IE 525 - Numerical Methods in Finance

Monte Carlo simulation - variance reduction

Liming Feng

Dept. of Industrial & Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Variance reduction

- For i.i.d. $\{X_i, i \geq 1\}$ with $\mu = \mathbb{E}[X_1], \sigma^2 = \text{var}(X_1)$

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \sim \frac{\sigma}{\sqrt{n}} N(0,1) \quad \text{approximately}$$

- Monte Carlo methods converge at rate $1/\sqrt{n}$: need to quadruple sample size to halve standard error
- Might be easier to reduce σ^2 (and hence σ) using **variance reduction** techniques
- In European call example: using **antithetic variates**, standard deviation nearly halves (would need to quadruple sample size to achieve the same standard error in the standard approach)

I. Control variates

- Interested in $\mathbb{E}[Y]$: generate i.i.d. $\{Y_i \sim Y, i \geq 1\}$. Standard approach: estimate $\mathbb{E}[Y]$ with

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

- Suppose
  - X and Y are correlated
  - $\mathbb{E}[X]$ is known (and hence the error $\mathbb{E}[X] - X_i$ for any $X_i \sim X$)
  - Error of Y_i is *"proportional"* to error of X_i

$$\mathbb{E}[Y] - Y_i \approx b(\mathbb{E}[X] - X_i)$$

- Using X_i's as **control variates**: generate i.i.d. $\{(X_i, Y_i) \sim (X, Y),\ i \geq 1\}$. Let

$$Y_i^b = Y_i + b(\mathbb{E}[X] - X_i)$$

  and corresponding estimator

$$\bar{Y}_n^b = \frac{1}{n}\sum_{i=1}^{n}(Y_i + b(\mathbb{E}[X] - X_i))$$

- By law of large numbers,

$$\bar{Y}_n^b \to \mathbb{E}[Y] + b(\mathbb{E}[X] - \mathbb{E}[X]) = \mathbb{E}[Y]$$

- Variance of $\bar{Y}_n^b$ given by

$$\mathrm{var}(\bar{Y}_n^b) = \frac{1}{n}(\sigma_Y^2 + b^2\sigma_X^2 - 2b\sigma_X\sigma_Y\rho_{XY})$$

where $\sigma_Y^2 = \mathrm{var}(Y)$, $\sigma_X^2 = \mathrm{var}(X)$, $\rho_{XY} = \mathrm{corr}(X, Y)$
- In contrast, without control variates ($b = 0$)

$$\mathrm{var}(\bar{Y}_n) = \frac{1}{n}\sigma_Y^2$$

- Better estimation when

$$b^2\sigma_X^2 - 2b\sigma_X\sigma_Y\rho_{XY} < 0$$

- Best performance when (**optimally controlled**)

$$b = b^* = \frac{\mathrm{cov}(X, Y)}{\mathrm{var}(X)}$$

$$\mathrm{var}(\bar{Y}_n^{b^*}) = (1 - \rho_{XY}^2)\mathrm{var}(\bar{Y}_n)$$

- High correlation preferred:

$$\begin{aligned} \rho_{XY} &= 0.9487, \quad 1 - \rho_{XY}^2 = 10\% \\ \rho_{XY} &= 0.8660, \quad 1 - \rho_{XY}^2 = 25\% \\ \rho_{XY} &= 0.7071, \quad 1 - \rho_{XY}^2 = 50\% \end{aligned}$$

- 94.87% correlation leads to a standard error that can only be achieved by increasing the sample size by 10 times in the standard approach

- Usually, b^* has to be estimated (e.g., using $J = 1000$ pairs of (X_i, Y_i))

$$\hat{b}_J := \frac{\sum_{i=1}^{J}(X_i - \bar{X}_J)(Y_i - \bar{Y}_J)}{\sum_{i=1}^{J}(X_i - \bar{X}_J)^2}$$

- By law of large numbers, with probability one

$$\hat{b}_J \to b^* = \frac{\mathrm{cov}(X, Y)}{\mathrm{var}(X)}, \quad J \to +\infty$$

