,X_{k,1},\ldots,X_{k,m})\,\right]$$

Algorithm: CRN for Sensitivity of Knockout Option

1. Fix "small" $\epsilon > 0$.
2. On each independent replication $k = 1,2,\ldots,n$,
   - Generate $X_{k,1},\ X_{k,2},\ldots,\ X_{k,m} \ \sim\ \mathcal{N}(\gamma,\psi^2)$ IID.
   - Compute
$$D_k \;=\; \frac{1}{2\epsilon}\big[\,h(S_0{+}\epsilon,X_{k,1},\ldots,X_{k,m}) - h(S_0{-}\epsilon,X_{k,1},\ldots,X_{k,m})\,\big]$$
where
$$h(S,X_{k,1},\ldots,X_{k,m})$$
$$= e^{-rT} I\left\{A < S\prod_{j=1}^{i} e^{X_{k,j}} < B,\ \forall\, i \le m\right\}\left(S\prod_{j=1}^{m} e^{X_{k,j}} - K\right)^+$$
3. Compute 95% CI for Δ as $[\bar{D}_n \pm 1.96 S_{D,n}/\sqrt{n}\,]$, where $\bar{D}_n = \frac{1}{n}\Sigma_{k=1}^{n} D_k$ and $S_{D,n}^2 = \frac{1}{n-1}\Sigma_{k=1}^{n}(D_k-\bar{D}_n)^2$.

Remarks: CRN for Finite Difference

- Central finite-difference (CFD) estimator
$$\widehat{\Delta}_\epsilon \;=\; \frac{\widehat{p}(S_0+\epsilon)\;-\;\widehat{p}(S_0-\epsilon)}{2\epsilon}$$
is less biased than forward finite-difference (FFD) estimator
$$\widetilde{\Delta}_\epsilon \;=\; \frac{\widehat{p}(S_0+\epsilon)\;-\;\widehat{p}(S_0)}{\epsilon}$$
  - But CFD requires simulating at $S_0\pm\epsilon$ and S_0
  - FFD requires simulating at only $S_0+\epsilon$ and S_0
- Tricky to choose appropriately small $\epsilon \equiv \epsilon_n > 0$
  - CFD with CRN: $\epsilon_n = O(n^{-1/5})$ asymptotically minimizes MSE.
- There are better unbiased sensitivity estimators
  - Perturbation analysis
  - Likelihood ratio method

Antithetic Variates (AV)

- Idea similar to CRN, but now only one system.
- **Goal:** estimate $\alpha = \mathcal{E}[X]$, where X is random output.
- **Approach:** generate n pairs of replications $(X_1, X_2),\ (X_3, X_4),\ \ldots,\ (X_{2n-1}, X_{2n})$
  - $X_1, X_2, \ldots, X_{2n}$ **identically distributed**
  - pairs are independent, but **dependence within each pair**
- Running n paired replications gives estimator

$$\alpha_{2n} \;=\; \frac{1}{2n} \sum_{i=1}^{2n} X_i$$

- Variance of α_{2n} is

$$\mathcal{V}[\alpha_{2n}] \;=\; \frac{1}{(2n)^2} \left(\sum_{i=1}^{2n} \mathcal{V}[X_i] \;+\; \sum_{k=1}^{n} 2\,\mathrm{Cov}[\,X_{2k-1}, X_{2k}\,] \right)$$

Antithetic Variates (AV)

- Recall: Variance of estimator α_{2n} is

$$\mathcal{V}[\alpha_{2n}] \;=\; \frac{1}{(2n)^2} \left(\sum_{i=1}^{2n} \mathcal{V}[X_i] \;+\; \sum_{k=1}^{n} 2\,\mathrm{Cov}[\,X_{2k-1}, X_{2k}\,] \right)$$

- Standard simulation: $X_1, \ldots, X_{2n}$ IID, so all Cov $= 0$.
- AV: within each pair (X_{2k-1}, X_{2k}), outputs generated dependently.
  - If all Cov < 0, then variance reduced.
  - "Opposite" of CRN, where we wanted positive correlation.
- **Intuition:**
  - For every replication with "favorable" random fluctuations, there is paired replication with "unfavorable" conditions.
  - Smooth out "good" and "bad" fluctuations in paired replications by taking average $Y_k = (X_{2k-1} + X_{2k})/2$.

Antithetic Variates (AV): One Approach

- Suppose generating output X requires m IID unif[0, 1] $U_1, \ldots, U_m$:

$$X \;=\; h(U_1, \ldots, U_m).$$

- Let's just consider first pair of replications (X_1, X_2).
- **Key Idea:** $U_i \sim$ unif[0, 1] implies $1 - U_i \sim$ unif[0, 1]
- Generate X_1 and X_2 using **same** $U_1, \ldots, U_m \sim$ unif[0, 1] IID:

$$\begin{aligned} X_1 &= h(U_1, \ldots, U_m), \\ X_2 &= h(1 - U_1, \ldots, 1 - U_m). \end{aligned}$$

- $X_1 \stackrel{\mathcal{D}}{=} X_2$, but X_1 and X_2 **dependent**
- Variance reduction when Cov$[X_1, X_2] < 0$
  - Cov$[X_1, X_2] < 0$ if h is monotone in each of its arguments.
  - May still have Cov < 0 if h not monotone, but no guarantee.

Antithetic Variates: Point Estimator

- General approach for applying AV to estimate $\alpha = \mathcal{E}[X]$:
  - Use AV to generate $(X_1, X_2),\ (X_3, X_4),\ \ldots,\ (X_{2n-1}, X_{2n})$ as IID pairs
  - Within kth pair, X_{2k-1} and X_{2k} are dependent from AV
  - AV estimator of α is

$$\alpha_{2n} \;=\; \frac{1}{2n} \sum_{i=1}^{2n} X_i$$

- How do we construct CI for α when using AV?
  - Need to show α_{2n} satisfies LLN and CLT.
  - **Key Idea:** express α_{2n} as sample average of IID RVs.

Antithetic Variates: LLN and CLT

- **Goal:** show AV estimator α_{2n} of $\alpha = \mathcal{E}[X]$ satisfies LLN and CLT.
- **Approach:** express α_{2n} as sample average of IID RVs.
  - Let $Y_k = \frac{1}{2}(X_{2k-1} + X_{2k})$ for each $k = 1, \ldots, n$.
  - Since $(X_1, X_2), (X_3, X_4), \ldots, (X_{2n-1}, X_{2n})$ are IID pairs, $Y_1, Y_2, \ldots, Y_n$ are IID, and

$$\alpha_{2n} \;=\; \frac{1}{2n}\sum_{i=1}^{2n} X_i \;=\; \frac{1}{n}\sum_{k=1}^{n} Y_k$$

  - $\mathcal{E}[Y_k] = \frac{1}{2}(\mathcal{E}[X_{2k-1}] + \mathcal{E}[X_{2k}]) = \alpha$.
  - **LLN**: If $\mathcal{E}[|Y_k|] < \infty$, then $\alpha_{2n} \approx \alpha$ for large n.
  - **CLT**: If $0 < \mathcal{V}[Y_k] < \infty$, then for large n,

$$\frac{\sqrt{n}}{\sqrt{\mathcal{V}[Y_k]}}[\alpha_{2n} - \alpha] \;\overset{\mathcal{D}}{\approx}\; \mathcal{N}(0,1).$$

Antithetic Variates (AV): Generic Algorithm

Algorithm: to construct CI for $\alpha = \mathcal{E}[X]$ when using AV.

1. For each $k = 1, 2, \ldots, n$,
   - Generate output pair (X_{2k-1}, X_{2k}) using AV, independently of all other pairs (X_{2j-1}, X_{2j}), $j \neq k$
     - X_{2k-1} and X_{2k} dependent since generated using AV.
   - Compute $Y_k = (X_{2k-1} + X_{2k})/2$.

2. Compute sample mean and sample variance of IID $Y_1, Y_2, \ldots, Y_n$:

$$\bar{Y}_n \;=\; \frac{1}{n}\sum_{k=1}^{n} Y_k \;=\; \frac{1}{2n}\sum_{i=1}^{2n} X_i, \qquad S_Y^2(n) \;=\; \frac{1}{n-1}\sum_{k=1}^{n} (Y_k - \bar{Y}_n)^2,$$

to get point estimate $\bar{Y}_n$ and 95% CI

$$\left[\; \bar{Y}_n \pm \frac{1.96\, S_Y(n)}{\sqrt{n}} \;\right]$$

for α.