Example: pricing European options

- Pricing a European option with payoff $G(S_T)$

$$e^{-rT}\mathbb{E}[G(S_T)]$$

  For a European vanilla call, $G(S_T) = (S_T - K)^+$

- Simple control variates: **asset price itself**

$$\mathbb{E}[S_T] = S_0 e^{(r-q)T}$$

- Computing $\mathbb{E}[Y] := \mathbb{E}[G(S_T)]$ using control variates:

$$Y_i = G(S_T^i), \quad X_i = S_T^i, \quad Y_i^b = Y_i + b(S_0 e^{(r-q)T} - X_i)$$

Example: pricing Asian options

- Pricing a discrete Asian call option with maturity T, strike K, m monitoring times $t_j = j\delta, 1 \leq j \leq m, \delta = T/m$

$$V = e^{-rT}\mathbb{E}[(\bar{S}_T - K)^+], \quad \bar{S}_T = \frac{1}{m}\sum_{j=1}^{m} S_{t_j}$$

- Simulate n sample paths, for ith sample path, compute

$$Y_i = \left(\frac{1}{m}\sum_{j=1}^{m} S^i_{t_j} - K\right)^+$$

- Correlation between $\left(\bar{S}_T - K\right)^+$ and S_T not strong
- Better control variates: **geometric Asian call**

$$X = \left(\left(\prod_{j=1}^{m} S_{t_j} \right)^{1/m} - K \right)^+$$

- The geometric average is log-normally distributed in BSM

$$\left(\prod_{j=1}^{m} S_{t_j} \right)^{1/m} = S_0 \exp \left((r - q - \frac{1}{2}\sigma^2) \frac{1}{m} \sum_{j=1}^{m} t_j + \frac{\sigma}{m} \sum_{j=1}^{m} B_{t_j} \right)$$

- Since $t_j = j\delta$,

$$\frac{1}{m}\sum_{j=1}^{m} t_j = \frac{1}{2}(m+1)\delta = \frac{1}{2}(T+\delta) := \tilde{T}$$

- Since $\{B_t, t \geq 0\}$ is a standard BM

$$\frac{\sigma}{m}\sum_{j=1}^{m} B_{t_j} = \frac{\sigma}{m}\Big(B_{m\delta} - B_{(m-1)\delta} + 2(B_{(m-1)\delta} - B_{(m-2)\delta}) + \cdots + mB_{\delta}\Big)$$

is normal with mean 0 and variance

$$\frac{\sigma^2}{m^2}(1+4+\cdots m^2)\delta = \frac{\sigma^2}{6m^2}m(m+1)(2m+1)\delta = \frac{2m+1}{3m}\sigma^2\,\tilde{T}$$

- Define $\tilde{\sigma}$ such that

$$\tilde{\sigma}^2 = \frac{2m+1}{3m}\sigma^2$$

and denote

$$\tilde{q} = q + \frac{1}{2}\sigma^2 - \frac{1}{2}\tilde{\sigma}^2$$

then

$$\begin{aligned}\left(\prod_{j=1}^{m} S_{t_j}\right)^{1/m} &\sim S_0 \exp\left(\left(r - q - \frac{1}{2}\sigma^2\right)\tilde{T} + \tilde{\sigma} B_{\tilde{T}}\right) \\ &= S_0 \exp\left(\left(r - \tilde{q} - \frac{1}{2}\tilde{\sigma}^2\right)\tilde{T} + \tilde{\sigma} B_{\tilde{T}}\right) := \tilde{S}_{\tilde{T}}\end{aligned}$$

- For the geometric Asian call

$$\mathbb{E}[X] = e^{r\tilde{T}}\mathbb{E}[e^{-r\tilde{T}}(\tilde{S}_{\tilde{T}} - K)^+] = e^{r\tilde{T}} \times \text{BS call price}$$

where the Black-Scholes call price is obtained with *interest rate* r*, continuous yield* $\tilde{q}$*, volatility* $\tilde{\sigma}$*, strike* K*, maturity* $\tilde{T}$*, initial asset price* S_0