Example: Antithetic Variates for G/G/1 Queue

- Consider one G/G/1 queue (as in CRN example)
  - Interarrival time CDF G
  - Service-time CDF H
- Want to estimate $\alpha = \mathcal{E}[X]$
  - X = average waiting time of first m customers.
- X_k = avg waiting time of first m customers in kth replication.
- Let's just focus on generating the first 2 replications (X_1, X_2).

Example: Antithetic Variates for G/G/1 Queue

- To generate (X_1, X_2),
  - Use one unif[0, 1] stream $(U_j : j \geq 1)$ to generate **interarrival times** for both replications 1 and 2.
  - Use another unif[0, 1] stream $(V_j : j \geq 1)$ to generate **service times** for both replications 1 and 2.
  - For replication 1,
    - ▷ generate j-th interarrival time and j-th service time as

$$G^{-1}(U_j) \quad \text{and} \quad H^{-1}(V_j)$$

  - For replication 2,
    - ▷ generate j-th interarrival time and j-th service time as

$$G^{-1}(1 - U_j) \quad \text{and} \quad H^{-1}(1 - V_j)$$

- Can show this will induce desired negative correlation in waiting times by using Lindley's eqn since G^{-1} and H^{-1} both m-increasing fcns.

Algorithm: AV for G/G/1 Queue

Goal: to construct CI for $\alpha = \mathcal{E}[X]$ when using AV

1. On each independent $k = 1, 2, \ldots, n$, do
   - Generate $U_{k,1}, U_{k,2}, \ldots, U_{k,m-1}$ and $V_{k,1}, V_{k,2}, \ldots, V_{k,m-1}$ as independent streams of IID unif[0, 1] random numbers.
   - Compute $Y_k = (X_k + X'_k)/2$, where

$$X_k = \frac{1}{m}\sum_{j=1}^{m} W_{k,j}, \qquad X'_k = \frac{1}{m}\sum_{j=1}^{m} W'_{k,j},$$

$W_{k,1} = W'_{k,1} = 0$; $W_{k,j} = (W_{k,j-1} + S_{k,j-1} - A_{k,j-1})^+$ and $W'_{k,j} = (W'_{k,j-1} + S'_{k,j-1} - A'_{k,j-1})^+$ for $2 \le j \le m$;

$$A_{k,j} = G^{-1}(U_{k,j}), \quad S_{k,j} = H^{-1}(V_{k,j}),$$
$$A'_{k,j} = G^{-1}(1 - U_{k,j}), \quad S'_{k,j} = H^{-1}(1 - V_{k,j}).$$

2. Compute 95% CI for α as $[\bar{Y}_n \pm 1.96 S_n/\sqrt{n}]$, where $\bar{Y}_n = \frac{1}{n}\Sigma_{k=1}^n Y_k$ and $S_n^2 = \frac{1}{n-1}\Sigma_{k=1}^n (Y_k - \bar{Y}_n)^2$.

Antithetic Variates

- Above idea works in wide variety of contexts.
- For G/G/1 queue, another way of inducing negative correlation:
  - Use $U_j \sim$ unif[0, 1] and inversion to generate
    - ▷ jth **service** time in replication 1
    - ▷ jth **interarrival** time for replication 2.
  - This works because
    - ▷ long interarrival time **decreases** delays
    - ▷ long service time **increases** delays.
- Monotonic response ensures antithetic variates works.
- **Warning:** CRN and AV can increase variance if not done carefully.

Example: Antithetic Variates for Knockout Option

European knockout (KO) call option

- T = expiration, which is m days
  - ΔT = one day; $T = m\,\Delta T$; $t_i = i\Delta T$ is end of ith day.
- K = strike price
- A = lower barrier, $B = +\infty$ (no upper barrier)
- S_t = price of underlying asset at time t; given S_0.
- σ = volatility
- r = risk-free rate
- Payoff: $I\{S_{t_i} > A,\ \forall\, i \le m\}\ (S_T - K)^+$
  - If $S_{t_i} > A\ \forall\ i = 1, 2, \ldots, m$, then payoff is $(S_T - K)^+$
  - Otherwise, payoff is 0.

Example: Antithetic Variates for Knockout Option

- Price of underlying asset follows GBM:

$$S_{t_i} \;=\; S_{t_{i-1}}\, e^{\gamma+\psi Z_i} \;=\; S_0 \prod_{j=1}^{i} e^{\gamma+\psi Z_j}$$

  - $Z_1, \ldots, Z_m \sim \mathcal{N}(0, 1)$ IID, so $\gamma + \psi Z_j \sim \mathcal{N}(\gamma, \psi^2)$
  - $\gamma = \left(r - \frac{\sigma^2}{2}\right)\Delta T$ (mean under risk-neutral valuation)
  - $\psi^2 = \sigma^2 \Delta T$ (variance)
- Discounted payoff of knockout option :

$$\begin{aligned} & e^{-rT}\, I\{S_{t_i} > A,\ \forall\, i \le m\}\ (S_T - K)^+ \\ &= e^{-rT}\, I\left\{S_0 \prod_{j=1}^{i} e^{\gamma+\psi Z_j} > A,\ \forall\, i \le m\right\} \left(S_0 \prod_{j=1}^{m} e^{\gamma+\psi Z_j} - K\right)^+ \\ &\equiv h(Z_1, \ldots, Z_m) \end{aligned}$$

- Want to estimate fair price $\alpha = \mathcal{E}[h(Z_1, \ldots, Z_m)]$ of KO option.

Example: Antithetic Variates for Knockout Option

- **Recall:** discounted payoff of knockout option

$$h(Z_1,\ldots,Z_m) = e^{-rT} I\left\{S_0 \prod_{j=1}^{i} e^{\gamma+\psi Z_j} > A,\ \forall\, i \le m\right\} \left(S_0 \prod_{j=1}^{m} e^{\gamma+\psi Z_j} - K\right)^+$$

where $Z_1,\ldots,Z_m \sim \mathcal{N}(0,1)$ IID.

- **Key Idea:** If $Z_j \sim \mathcal{N}(0,1)$, then $-Z_j \sim \mathcal{N}(0,1)$, so

$$h(Z_1,\ldots,Z_m) \stackrel{\mathcal{D}}{=} h(-Z_1,\ldots,-Z_m)$$

- **Claim:** h is monotone in each argument Z_ℓ
  - Suppose we increase Z_ℓ and keep all other Z_p fixed.
  - $\gamma + \psi Z_\ell$ increases since $\psi > 0$.
  - $S_{t_k} = S_0 \prod_{j=1}^{k} e^{\gamma+\psi Z_j}$ increases for each $k \ge \ell$.
  - $I\{\cdots\}$ and $(\cdots)^+$ increase monotonically, so h is monotone in Z_ℓ.

Algorithm: Antithetic Variates for Knockout Option

1. On each independent replication $k = 1, 2, \ldots, n$,
   - Generate $Z_{k,1}, Z_{k,2}, \ldots, Z_{k,m} \sim \mathcal{N}(0,1)$ IID.
   - Compute average of the 2 discounted payoffs :

$$Y_k = \frac{1}{2}\left[h(Z_{k,1},\ldots,Z_{k,m}) + h(-Z_{k,1},\ldots,-Z_{k,m})\right]$$

   where

$$h(z_1,\ldots,z_m) = e^{-rT} I\left\{S_0 \prod_{j=1}^{i} e^{\gamma+\psi z_j} > A, \forall\, i \le m\right\} \left(S_0 \prod_{j=1}^{m} e^{\gamma+\psi z_j} - K\right)^+$$

2. Compute 95% CI for price $\alpha = \mathcal{E}[h(Z_1,\ldots,Z_m)]$ of KO option as

$$[\ \bar{Y}_n \pm 1.96 S_Y/\sqrt{n}\ ],$$

where $\bar{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$ and $S_Y^2 = \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y}_n)^2$.