- **Control variate method for pricing discrete Asian call options**

$$Y_i^b = Y_i + b(\mathbb{E}[X] - X_i)$$

where $\mathbb{E}[X]$ is computed as above

Effectiveness

- Effectiveness depends on correlation ρ_{XY}, which may depend on parameters in the problem: e.g., for European call in the BSM ($r = 5\%, \sigma = 0.3, S_0 = 50, T = 0.25$), the lower K, the greater the correlation between $Y = (S_T - K)^+$ and $X = S_T$

K	40	50	60
ρ	99.5%	89.5%	60.4%
$1 - \rho^2$	1%	20%	64%

- For Asian call in the BSM ($r = 5\%$, $\sigma = 0.3$, $S_0 = 50$, $T = 0.25$, $K = 50$, $m = 13$): $Y = \left(\frac{1}{m}\sum_{j=1}^{m} S_{t_j} - K\right)^+$, $X = S_T$ or $(S_T - K)^+$ or $\left(\left(\prod_{j=1}^{m} S_{t_j}\right)^{1/m} - K\right)^+$

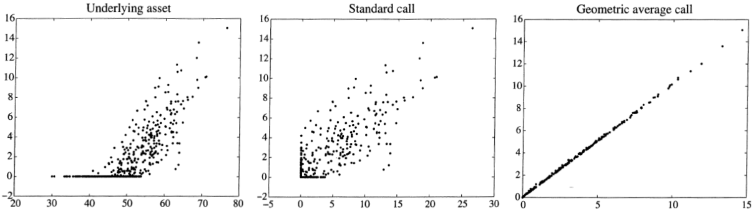


Figure: Correlations are 0.79, 0.85, > 0.99 respectively

- Using geometric Asian call as control is the most efficient

Example: pricing in non-BSM models

- Want to price a contract in the following model

$$dS_t = (r - q)S_t dt + \sigma_t S_t dB_t$$

  where σ_t is a function of t and/or S_t as in **local volatility** models or governed by a separate stochastic process as in **stochastic volatility** models

- Using Ito formula, $X_t = \ln(S_t)$ is governed by

$$dX_t = (r - q - \frac{1}{2}\sigma_t^2)dt + \sigma_t dB_t$$

- For $\delta = T/m, t_i = i\delta$, **Euler discretization** leads to

$$X_{t_{i+1}} - X_{t_i} = (r - q - \frac{1}{2}\sigma_{t_i}^2)\delta + \sigma_{t_i}(B_{t_{i+1}} - B_{t_i})$$

  or equivalently

$$S_{t_{i+1}} = S_{t_i} \exp\left((r - q - \frac{1}{2}\sigma_{t_i}^2)\delta + \sigma_{t_i}(B_{t_{i+1}} - B_{t_i})\right)$$

- Starting from S_0, $\{S_{t_1}, \cdots, S_{t_m}\}$ can be simulated by replacing $B_{t_{i+1}} - B_{t_i}$ by $\sqrt{\delta}Z_{i+1}$ for $Z_{i+1} \sim N(0,1)$

- For European call with strike K, one generates replicates for $Y = (S_{t_m} - K)^+$ as above
- For a control, starting from S_0, generate a sequence $\{\hat{S}_{t_1}, \cdots, \hat{S}_{t_m}\}$ as follows:

$$\hat{S}_{t_{i+1}} = \hat{S}_{t_i} \exp\left((r - q - \frac{1}{2}\sigma^2)\delta + \sigma\sqrt{\delta} Z_{i+1}\right)$$

  where σ is a constant and typical value of σ_t.
- Let $X = (\hat{S}_{t_m} - K)^+$. $\mathbb{E}[X]$ is known through Black-Scholes formula. X serves as the control

- Would like to compute

$$\mathbb{E}_f[h(X)] = \int_{\mathbb{R}^m} h(x)f(x)dx$$

  where X is a m-dimensional random vector with joint density $f(x)$

- Direct approach: simulate $\{X^i, i \geq 1\}$ from $f(x)$, estimate $\mathbb{E}_f[h(X)]$ by

$$\frac{1}{n}\sum_{i=1}^{n} h(X^i)$$

- For any density function $g(x)$ that satisfies $f(x)h(x) > 0 \Rightarrow g(x) > 0$

$$
\begin{aligned}
\mathbb{E}_f[h(X)] &= \int_{\mathbb{R}^m} h(x)f(x)dx \\
&= \int_{\mathbb{R}^m} \frac{h(x)f(x)}{g(x)} g(x)dx = \mathbb{E}_g\Big[\frac{h(X)f(X)}{g(X)}\Big]
\end{aligned}
$$

where $\mathbb{E}_g$ corresponds to the expectation when X follows a distribution with density $g(x)$

- Simulate $\{X^i, i \geq 1\}$ from $g(x)$, estimate $\mathbb{E}_f[h(X)]$ by

$$
\frac{1}{n}\sum_{i=1}^{n} \frac{h(X^i)f(X^i)}{g(X^i)}
$$

- Computing $\frac{h(X^i)f(X^i)}{g(X^i)}$ is more time consuming than $h(X^i)$
- New method could be attractive if simulating from g is sufficiently faster and/or

$$\text{var}_g\left(\frac{h(X)f(X)}{g(X)}\right) < \text{var}_f(h(X))$$

- Assuming positive h, if can find $g(x)$ that is nearly proportional to $h(x)f(x)$, then $\text{var}_g\left(\frac{h(X)f(X)}{g(X)}\right)$ could be much smaller
- $f(X^i)/g(X^i)$ - likelihood ratio: weight assigned to $h(X^i)$
- More weight to more important regions: if $h(x)$ is zero in a certain region, find g so that it's less likely to generate X^i's in that region

Example: asset-or-nothing call

- Consider an asset-or-nothing call option with payoff $S_T \mathbf{1}_{\{S_T > K\}}$ in model $S_t = S_0 e^{X_t}$, $X_T \sim f(x)$

$$\mathbb{E}_f[S_0 e^{X_T} \mathbf{1}_{\{X_T > \ln(K/S_0)\}}]$$

- For large $\ln(K/S_0)$, direct simulation returns many 0's and some large values, hence large variance; large sample size needed
- With importance sampling,

$$\mathbb{E}_g[S_0 e^{X_T} \mathbf{1}_{\{X_T > \ln(K/S_0)\}} \frac{f(X_T)}{g(X_T)}]$$

- Use a distribution $g(x)$ whose center is more to the right

- E.g., for normal f, one may use normal g with a larger mean
- In BSM, $X_T = (r - q - \frac{1}{2}\sigma^2)T + \sigma B_T \sim N(\mu, \sigma^2 T)$, $\mu = (r - q - \frac{1}{2}\sigma^2)T$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2 T}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2 T} \Big)$$

- Let $g(x)$ be the pdf of $N(\hat{\mu}, \sigma^2 T)$

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2 T}} \exp\Big(- \frac{(x-\hat{\mu})^2}{2\sigma^2 T} \Big)$$

- Likelihood ratio is given by

$$\frac{f(x)}{g(x)} = \exp\Big(- \frac{1}{2\sigma^2 T}(\mu^2 - \hat{\mu}^2 + 2(\hat{\mu} - \mu)x) \Big)$$

Exponential tilting

- The previous was a special case of exponential tilting
- Suppose the exponential moments of $f(x)$ exist:

$$M(\theta) = \int_{\mathbb{R}} e^{\theta x} f(x) dx < \infty, \quad \psi(\theta) = \ln(M(\theta))$$

  $\psi(\theta)$ is called the cumulant generating function
- Define $g(x) = e^{\theta x - \psi(\theta)} f(x)$. It's a density function. The corresponding likelihood ratio is

$$\frac{f(x)}{g(x)} = e^{-\theta x + \psi(\theta)}$$

  $\psi'(\theta)$ is the expectation of $g(x)$

- Exponential tilting for normal distributions: let $f(x)$ be the pdf of $N(\mu, \sigma^2)$

$$M(\theta) = \int_{\mathbb{R}} e^{\theta x} f(x) dx = \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2), \ \ \psi(\theta) = \mu\theta + \frac{1}{2}\sigma^2\theta^2$$

$$\begin{aligned} g(x) &= e^{\theta x - \psi(\theta)} f(x) \\ &= \exp\Big(\theta x - \mu\theta - \frac{1}{2}\sigma^2\theta^2\Big) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu-\sigma^2\theta)^2\Big) \end{aligned}$$