Control Variates

- **Goal:** to estimate $\alpha = \mathcal{E}[X]$
- Suppose $\exists$ output random variable (RV) Y such that
  - can generate Y in **same simulation** when generating output X
  - Mean $\nu = \mathcal{E}[Y]$ is **known**.
- **Example 1:** Queueing simulation
  - $X =$ average **waiting** time of first m customers
  - $Y =$ average **service** time of first $m-1$ customers
  - Since we know service-time distribution, we also know $\mathcal{E}[Y]$
- **Example 2:** Pricing European KO call option under GBM
  - $X =$ discounted payoff of **KO** call option
  - $Y =$ discounted payoff of **standard European** call option.
  - Black-Scholes formula gives $\mathcal{E}[Y]$.
- Can we use knowledge of $\mathcal{E}[Y]$ to get better estimate of $\alpha = \mathcal{E}[X]$?

Control Variates

- Recall:
  - Want to estimate $\alpha = \mathcal{E}[X]$
  - $\nu = \mathcal{E}[Y]$ is known.
- We call Y a "control variate" (CV)
- Define "controlled estimator"

$$X' = X - (Y - \nu)$$

- X' is unbiased estimator of α since

$$\begin{aligned}\mathcal{E}[X'] &= \mathcal{E}[X] - (\mathcal{E}[Y] - \nu)\\ &= \mathcal{E}[X]\\ &= \alpha\end{aligned}$$

Control Variates

- Variance of controlled estimator $X' = X - (Y - \nu)$ is
$$\mathcal{V}[X'] \;=\; \mathcal{V}[X] \,+\, \mathcal{V}[Y] \,-\, 2\,\mathsf{Cov}[X,Y]$$
- Thus, variance is reduced (relative to sampling just X) if
$$2\,\mathsf{Cov}[X,Y] \;>\; \mathcal{V}[Y].$$
- **Intuition:** Suppose X and Y are **positively correlated**.
  - $\mathsf{Cov}[X,Y] = \mathcal{E}[(X-\alpha)(Y-\nu)] > 0$.
  - If $Y > \mathcal{E}[Y] = \nu$, then typically $X > \mathcal{E}[X]$, and
$$X' \;=\; X - (Y-\nu) \;<\; X.$$
  - If $Y < \mathcal{E}[Y] = \nu$, then typically $X < \mathcal{E}[X]$, and
$$X' \;=\; X - (Y-\nu) \;>\; X.$$
  - If X, Y negatively correlated, let $X' = X + (Y - \nu)$.

Control Variates: General Approach

- More generally, consider controlled estimator
$$X(\lambda) \;=\; X \,-\, \lambda(Y-\nu)$$
for some constant λ.
  - Before, we implicitly assumed $\lambda = 1$.
- Since $\mathcal{E}[Y-\nu] = 0$,
$$\mathcal{E}[\,X(\lambda)\,] \;=\; \alpha \quad \text{for all } \lambda,$$
so $X(\lambda)$ is an unbiased estimator of α for all λ.
- Variance of estimator $X(\lambda)$ is
$$v(\lambda) \;=\; \mathcal{V}[X] \,+\, \lambda^2\mathcal{V}[Y] \,-\, 2\lambda\mathsf{Cov}[X,Y]$$
- Choose λ to minimize variance of $X(\lambda)$.

Control Variates: Choosing λ

- Recall: Variance of controlled estimator $X(\lambda) = X - \lambda(Y-\nu)$ is
$$v(\lambda) \;=\; \mathcal{V}[X] \,+\, \lambda^2\mathcal{V}[Y] \,-\, 2\lambda\mathsf{Cov}[X,Y]$$
- $v(\lambda)$ is a quadratic function in λ with minimum at
$$\lambda^* \;=\; \frac{\mathsf{Cov}[X,Y]}{\mathcal{V}[Y]} \;=\; \frac{\mathcal{E}[\,(X-\alpha)\,(Y-\nu)\,]}{\mathcal{V}[Y]}$$
  - λ^* has same sign as $\mathsf{Cov}[X,Y]$
- Minimizing variance is
$$v(\lambda^*) \;=\; (1-\rho^2)\,\mathcal{V}[X]$$
  - ρ is correlation between X and Y:
$$\rho \;=\; \frac{\mathsf{Cov}[X,Y]}{\sqrt{\mathcal{V}[X]\cdot\mathcal{V}[Y]}}$$
  - CV with λ_* reduces variance when $\rho \neq 0$.

Control Variates: Choosing λ

- Recall
$$\lambda^* = \frac{\mathsf{Cov}[X,Y]}{\mathcal{V}[Y]}, \qquad v(\lambda^*) \;=\; (1-\rho^2)\,\mathcal{V}[X].$$
- Significant variance reduction when $\rho \approx 1$ or $\rho \approx -1$
  - i.e., if X and Y have strong **linear** relationship
- λ^* unknown since $\mathsf{Cov}[X,Y] = \mathcal{E}[(X-\alpha)(Y-\nu)]$ unknown.
  - Estimate λ^* by
$$\lambda_n^* \;=\; \frac{\frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{\frac{1}{n-1}\sum_{k=1}^n (Y_k - \bar{Y}_n)^2}$$
where $(X_1,Y_1),\ldots,(X_n,Y_n)$ are IID replicates of (X,Y).
  - In Excel, functions `COVAR`, `CORREL` compute sample covariance, ρ.
- Compute sample mean and variance of $X'_1,\ldots,X'_n$, where $X'_k = X_k - \lambda_n^*(Y_k - \nu)$. (What is the slight issue with this?)

Control Variates: Point Estimator

- General approach to estimate $\alpha = \mathcal{E}[X]$ when using CV Y with (known) mean ν:
  - Generate (X_1, Y_1), (X_2, Y_2), $\ldots$, (X_n, Y_n) as IID pairs.
  - For each k, generate X_k and Y_k using **same** simulation, so they're dependent.
  - CV estimator of α is

$$\begin{aligned} \alpha_n &= \frac{1}{n}\sum_{k=1}^{n}[X_k - \lambda_n^*(Y_k - \nu)] \\ &= \bar{X}_n - \lambda_n^*(\bar{Y}_n - \nu), \end{aligned}$$

    where λ_n^* is defined on previous slide.
- How do we construct CI for α when using CV?
  - Need to show α_n satisfies LLN and CLT.
  - **Key Idea:** approximate α_n as sample average of IID RVs.

Control Variates: LLN and CLT

Approach: show $\alpha_n \approx$ sample average of IID RVs.

- First suppose λ^* is known.
- Let $X_k'' = X_k - \lambda^*(Y_k - \nu)$ for each $k = 1, \ldots, n$.
- Since $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are IID pairs,

$$\bar{X}_n'' = \frac{1}{n}\sum_{k=1}^{n} X_k'',$$

  is sample average of IID RVs $X_1'', X_2'', \ldots, X_n''$.
- $\mathcal{E}[X_k''] = \mathcal{E}[X_k] - \lambda^*(\mathcal{E}[Y_k] - \nu) = \alpha$.
- **LLN**: If $\mathcal{E}[|X_k''|] < \infty$, then $\bar{X}_n'' \approx \alpha$ for large n.
- **CLT**: If $0 < \mathcal{V}[X_k''] < \infty$, then for large n,

$$\frac{\sqrt{n}}{\sqrt{\mathcal{V}[X_k'']}}[\bar{X}_n'' - \alpha] \overset{\mathcal{D}}{\approx} \mathcal{N}(0,1).$$

Control Variates: LLN and CLT

- But $\lambda^* = \frac{\text{Cov}[X,Y]}{\mathcal{V}[Y]}$ is not known and is estimated by

$$\lambda_n^* = \frac{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{\frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y}_n)^2}.$$

- However, LLN implies for large n,

$$\begin{aligned} \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)(Y_k - \bar{Y}_n) &\approx \text{Cov}[X,Y], \\ \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y}_n)^2 &\approx \mathcal{V}[Y]. \end{aligned}$$

- Thus, $\lambda_n^* \approx \lambda^*$ for large n.
- So previous LLN and CLT still hold when λ_n^* replaces λ^*.

Control Variates: Generic Algorithm

Algorithm: to construct CI for $\alpha = \mathcal{E}[X]$ using CV Y with (known) mean ν.

1. For each independent replication $k = 1, 2, \ldots, n$,
   - Generate output pair (X_k, Y_k) from **same simulation**.
     - Since X_k and Y_k generated together, they are dependent.
2. After running all n IID reps, compute $\lambda_n^* = \frac{\sum_{k=1}^{n}(X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{\sum_{k=1}^{n}(Y_k - \bar{Y}_n)^2}$, where $\bar{X}_n = (1/n)\sum_{k=1}^{n} X_k$ and $\bar{Y}_n = (1/n)\sum_{k=1}^{n} Y_k$.
3. Let $X_k' = X_k - \lambda_n^*(Y_k - \nu)$ for all $k = 1, \ldots, n$.
4. Compute 95% CI for α as

$$\left[\bar{X}_n' \pm \frac{1.96 S_n'}{\sqrt{n}}\right],$$

   where $\bar{X}_n' = \frac{1}{n}\sum_{k=1}^{n} X_k'$ and $S_n'^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k' - \bar{X}_n')^2$.

Example: Control Variates for G/G/1 Queue

- **Goal:** to estimate expected average wait of first m customers.
- Suppose mean service time is known to be $\nu = 5$.
- For kth replication,
  - $A_{k,1}, \ldots, A_{k,m}$ are first m interarrival times
  - $S_{k,1}, \ldots, S_{k,m}$ are first m service times
  - $W_{k,1}, \ldots, W_{k,m}$ are first m waiting times, with $W_{k,1} = 0$ and $W_{k,j} = (W_{k,j-1} + S_{k,j-1} - A_{k,j})^+$ for $j \geq 2$
  - $X_k = (1/m) \sum_{j=1}^{m} W_{k,j}$ is average wait of first m customers
  - With **same** $S_{k,1}, \ldots, S_{k,m}$ used to compute $W_{k,j}$, compute CV

$$Y_k = \frac{1}{m-1} \sum_{j=1}^{m-1} S_{k,j}$$

- After running all n IID reps, compute kth controlled output

$$X'_k = X_k - \lambda_n^*(Y_k - 5)$$

Example: CV for Knockout Option under GBM

- When estimating fair price α of **knockout** call option under GBM, can use discounted payoff Y of **standard** Euro call option as CV:

$$\begin{aligned} Y &= e^{-rT}\left(S_0 \prod_{j=1}^{m} e^{X_j} - K\right)^+ \\ &\equiv g(X_1, \ldots, X_m) \end{aligned}$$

- By Black-Scholes formula,

$$\nu = \mathcal{E}[Y] = S_0\,\Phi(d_1) - K\,e^{-rT}\Phi(d_2)$$

where Φ is CDF of $\mathcal{N}(0,1)$,

$$\begin{aligned} d_1 &= \frac{\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}. \end{aligned}$$

Algorithm: CV for Knockout Option under GBM

1. On each independent replication $k = 1, 2, \ldots, n$,
   - Generate $X_{k,1}, \ldots, X_{k,m} \sim \mathcal{N}(\gamma, \psi^2)$ IID.
   - On kth replication, compute discounted payoffs of knockout option and standard European call option using **same** $(X_{k,1}, \ldots, X_{k,m})$:

$$W_k = h(X_{k,1}, \ldots, X_{k,m}), \qquad Y_k = g(X_{k,1}, \ldots, X_{k,m}),$$

   where

$$g(X_{k,1}, \ldots, X_{k,m}) = e^{-rT}\left(S_0 \prod_{j=1}^{m} e^{X_{k,j}} - K\right)^+,$$

$$h(X_{k,1}, \ldots, X_{k,m}) = g(X_{k,1}, \ldots, X_{k,m})\, I\left\{A < S_0 \prod_{j=1}^{i} e^{X_{k,j}} < B, \forall\, i \leq m\right\}.$$

2. After completing all n reps, let $W'_k = W_k - \lambda_n^*(Y_k - \nu)$ for $k = 1, \ldots, n$, where $\nu = S_0\,\Phi(d_1) - K\,e^{-rT}\Phi(d_2)$ with d_1 and d_2 on previous slide, and

$$\lambda_n^* = \frac{\sum_{k=1}^{n} (W_k - \bar{W}_n)(Y_k - \bar{Y}_n)}{\sum_{k=1}^{n} (Y_k - \bar{Y}_n)^2}.$$

3. Compute 95% CI for fair price α of knockout option as $[\bar{W}'_n \pm 1.96 S'_n/\sqrt{n}\,]$, where $\bar{W}'_n = \frac{1}{n}\sum_{k=1}^{n} W'_k$ and $S'^2_n = \frac{1}{n-1}\sum_{k=1}^{n}(W'_k - \bar{W}'_n)^2$.

Multiple Control Variates

- Previous idea can be extended to multiple control variates

$$Y(1), \ldots, Y(r)$$

with known $\mathcal{E}[Y(j)] = \nu_j$ for $j = 1, \ldots, r$.

- Controlled estimator is

$$X(\lambda_1, \ldots, \lambda_r) = X - \lambda_1[Y(1) - \nu_1] - \cdots - \lambda_r[Y(r) - \nu_r]$$

- Formulas for optimal $(\lambda_1^*, \ldots, \lambda_r^*)$ are now found by solving system of r linear equations.
  - Essentially fitting regression model
  - Simulating for "leftover" uncertainty
  - Ref: Ross, p. 168, or Glasserman, Section 4.1.2.

Importance Sampling: Example

- Suppose $X \sim \mathcal{N}(0,1)$, with density $f(x) = (1/\sqrt{2\pi})e^{-x^2/2}$.
- **Goal:** to estimate

$$\begin{aligned}\alpha &= \mathcal{P}_f(X \geq 3) = \int_{-\infty}^{\infty} I\{x \geq 3\}\, f(x)\, dx \\ &= \mathcal{E}_f\,[\, I\{X \geq 3\}\,]\end{aligned}$$

  - $\mathcal{P}_f$ and $\mathcal{E}_f$ are probability and expectation when X has density f.
- $\alpha = 0.00135$ (from z-table), but suppose we didn't know this.
- **Standard simulation algorithm** to estimate α
  1. Generate $X_1, X_2, \ldots, X_n \sim \mathcal{N}(0,1)$ IID
  2. Compute 95% CI $[\alpha_n \pm 1.96 S_n/\sqrt{n}]$ for α, where

$$\alpha_n = \frac{1}{n}\sum_{k=1}^{n} I\{X_k \geq 3\}, \quad S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} (I\{X_k \geq 3\} - \alpha_n)^2$$

Importance Sampling: Example

- Recall: $X \sim \mathcal{N}(0,1)$ and

$$\alpha = \mathcal{P}_f\{X \geq 3\} = \mathcal{E}_f\,[\, I\{X \geq 3\}\,] = 0.00135$$

  from z-table.
- Running $n = 10^4$ IID replications, we get 95% CI for α as

$$0.00170 \pm 47.5\%$$

- $A \equiv \{X \geq 3\}$ is a **rare event** since $\mathcal{P}_f\{X \geq 3\} \approx 0$.
  - $I\{X_k \geq 3\} = 0$ for most outputs X_k.
- **Basic idea of importance sampling:**
  Change distribution of X to increase probability of event A.

Importance Sampling: Example

- Let $g(x) = (1/\sqrt{2\pi})e^{-(x-3)^2/2}$ be density of $\mathcal{N}(3,1)$
- Observe that

$$\begin{aligned}\alpha &= \mathcal{E}_f\,[\, I\{X \geq 3\}\,] = \int_{-\infty}^{\infty} I\{x \geq 3\}\, f(x)\, dx \\ &= \int_{-\infty}^{\infty} I\{x \geq 3\}\, \frac{f(x)}{g(x)}\, g(x)\, dx \\ &= \int_{-\infty}^{\infty} I\{x \geq 3\}\, L(x)\, g(x)\, dx = \mathcal{E}_g\,[\, I\{X \geq 3\}\, L(X)\,]\end{aligned}$$

  - $L(x)$ is **likelihood ratio (LR)**:

$$L(x) = \frac{f(x)}{g(x)} = \frac{(1/\sqrt{2\pi})e^{-x^2/2}}{(1/\sqrt{2\pi})e^{-(x-3)^2/2}} = e^{(9-6x)/2}$$

- **Key idea of importance sampling**
  - Since $\alpha = \mathcal{E}_g[\, I\{X \geq 3\}\, L(X)\,]$, can estimate α by generating IID observations of $X \sim g$ and averaging $I\{X \geq 3\}\, L(X)$.