- Exponential tilting for a normal distribution shifts the mean by $\sigma^2\theta$

- Suppose we want to compute $E_f[h(X_1, \cdots, X_m)]$ for i.i.d. X_j's with density $f(x)$ for X_j. Using importance sampling, we generate X_j's from $g(x)$. The likelihood ratio is then

$$\frac{f(X_1)f(X_2)\cdots f(X_m)}{g(X_1)g(X_2)\cdots g(X_m)} = \Pi_{j=1}^m \frac{f(X_j)}{g(X_j)}$$

- Exponential tilting for i.i.d. $\{X_j, 1 \le j \le m\}$: the likelihood ratio is

$$\Pi_{j=1}^m \frac{f(X_j)}{g(X_j)} = \Pi_{j=1}^m \exp(-\theta X_j + \psi(\theta)) = \exp\left(-\theta \sum_{j=1}^m X_j + m\psi(\theta)\right)$$

Example: knock-in digital call

- Consider a down-and-in digital call option with lower barrier L ($L < S_0$) and payoff

$$\mathbf{1}_{\{S_T > K, \ \min_{0<k<m} S_{t_k} < L\}}$$

where $T = m\delta, t_k = k\delta$. Assume BSM with

$$S_{t_k} = S_{t_{k-1}} \exp(X_k), X_k \sim N((r - q - \frac{1}{2}\sigma^2)\delta, \sigma^2\delta)$$

- Want to compute

$$V = \mathbb{E}_f[\mathbf{1}_{\{S_T > K, \ \min_{0<k<m} S_{t_k} < L\}}]$$

- When $L << S_0$ and/or $S_0 << K$, direct simulation returns mainly 0's

- Let τ be the first time the asset price drops below L

$$\tau = \inf\{j : S_{t_j} < L\} = \inf\{j : X_1 + \cdots + X_j < \ln(L/S_0)\}$$

- Denote $b = -\ln(L/S_0)$, $c = \ln(K/S_0)$

$$V = \mathbb{E}_f[\mathbf{1}_{\{X_1+\cdots+X_m>c, \quad \tau<m\}}]$$

  f is the pdf of $N((r - q - \frac{1}{2}\sigma^2)\delta, \sigma^2\delta)$
- Use importance sampling and change the distribution of X_j's to make $\{X_1 + \cdots + X_m > c, \quad \tau < m\}$ more likely

- Denote $L_n = \sum_{j=1}^{n} X_j$. To have a non-zero payoff, need to drive L_n down toward $-b$ for knock-in and then up toward c for positive payoff
- Shift the distributions of $X_1, \cdots, X_\tau$ leftward using exponential tilting with parameter θ_- until the barrier is crossed at τ, then shift the distributions of $X_{\tau+1}, \cdots, X_m$ rightward using exponential tilting with parameter θ_+
- When $\tau < m$, the likelihood ratio is (make $\psi(\theta_+) = \psi(\theta_-)$)

$$\Pi_{j=1}^{\tau} \exp(-\theta_- X_j + \psi(\theta_-)) \; \Pi_{j=\tau+1}^{m} \exp(-\theta_+ X_j + \psi(\theta_+))$$

$$= \exp\Big(-\theta_- L_\tau + \tau\psi(\theta_-)\Big) \; \exp\Big(-\theta_+ (L_m - L_\tau) + (m-\tau)\psi(\theta_+)\Big)$$

$$= \exp\Big((\theta_+ - \theta_-) L_\tau - \theta_+ L_m + m\psi(\theta_+)\Big)$$