Importance Sampling: Example

- Recall: $\alpha = \mathcal{E}_g\,[\, I\{X \geq 3\}\, L(X)\,]$ with $L(X) = f(X)/g(X)$.
- **Importance sampling algorithm** to estimate α
  1. Generate IID outputs $X_1, \ldots, X_n$ using density g of $\mathcal{N}(3,1)$
  2. For each output X_k, compute LR

$$L(X_k) = \frac{f(X_k)}{g(X_k)} = e^{(9-6X_k)/2}$$

  3. Compute 95% CI $[\tilde{\alpha}_n \pm 1.96 \tilde{S}_n/\sqrt{n}]$ for α, with

$$\begin{aligned}\tilde{\alpha}_n &= \frac{1}{n}\sum_{k=1}^{n} I\{X_k \geq 3\}\, L(X_k), \\ \tilde{S}_n^2 &= \frac{1}{n-1}\sum_{k=1}^{n} [\, I\{X_k \geq 3\}\, L(X_k) - \tilde{\alpha}_n\,]^2.\end{aligned}$$

- In experiment with $n = 10^4$ IID replications, 95% CI for α is

$$0.00137 \pm 3.6\%$$

  - Std sim CI was $\pm 47.5\%$, so variance reduced by factor of 267.

Importance Sampling (IS): Single RV

- **Goal:** given some function h, we want to estimate

$$\alpha \;=\; \int h(x)\, f(x)\, dx \;=\; \mathcal{E}_f\,[\,h(X)\,]$$

  - f is density function of RV X, so $X \sim f$
  - $\mathcal{E}_f$ and $\mathcal{V}_f$ are expectation and variance when X has density f

- **Standard simulation algorithm:**

  1. Generate IID replicates $X_1, \ldots, X_n$ of $X \sim f$
  2. Construct 95% CI $[\alpha_n \pm 1.96 S_n/\sqrt{n}]$ for α from sample mean and sample variance of IID outputs $h(X_1), \ldots, h(X_n)$:

$$\alpha_n \;=\; \frac{1}{n}\sum_{k=1}^{n} h(X_k),$$

$$S_n^2 \;=\; \frac{1}{n-1}\sum_{k=1}^{n} [\; h(X_k) \;-\; \alpha_n \;]^2$$

Importance Sampling (IS): Change of Measure

- Consider another density g satisfying

$$g(x) \;>\; 0 \quad \text{whenever} \quad |h(x)|\, f(x) \;>\; 0$$

- Then apply **change of measure**:

$$\alpha \;=\; \mathcal{E}_f\,[\,h(X)\,] \;=\; \int h(x)\, f(x)\, dx \;=\; \int h(x)\, \frac{f(x)}{g(x)}\, g(x)\, dx$$

$$\;=\; \int h(x)\, L(x)\, g(x)\, dx \;=\; \mathcal{E}_g\,[\,h(X)\, L(X)\,]$$

  - $L(x) = f(x)/g(x)$ is **likelihood ratio** (LR)
  - $\mathcal{E}_g$ and $\mathcal{V}_g$ are expectation and variance when X has density g.

- Thus, $h(X)\, L(X)$ is an **unbiased** estimator of α under density g.
- Hence, we can estimate α by using density g to generate IID outputs of $X \sim g$ and averaging $h(X)\, L(X)$.

Importance Sampling

- General approach for applying IS to estimate

$$\alpha \;=\; \mathcal{E}_f[\,h(X)\,] \;=\; \mathcal{E}_g[\,h(X)\, L(X)\,]$$

  - Use g to generate n IID outputs $X_1, X_2, \ldots, X_n$.
  - In kth replication, compute LR as $L(X_k) = f(X_k)/g(X_k)$.
  - IS estimator of α is

$$\tilde{\alpha}_n \;=\; \frac{1}{n}\sum_{k=1}^{n} h(X_k)\, L(X_k)$$

- How do we construct CI for α when using IS?
  - Need to show $\tilde{\alpha}_n$ satisfies LLN and CLT.
  - **Key Idea:** use fact that $\tilde{\alpha}_n$ is sample average of IID RVs.

Importance Sampling: LLN and CLT

- **Goal:** show LLN and CLT for IS estimator $\tilde{\alpha}_n$ of

$$\alpha \;=\; \mathcal{E}_f[\,h(X)\,] \;=\; \mathcal{E}_g[\,h(X)\, L(X)\,]$$

- **Approach:** use fact that $\tilde{\alpha}_n$ is sample average of IID RVs.
  - IS estimator $\tilde{\alpha}_n$ is sample average of IID RVs $Z_k \equiv h(X_k)L(X_k)$, $k = 1, 2, \ldots, n$:

$$\tilde{\alpha}_n \;=\; \frac{1}{n}\sum_{k=1}^{n} h(X_k)L(X_k) \;=\; \frac{1}{n}\sum_{k=1}^{n} Z_k$$

  - $\mathcal{E}_g[h(X_k)L(X_k)] = \alpha$.
  - **LLN**: If $\mathcal{E}_g[|h(X_k)L(X_k)|] < \infty$, then $\tilde{\alpha}_n \approx \alpha$ for large n.
  - **CLT**: If $0 < \mathcal{V}_g[h(X_k)L(X_k)] < \infty$, then for large n,

$$\frac{\sqrt{n}}{\sqrt{\mathcal{V}_g[h(X_k)L(X_k)]}}[\tilde{\alpha}_n - \alpha] \;\overset{\mathcal{D}}{\approx}\; \mathcal{N}(0,1).$$

Importance Sampling (IS): Generic Algorithm

- Recall: $\alpha = \mathcal{E}_f[h(X)] = \mathcal{E}_g[h(X)L(X)]$
- **Importance sampling algorithm:** to estimate α
  1. Generate IID replicates $X_1, \ldots, X_n$ of $X \sim g$, where LR on kth replication is $L(X_k) = f(X_k)/g(X_k)$
  2. Compute sample mean and sample variance of IID outputs $h(X_k)L(X_k)$, $k = 1, \ldots, n$:

$$\tilde{\alpha}_n = \frac{1}{n}\sum_{k=1}^{n} h(X_k)L(X_k),$$

$$\tilde{S}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} [\,h(X_k)L(X_k) - \tilde{\alpha}_n\,]^2,$$

  so 95% CI for α is $[\,\tilde{\alpha}_n \pm 1.96\tilde{S}_n/\sqrt{n}\,]$.
- Choose "change of measure" g to get a variance reduction:

$$\mathcal{V}_g[h(X)L(X)] < \mathcal{V}_f[h(X)]$$

Importance Sampling: Choosing IS Density g

- Recall:

$$\alpha = \int h(x)f(x)\,dx = \mathcal{E}_f[h(X)] = \mathcal{E}_g[h(X)L(X)]$$

- Suppose $h(x) \geq 0$ for all x, and $0 < \alpha < \infty$.
- Define change of measure g^* as

$$g^*(x) = \frac{h(x)f(x)}{\alpha}$$

  - g^* is a density because $g^*(x) \geq 0$ and

$$\int g^*(x)\,dx = \frac{1}{\alpha}\int h(x)f(x)\,dx = \frac{1}{\alpha}\alpha = 1$$

- Using IS to generate X with density g^*, each output is

$$h(X)L(X) = h(X)\frac{f(X)}{g^*(X)} = h(X)\frac{f(X)}{h(X)f(X)/\alpha} = \alpha$$

  so estimator $\tilde{\alpha}_n$ has variance 0 !

Importance Sampling: Zero-Variance Estimator

- Recall:
  - Want to estimate $\alpha = \mathcal{E}_f[h(X)] = \mathcal{E}_g[h(X)L(X)]$
  - Zero-variance estimator: use optimal density

$$g^*(x) = \frac{h(x)f(x)}{\alpha}$$

- Unfortunately, g^* is not implementable
  - g^* requires knowledge of α, which we are trying to estimate.
- But g^* gives us some guidance on how to choose a "good" g
  - Since

$$\alpha = \int h(x)f(x)\,dx,$$

    we should choose g to give more weight to outputs x when $h(x)f(x)$ is large since those contribute more to value of α.
  - If $h(x) = I\{x \in A\}$, then g^* is conditional density given $X \in A$.

Importance Sampling: Function of Multiple RVs

- **Goal:** to estimate

$$\alpha = \mathcal{E}_p[h(X_1, X_2, \ldots, X_m)]$$

  where $X_1, X_2, \ldots, X_m$ are IID copies of a discrete RV X with
  - support S
  - PMF $p(\cdot)$
  - $\mathcal{E}_p$ is expectation when $X_1, X_2, \ldots, X_m \sim$ PMF p IID

Importance Sampling: Function of Multiple RVs

Standard algorithm to estimate $\alpha = \mathcal{E}_p[\,h(X_1,\ldots,X_m)\,]$

1. Using PMF p, generate n IID replicates
   $(X_{1,1},\ldots,X_{1,m}),\ (X_{2,1},\ldots,X_{2,m}),\ \ldots,\ (X_{n,1},\ldots,X_{n,m})$
   of $(X_1,\ldots,X_m)$.