- In addition to making $\psi(\theta_+) = \psi(\theta_-)$, we also select $\theta_\pm$ so that

$$\frac{-b}{\psi'(\theta_-)} + \frac{c+b}{\psi'(\theta_+)} = m$$

- Note that $\psi'(\theta_-)$ is the mean of $X_1, \cdots, X_\tau$, $\psi'(\theta_+)$ is the mean of $X_{\tau+1}, \cdots, X_m$. Roughly, $L_n = X_1 + \cdots + X_n$ travels to $-b$ and then to c in m steps
- Note that $X_j \sim N(\mu, \sigma^2\delta)$, $\mu = (r - q - \frac{1}{2}\sigma^2)\delta$, $\psi(\theta) = \mu\theta + \frac{1}{2}\sigma^2\delta\theta^2$. The above two equations lead to

$$\theta_\pm = -\frac{\mu}{\sigma^2\delta} \pm \frac{2b+c}{\sigma^2 T}$$

Numerical performance

- With $r = 5\%, q = 0, \sigma = 0.15, S_0 = 95$, $T = 0.25, m = 50$, we obtain the following variance ratio (variance of the direct approach/variance of the importance sampling approach)

L	K	variance ratio
90	96	10
85	96	477
90	106	177

- Importance sampling significantly reduces variance for low L or high K

III. Stratified sampling

- Want to compute $\mathbb{E}[Y]$. **Direct approach**: simulate i.i.d. $\{Y_1, \cdots, Y_n\}$ and estimate $\mathbb{E}[Y]$ by $\bar{Y}_n$
- Let disjoint sets $\{A_1, \cdots, A_K\}$ be such that $\mathbb{P}(Y \in \cup_i A_i) = 1$. Denote $p_i = \mathbb{P}(Y \in A_i)$.

$$\mathbb{E}[Y] = \sum_{i=1}^{K} \mathbb{P}(Y \in A_i)\mathbb{E}[Y|Y \in A_i] = \sum_{i=1}^{K} p_i \mathbb{E}[Y|Y \in A_i]$$

- Simulate n_i i.i.d. replicates of Y conditional on $Y \in A_i$: $\{Y_{ij}, 1 \le j \le n_i\}$. Denote $n = n_1 + \cdots + n_K$, $q_i = n_i/n$. Estimate $\mathbb{E}[Y]$ by

$$\sum_{i=1}^{K} \left(p_i \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \right) = \frac{1}{n} \sum_{i=1}^{K} \left(\frac{p_i}{q_i} \sum_{j=1}^{n_i} Y_{ij} \right)$$

- Each A_i is a **stratum**, $\{A_1, \cdots, A_K\}$ is the set of all **strata**
- Stratified sampling eliminates sampling variability across strata
- In the direct approach, we don't know how many replicates will be from each stratum
- In stratified sampling, we predetermine the number of replicates from each stratum strategically to minimize variance

- More generally, for any X and $\{A_1, \cdots, A_K\}$ such that $\mathbb{P}(X \in \cup_i A_i) = 1$, $p_i = \mathbb{P}(X \in A_i)$,

$$\mathbb{E}[Y] = \sum_{i=1}^{K} p_i \mathbb{E}[Y | X \in A_i]$$

The **stratified sampling** approach simulates n_i i.i.d. replicates of (X, Y) conditional on $X \in A_i$ and estimates $\mathbb{E}[Y]$ by

$$\frac{1}{n} \sum_{i=1}^{K} \left(\frac{p_i}{q_i} \sum_{j=1}^{n_i} Y_{ij} \right)$$

Again, $n = n_1 + \cdots + n_K$, $q_i = n_i / n$.

Two issues

- 1. choosing the **stratification variable** X, the strata $\{A_1, \cdots, A_K\}$ and determining the allocation rule $n_1, \cdots, n_K$

  Goal is to make the variability of Y within each stratum small

- 2. simulating (X, Y) conditional on $X \in A_i$

- To construct a strata when X is continuous and has cdf F, first specify probabilities $p_1, \cdots, p_K > 0$ satisfying $\sum_i p_i = 1$, then compute the quantiles

$$a_1 = F^{-1}(p_1), a_2 = F^{-1}(p_1 + p_2), \cdots,$$

$$a_K = F^{-1}(p_1 + \cdots + p_K) = F^{-1}(1)$$

  a_K might be $+\infty$
- Define strata

$$A_1 = (-\infty, a_1], A_2 = (a_1, a_2], \cdots, A_K = (a_{K-1}, a_K]$$

- Denote the stratified sampling estimator by $\hat{Y}_n$:

$$\hat{Y}_n = \frac{1}{n}\sum_{i=1}^{K}\Big(\frac{p_i}{q_i}\sum_{j=1}^{n_i} Y_{ij}\Big)$$

- Denote $\sigma_i^2 = \mathrm{var}(Y_{ij}) = \mathrm{var}(Y|X \in A_i)$
- $\hat{Y}_n$ is unbiased estimator of $\mathbb{E}[Y]$, with variance

$$\begin{aligned}\mathrm{var}(\hat{Y}_n) &= \frac{1}{n^2}\sum_{i=1}^{K}\frac{p_i^2}{q_i^2} n_i \sigma_i^2 \\ &= \frac{1}{n}\sum_{i=1}^{K}\frac{p_i^2}{q_i}\sigma_i^2 = \frac{\sigma^2(q)}{n}\end{aligned}$$

where $\sigma(q)^2 = \sum_{i=1}^{K}\frac{p_i^2}{q_i}\sigma_i^2$

- Consider a proportional allocation with $q_i \approx p_i$

$$\sigma(q)^2 = \sum_{i=1}^{K} \frac{p_i^2}{q_i} \sigma_i^2 \approx \sum_{i=1}^{K} p_i \sigma_i^2$$

- Intuitively, $\sigma(p)^2$ is smaller than $\sigma^2 = \mathrm{var}(Y)$ as long as $\sigma_i^2 < \sigma^2$. Variance reduction should be effective
- Could be shown rigorously
- Can achieve better performance by selecting q_i's strategically

Why stratified sampling works

- Denote $\mu = \mathbb{E}[Y]$ and $\mu_i = \mathbb{E}[Y_{ij}] = \mathbb{E}[Y|X \in A_i]$

$$\mu = \sum_{i=1}^{K} p_i \mu_i$$

$$\mathbb{E}[Y^2] = \sum_{i=1}^{K} p_i \mathbb{E}[Y^2|X \in A_i] = \sum_{i=1}^{K} p_i(\sigma_i^2 + \mu_i^2)$$

- Then

$$\sigma^2 = \mathbb{E}[Y^2] - \mu^2 = \sum_{i=1}^{K} p_i \sigma_i^2 + \sum_{i=1}^{K} p_i \mu_i^2 - (\sum_{i=1}^{K} p_i \mu_i)^2 \geq \sum_{i=1}^{K} p_i \sigma_i^2$$

Variance decomposition

- The following decomposition

$$\sigma^2 = \sum_{i=1}^{K} p_i \sigma_i^2 + \Big(\sum_{i=1}^{K} p_i \mu_i^2 - (\sum_{i=1}^{K} p_i \mu_i)^2 \Big)$$

  shows that there are two parts in σ^2: the first part represents **variability within strata**, the second part represents **variability across strata**

- Stratified sampling effectively **removes variability across strata**
- **How to select strata**: use a strata such that variability across strata is large and variability within strata is small

Optimal allocation

- It was shown that **stratified sampling with a proportional allocation can only decrease variance**
- Optimize allocation rule for further variance reduction

$$\min_{q_i} \sum_{i=1}^{K} \frac{p_i^2}{q_i} \sigma_i^2$$

subject to $0 \leq q_i \leq 1, \sum_{i=1}^{K} q_i = 1$

- **Optimal allocation rule**

$$q_i^* = \frac{p_i \sigma_i}{\sum_{j=1}^{K} p_j \sigma_j}, \quad \sigma(q^*)^2 = (\sum_{i=1}^{K} p_i \sigma_i)^2$$

Computing optimal allocation

- Why is $\{q_i^*, 1 \leq i \leq K\}$ optimal? $\sigma(q)^2$ has the following lower bound

$$\sigma(q)^2 = \sum_{i=1}^{K} \frac{p_i^2}{q_i}\sigma_i^2 = \sum_{i=1}^{K} q_i \left(\frac{p_i}{q_i}\sigma_i\right)^2 \geq \left(\sum_{i=1}^{K} q_i \frac{p_i}{q_i}\sigma_i\right)^2 = (\sum_{i=1}^{K} p_i\sigma_i)^2$$

  and the lower bound is achieved with $q_i^* = \frac{p_i\sigma_i}{\sum_{j=1}^{K} p_j\sigma_j}$
- σ_i typically not known, but can be approximated by **pilot runs**: simulate a few Y conditional on $X \in A_i$ and estimate σ_i by sample standard deviation