2. For each replication $k = 1, 2, \ldots, n$, compute output
$$\alpha(k) \;=\; h(X_{k,1},\ldots,X_{k,m})$$

3. Construct 95% CI for α as $[\,\alpha_n \pm 1.96 S_n/\sqrt{n}\,]$ from sample avg and sample variance of IID outputs $\alpha(1),\ldots,\alpha(n)$:
$$\alpha_n \;=\; \frac{1}{n}\sum_{k=1}^{n}\alpha(k),$$
$$S_n^2 \;=\; \frac{1}{n-1}\sum_{k=1}^{n}[\,\alpha(k) - \alpha_n\,]^2$$

Importance Sampling: Change of Measure

- Consider another PMF $q(\,\cdot\,)$ with same support S as p and satisfying
$$q(x) > 0 \quad \text{whenever} \quad p(x) > 0$$
- **Claim:**
$$\alpha \;=\; \mathcal{E}_p[\,h(X_1,\ldots,X_m)\,] \;=\; \mathcal{E}_q[\,h(X_1,\ldots,X_m)\,L\,]$$
  - $\mathcal{E}_p$ is expectation when $X_1,\ldots,X_m$ IID with PMF p
  - $\mathcal{E}_q$ is expectation when $X_1,\ldots,X_m$ IID with PMF q
  - L is likelihood ratio (LR):
$$L \;\equiv\; L(X_1,\ldots,X_m) \;=\; \prod_{i=1}^{m}\frac{p(X_i)}{q(X_i)}$$

Importance Sampling: Change of Measure

Proof:

$$\begin{aligned}
\alpha &= \mathcal{E}_p[\,h(X_1,\ldots,X_m)\,] \\
&= \sum_{x_1\in S}\cdots\sum_{x_m\in S} h(x_1,\ldots,x_m)\prod_{i=1}^{m}p(x_i) \\
&= \sum_{x_1\in S}\cdots\sum_{x_m\in S} h(x_1,\ldots,x_m)\frac{\prod_{i=1}^{m}p(x_i)}{\prod_{i=1}^{m}q(x_i)}\prod_{i=1}^{m}q(x_i) \\
&= \mathcal{E}_q\left[\,h(X_1,\ldots,X_m)\,\frac{\prod_{i=1}^{m}p(X_i)}{\prod_{i=1}^{m}q(X_i)}\,\right] \\
&= \mathcal{E}_q[\,h(X_1,\ldots,X_m)\,L\,],
\end{aligned}$$

where L is likelihood ratio,
$$L \;=\; \prod_{i=1}^{m}\frac{p(X_i)}{q(X_i)}$$

Importance Sampling: Algorithm

Since $\alpha = \mathcal{E}_q[\,h(X_1,\ldots,X_m)\,L\,]$, we can estimate α as follows:

1. Using PMF q, generate n IID replicates
   $(X_{1,1},\ldots,X_{1,m}),\ (X_{2,1},\ldots,X_{2,m}),\ \ldots,\ (X_{n,1},\ldots,X_{n,m})$
   of $(X_1,\ldots,X_m)$.

2. For each replication $k = 1, 2, \ldots, n$, compute output
$$\tilde{\alpha}(k) \;=\; h(X_{k,1},\ldots,X_{k,m})\,L_k$$
   where $L_k = \prod_{i=1}^{m} p(X_{k,i})/q(X_{k,i})$ is LR on kth replication.

3. Construct 95% CI for α as $[\,\tilde{\alpha}_n \pm 1.96\tilde{S}_n/\sqrt{n}\,]$ from sample avg and sample variance of IID outputs $\tilde{\alpha}(1),\ldots,\tilde{\alpha}(n)$:
$$\tilde{\alpha}_n \;=\; \frac{1}{n}\sum_{k=1}^{n}\tilde{\alpha}(k),$$
$$\tilde{S}_n^2 \;=\; \frac{1}{n-1}\sum_{k=1}^{n}[\,\tilde{\alpha}(k) - \tilde{\alpha}_n\,]^2$$

Remarks on Importance Sampling (IS)

- IS applies much more generally.
- IS often used to estimate probability of **rare event**
  - For queueing problems, buffer overflows are rare.
  - For reliability problems, system failures are rare.
  - For option pricing, hitting distant barrier or strike is rare.
- Good choice for change of measure is **problem specific**.
  - For queueing and finance problems, use **exponential tilting**.
  - For reliability problems, use **failure biasing**.
- Choose change of measure to **make rare event occur in most likely way it would have occurred under original measure**.
- Poor choice of change of measure can yield larger (or ∞) variance.

Importance Sampling: Exponential Change of Measure

- In many queueing and finance problems, choose change of measure using **exponential tilting** or **exponential twisting**.
- Let X be RV with density function f.
- Define **moment generating function (MGF)** of $X \sim f$:

$$\mathcal{M}(\theta) \;=\; \mathcal{E}_f\left[e^{\theta X}\right] \;=\; \int_{-\infty}^{\infty} e^{\theta x} f(x)\, dx$$

  and let $\Theta = \{\theta : \mathcal{M}(\theta) < \infty\}$.
- **Exponentially tilted density**: for $\theta \in \Theta$,

$$g_\theta(x) \;=\; \frac{e^{\theta x} f(x)}{\mathcal{M}(\theta)} \;=\; e^{\theta x - \kappa(\theta)} f(x)$$

  - **Cumulant generating function (CGF)**

$$\kappa(\theta) \;=\; \ln \mathcal{M}(\theta)$$

Importance Sampling: Exponential Change of Measure

- Recall: **Exponentially tilted density**:

$$g_\theta(x) \;=\; \frac{e^{\theta x} f(x)}{\mathcal{M}(\theta)} \;=\; e^{\theta x - \kappa(\theta)} f(x)$$

  - MGF: $\mathcal{M}(\theta) \;=\; \mathcal{E}_f[e^{\theta X}] \;=\; \int_{-\infty}^{\infty} e^{\theta x} f(x)\, dx$
  - CGF: $\kappa(\theta) \;=\; \ln \mathcal{M}(\theta)$
- Can show: $\mathcal{E}_{g_\theta}[X] \;=\; \int_{-\infty}^{\infty} x\, g_\theta(x)\, dx \;=\; \kappa'(\theta) \;\equiv\; \frac{d}{d\theta}\kappa(\theta)$.
- Suppose $X_1, \ldots, X_m$ are IID with density f.
  - Change of measure: $X_1, \ldots, X_m$ IID with density g_θ
  - Likelihood ratio

$$\begin{aligned} L \;&=\; \prod_{i=1}^{m} \frac{f(X_i)}{g_\theta(X_i)} \;=\; \prod_{i=1}^{m} \frac{f(X_i)}{e^{\theta X_i - \kappa(\theta)} f(X_i)} \\ &=\; \prod_{i=1}^{m} e^{-\theta X_i + \kappa(\theta)} \;=\; \exp\left(-\theta \sum_{i=1}^{m} X_i + m\kappa(\theta)\right) \end{aligned}$$

Importance Sampling: Exponential Change of Measure

- **Example:** $X \sim \mathcal{N}(\mu, \sigma^2)$, density $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$
- Can show that MGF of $X \sim f$ is

$$\begin{aligned} \mathcal{M}(\theta) \;&=\; \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &=\; \exp\left(\mu\theta + \frac{\sigma^2\theta^2}{2}\right) \end{aligned}$$

- CGF: $\kappa(\theta) \;=\; \ln \mathcal{M}(\theta) \;=\; \mu\theta + \frac{\sigma^2\theta^2}{2}$
- Can show exponentially tilted $\mathcal{N}(\mu, \sigma^2)$ density:

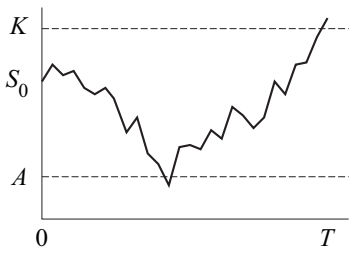
$$g_\theta(x) \;=\; e^{\theta x - \kappa(\theta)} f(x) \;\sim\; \mathcal{N}\left(\kappa'(\theta),\, \sigma^2\right)$$

  where $\kappa'(\theta) = \mu + \theta\sigma^2$.
  - Exponentially tilting a normal just changes its mean

Example: Importance Sampling for Knock-In Option

Down-and-in barrier (or knock-in) European call option

- T = expiration, which is m days
  - ΔT = one day; $T = m\,\Delta T$; $t_i = i\Delta T$ is end of ith day.
- K = strike price
- A = lower barrier, where $A \ll K$
- S_t = price of underlying asset at time t
- Given $S_0 \gg A$.
- σ = volatility
- r = risk-free rate



- Payoff: $I\{S_{t_i} \le A \text{ for some } i < m\}\ (S_T - K)^+$
  - If $S_{t_i} \le A$ for some $i = 1, 2, \ldots, m-1$, then payoff is $(S_T - K)^+$
  - Otherwise, payoff is 0.

Example: Importance Sampling for Knock-In Option

- Price of underlying asset follows geometric Brownian motion:

$$S_{t_i} = S_{t_{i-1}}\, e^{X_i}$$

  - $X_1, X_2, \ldots, X_m \sim \mathcal{N}(\gamma, \psi^2)$ IID
  - $\gamma = \left(r - \frac{\sigma^2}{2}\right)\Delta T$ (mean under risk-neutral valuation)
  - $\psi^2 = \sigma^2 \Delta T$ (variance)
- Price of underlying asset on day i:

$$S_{t_i} \;=\; S_0 \prod_{j=1}^{i} e^{X_j} \;=\; S_0 \exp\Bigl(\sum_{j=1}^{i} X_j\Bigr).$$

- **Positive payoff** when $S_{t_i} \le A$ on some day $i < m$ and $S_{t_m} > K$:

$$\sum_{j=1}^{i} X_j \;\le\; \ln(A/S_0) \;\equiv\; b \quad \text{and} \quad \sum_{j=1}^{m} X_j \;>\; \ln(K/S_0) \;\equiv\; c.$$

Example: Importance Sampling for Knock-In Option

- Discounted payoff of knock-in option is

$$\begin{aligned} & e^{-rT}\, I\{S_{t_i} \le A \text{ for some } i < m\}\ (S_T - K)^+ \\ & = e^{-rT}\, I\left\{S_0 \exp\Bigl(\sum_{j=1}^{i} X_j\Bigr) \le A \text{ for some } i < m\right\} \left(S_0 \exp\Bigl(\sum_{j=1}^{m} X_j\Bigr) - K\right)^+ \\ & \equiv\; h(X_1, \ldots, X_m) \end{aligned}$$

- Thus, fair price of knock-in option is

$$\begin{aligned} \alpha \;&=\; \mathcal{E}_f\,[\,h(X_1, \ldots, X_m)\,] \\ &=\; \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_m) \prod_{j=1}^{m} f(x_j)\, dx_j \end{aligned}$$

  - $\mathcal{E}_f$ is expectation when X_j are IID with density $f \sim \mathcal{N}(\gamma, \psi^2)$

Example: Importance Sampling for Knock-In Option

- Payoff: $I\{S_{t_i} \le A \text{ for some } i < m\}\ (S_T - K)^+$
- Barrier $A \ll S_0$ and strike $K \gg A$
  - For payoff > 0, asset price must
    - ▷ first decrease to hit barrier A
    - ▷ then end above strike K
  - **Positive payoff is rare event.**



- **Main idea of importance sampling**
  - Change means of RVs driving asset prices to make payoff > 0 more likely
    - ▷ Initially price has **negative drift** to drive it down to barrier A.
    - ▷ Then price has **positive drift** to drive it above strike K.

Example: Importance Sampling for Knock-In Option

- Asset price at time t_i is

$$S_{t_i} \;=\; S_0 \exp\Bigl(\sum_{j=1}^{i} X_j\Bigr)$$

- **Original measure:** X_j are IID with density $f \sim \mathcal{N}(\gamma, \psi^2)$
- **IS with exponential tilting:** X_j are independent with density

$$g_\theta(x) \;=\; e^{\theta x - \kappa(\theta)} f(x) \;\sim\; \mathcal{N}(\kappa'(\theta), \psi^2)$$

  - CGF $\kappa(\theta) = \gamma\theta + \frac{\psi^2\theta^2}{2}$
  - Mean $\kappa'(\theta) = \gamma + \theta\psi^2$
- Use different θ before and after hitting barrier A.
  - Before: use $\theta = \theta_1$ so that mean $\kappa'(\theta_1) < 0$.
  - After: use $\theta = \theta_2$ so that mean $\kappa'(\theta_2) > 0$.
- **Main issue:** choosing θ_1 and θ_2.

Example: Importance Sampling for Knock-In Option

- Let $\tilde{\tau} = \min\{i : S_{t_i} \le A\}$ = first day to go below barrier.
- Let $\tau = \min(\tilde{\tau}, m)$.
- For payoff > 0, need $\tau < m$ (asset hits barrier A before day m)
- Under importance sampling
  - $X_j \sim g_{\theta_1}$ for $j \le \tau$
  - $X_j \sim g_{\theta_2}$ for $j > \tau$
  - $\mathcal{E}_g$ is corresponding expectation
- Using change of measure, can express fair price of knock-in option as

$$\begin{aligned} \alpha &= \mathcal{E}_f\left[h(X_1, \ldots, X_m)\right] \\ &= \mathcal{E}_g\left[h(X_1, \ldots, X_m)\, L\right] \end{aligned}$$

  - Likelihood ratio

$$L \;=\; \prod_{j=1}^{\tau} \frac{f(X_j)}{g_{\theta_1}(X_j)} \prod_{j=\tau+1}^{m} \frac{f(X_j)}{g_{\theta_2}(X_j)}$$

Example: Importance Sampling for Knock-In Option

- Since

$$\frac{f(x)}{g_\theta(x)} \;=\; \frac{f(x)}{e^{\theta x - \kappa(\theta)} f(x)} \;=\; \exp\left\{-\theta x + \kappa(\theta)\right\},$$

likelihood ratio becomes

$$\begin{aligned} L \;&=\; \prod_{j=1}^{\tau} \frac{f(X_j)}{g_{\theta_1}(X_j)} \prod_{j=\tau+1}^{m} \frac{f(X_j)}{g_{\theta_2}(X_j)} \\ &=\; \exp\Bigl\{-\theta_1 \sum_{j=1}^{\tau} X_j + \tau\kappa(\theta_1)\Bigr\} \\ &\times\; \exp\Bigl\{-\theta_2 \sum_{j=\tau+1}^{m} X_j + (m-\tau)\,\kappa(\theta_2)\Bigr\} \end{aligned}$$

- Since $\sum_{j=\tau+1}^{m} = \sum_{j=1}^{m} - \sum_{j=1}^{\tau}$, we get (after some algebra)

$$L \;=\; \exp\Bigl\{(\theta_2-\theta_1)\sum_{j=1}^{\tau} X_j \;+\; \tau\left[\kappa(\theta_1)-\kappa(\theta_2)\right] \;-\; \theta_2 \sum_{j=1}^{m} X_j \;+\; m\kappa(\theta_2)\Bigr\}$$

Example: Importance Sampling for Knock-In Option

- Recall likelihood ratio

$$L = \exp\Bigl\{(\theta_2-\theta_1)\sum_{j=1}^{\tau} X_j \;+\; \tau\left[\kappa(\theta_1)-\kappa(\theta_2)\right] \;-\; \theta_2 \sum_{j=1}^{m} X_j \;+\; m\kappa(\theta_2)\Bigr\}$$

- To choose θ_1 and θ_2, need to understand how payoff > 0 typically occurs under original measure $\mathcal{P}_f$
  - i.e., when $S_{t_i} = S_0 \exp\left(\sum_{j=1}^{i} X_j\right)$ with $X_j \sim f = \mathcal{N}(\gamma, \psi^2)$
- Most of variability in L comes from barrier crossing time τ since when payoff > 0 occurs under $\mathcal{P}_f$,

$\sum_{j=1}^{\tau} X_j \;\approx\; b \equiv \ln(A/S_0)$ and $\sum_{j=1}^{m} X_j \;\approx\; c \equiv \ln(K/S_0)$.

- If we choose θ_1 and θ_2 so that $\kappa(\theta_1) = \kappa(\theta_2)$, then

$$L \;=\; \exp\Bigl\{(\theta_2-\theta_1)\sum_{j=1}^{\tau} X_j \;-\; \theta_2 \sum_{j=1}^{m} X_j \;+\; m\kappa(\theta_2)\Bigr\}$$

- How do we choose θ_1 and θ_2 with $\kappa(\theta_1) = \kappa(\theta_2)$?

Example: Importance Sampling for Knock-In Option

- Recall $b = \ln(A/S_0)$ and $c = \ln(K/S_0)$.
- Under original measure P_f,

$$\begin{aligned}
&\mathcal{P}_f\{\text{payoff} > 0\} \;=\; \mathcal{P}_f\{S_{t_i} \le A \text{ for some } i < m;\; S_T > K\} \\
&= \mathcal{P}_f\Big\{S_0 \exp\Big(\sum_{j=1}^{i} X_j\Big) \le A \text{ for some } i < m;\; S_0 \exp\Big(\sum_{j=1}^{m} X_j\Big) > K\Big\} \\
&= \mathcal{P}_f\Big\{\sum_{j=1}^{i} X_j \le b \text{ for some } i < m;\; \sum_{j=1}^{m} X_j > c\Big\} \qquad (1) \\
&\approx \mathcal{P}_f\Big\{\sum_{j=1}^{\tau} X_j \approx b;\; \sum_{j=1}^{m} X_j \approx c\Big\} \\
&\approx \mathcal{P}_f\Big\{\sum_{j=1}^{\tau} X_j \approx b;\; \sum_{j=\tau+1}^{m} X_j \approx c-b\Big\}
\end{aligned}$$

since most likely way (under $\mathcal{P}_f$) for event in eqn (1) to occur is when

$$\sum_{j=1}^{\tau} X_j \;\approx\; b \qquad \text{and} \qquad \sum_{j=1}^{m} X_j \;\approx\; c.$$

Example: Importance Sampling for Knock-In Option

- So under IS, choose θ so that means of $X_j \sim g_\theta$ lead to

$$\sum_{j=1}^{\tau} X_j \;\approx\; b \qquad \text{and} \qquad \sum_{j=\tau+1}^{m} X_j \;\approx\; c-b$$

where $b = \ln(A/S_0)$ and $c = \ln(K/S_0)$.

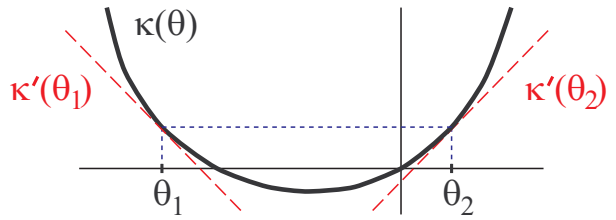
  - Recall mean of g_θ is $\kappa'(\theta) \equiv \frac{d}{d\theta}\kappa(\theta)$
  - $\kappa'(\theta_1)$ is mean of X_j up to τ
  - $\kappa'(\theta_2)$ is mean of X_j after τ, so

$$\sum_{j=1}^{\tau} X_j \;\approx\; \tau\,\kappa'(\theta_1) \qquad \text{and} \qquad \sum_{j=\tau+1}^{m} X_j \;\approx\; (m-\tau)\,\kappa'(\theta_2)$$

- Equating first and last displays leads to choosing θ_1 and θ_2 such that

$$\tau\,\kappa'(\theta_1) \;=\; b \qquad \text{and} \qquad (m-\tau)\,\kappa'(\theta_2) \;=\; c-b$$

Example: Importance Sampling for Knock-In Option



$$\kappa(\theta) \;=\; \gamma\theta + \frac{\psi^2\theta^2}{2} \qquad \text{and} \qquad \kappa'(\theta) \;=\; \gamma + \psi^2\theta$$

- Since $\kappa(\theta)$ is quadratic in θ
  - $\kappa(\theta)$ is symmetric about its minimum
  - we chose θ_1 and θ_2 such that $\kappa(\theta_1) = \kappa(\theta_2)$, so

$$\kappa'(\theta_1) \;=\; -\kappa'(\theta_2)$$

Example: Importance Sampling for Knock-In Option

- Recall
  1. $\tau\,\kappa'(\theta_1) \;=\; b$ and $(m-\tau)\,\kappa'(\theta_2) \;=\; c-b$
  2. $\kappa'(\theta_1) \;=\; -\kappa'(\theta_2)$
  3. $\kappa'(\theta) \;=\; \gamma + \psi^2\theta$
- Take difference of equations in item 1 and use item 2:

$$\kappa'(\theta_2) \;=\; \frac{c-2b}{m} \;=\; -\kappa'(\theta_1)$$

- Item 3 then yields

$$\theta_1 \;=\; -\frac{1}{\psi^2}\left(\frac{c-2b}{m} + \gamma\right), \qquad \theta_2 \;=\; \frac{1}{\psi^2}\left(\frac{c-2b}{m} - \gamma\right)$$

- Use $\kappa'(\theta_1)$ and $\kappa'(\theta_2)$, resp., as means of X_j up to and after τ.

Algorithm: Importance Sampling for Knock-In Option

1. On each independent replication $k = 1, 2, \ldots, n$,
   - Generate $(X_{k,1}, \ldots, X_{k,m})$, where
     - $X_{k,j} \sim \mathcal{N}\left(-\frac{c-2b}{m}, \psi^2\right)$ for $j \le \tau_k$
     - $X_{k,j} \sim \mathcal{N}\left(\frac{c-2b}{m}, \psi^2\right)$ for $j > \tau_k$
     - $\tilde{\tau}_k = \min\left(i : S_0 \prod_{j=1}^{i} e^{X_{k,j}} \le A\right)$, $\tau_k = \min(\tilde{\tau}_k, m)$
     - $b = \ln(A/S_0)$ and $c = \ln(K/S_0)$
   - Compute

$$\tilde{\alpha}(k) \;=\; e^{-rT} I\{\tau_k < m\} \left(S_0 \exp\left(\sum_{j=1}^{m} X_{k,j}\right) - K\right)^+ L_k$$

   where

$$L_k = \exp\left\{(\theta_2 - \theta_1) \sum_{j=1}^{\tau_k} X_{k,j} - \theta_2 \sum_{j=1}^{m} X_{k,j} + m\kappa(\theta_2)\right\},$$

$\theta_1 = \frac{-1}{\psi^2}\left(\frac{c-2b}{m} + \gamma\right)$, $\theta_2 = \frac{1}{\psi^2}\left(\frac{c-2b}{m} - \gamma\right)$ and $\kappa(\theta_2) = \gamma\theta_2 + \frac{\psi^2\theta_2^2}{2}$.

2. Compute sample mean $\tilde{\alpha}_n$ and sample variance $\tilde{S}_n^2$ of $\tilde{\alpha}(1), \ldots, \tilde{\alpha}(n)$, and construct 95% CI $[\,\tilde{\alpha}_n \pm 1.96\tilde{S}_n/\sqrt{n}\,]$ for fair price α of knock-in barrier option.

References for Importance Sampling

- General and queueing
  - Asmussen, S. and P. W. Glynn. 2007. *Stochastic Simulation*, Chapters V and VI, Springer.
  - Heidelberger, P. 1995. "Fast Simulation of Rare Events in Queueing and Reliability Models." *ACM Transactions on Modeling and Computer Simulation* 5, 1, 43–85.
- Reliability
  - Nicola, V. F., Shahabuddin, P., and Nakayama, M. K. 2001. "Techniques for fast simulation of highly dependable systems." *IEEE Transactions on Reliability* 50, 3, 246-264.
- Finance
  - Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*, Section 4.6, Springer.

Final Remarks on Variance-Reduction Techniques (VRTs)

- VRTs guaranteed to reduce variance compared to standard sim under certain conditions (e.g., CRN and AV if monotonicity holds)
  - But VRTs can increase variance if not applied properly (e.g., CRN, AV and IS)
- Can combine VRTs
  - e.g., combine antithetic variates and control variates
  - Not necessarily guaranteed to do better than using just one VRT.
- Other VRTs
  - Conditional Monte Carlo
  - Stratified sampling
  - Latin hypercube sampling