- $\hat{Y}_n$ is approximately normal (since it is a linear combination of independent normals) with mean $\mu = \mathbb{E}[Y]$ and variance $\sigma(q)^2/n$

$$\hat{Y}_n = \frac{1}{n}\sum_{i=1}^{K}\left(\frac{p_i}{q_i}\sum_{j=1}^{n_i} Y_{ij}\right) = \sum_{i=1}^{K}\left(p_i\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}\right) \sim N(\mu, \frac{\sigma(q)^2}{n})$$

- $1-\alpha$ **confidence interval** is therefore

$$\hat{Y}_n \pm z_{\alpha/2}\frac{\sigma(q)}{\sqrt{n}}$$ standard error

where σ_i in $\sigma(q)^2 = \sum_{i=1}^{K}\frac{p_i^2}{q_i}\sigma_i^2$ is approximated by sample standard deviation of replicates in each stratum

Example: simulating a Poisson process

- Simulate **arrival times** of a Poisson process on $[0, T]$ with arrival rate λ
- **Inter-arrival time** has an exponential distribution with parameter λ
- Simulate $\tau_1, \tau_2, \cdots$ from $Exp(\lambda)$. Number of arrivals in $[0, T]$:

$$N = \max\{n : \tau_1 + \cdots + \tau_n \leq T\}$$

Arrival times are $T_1 = \tau_1,\ T_2 = \tau_1 + \tau_2,\ \cdots,$
$T_N = \tau_1 + \cdots + \tau_N$

With stratification

- Let N be the number of arrivals, N has a Poisson distribution with pmf

$$\mathbb{P}(N = k) = e^{-\lambda T} \frac{(\lambda T)^k}{k!}, \; k = 0, 1, \cdots$$

- We use **N as the stratification variable**. Given strata

$$A_1 = \{0, \cdots, m_1\}, A_2 = \{m_1 + 1, \cdots, m_2\}, \cdots,$$

$$A_K = \{m_{K-1} + 1, \cdots\}$$

Probabilities $p_k = \mathbb{P}(N \in A_k)$ can be computed

Simulating N

- **Simulating N conditional on $N \in A_k$** (that is $m_{k-1} < N \leq m_k$, with the convention that $m_K = +\infty$)

  Let U be uniform on $[0, 1]$,

$$V = F(m_{k-1}) + (F(m_k) - F(m_{k-1}))U = \sum_{i=1}^{k-1} p_i + p_k U.$$

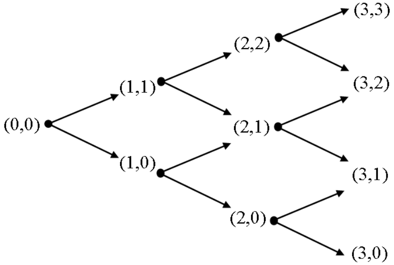
  Return $F^{-1}(V)$. Here F is the cdf of the Poisson distribution with parameter λT

Arrival times conditional on N

- Conditional on $N = m \in A_k$ (i.e., m arrivals on $[0, T]$), arrival times $\{T_1, \cdots, T_m\}$ has the same distribution as $\{U_{(1)}, \cdots, U_{(m)}\}$, where $U_1, \cdots, U_m$ are i.i.d. and uniform on $[0, T]$ and $\{U_{(1)}, \cdots, U_{(m)}\}$ is the corresponding ordered ascending sequence
- Therefore, one generates m replicates from the uniform distribution on $[0, 1]$, multiply them by T, and sort them in ascending order to get $\{T_1, \cdots, T_m\}$

Example: binomial tree

- Consider a $M-$step binomial tree on $[0, T = M\delta]$



- Node (m, j): time $m\delta$, j is the number of up moves in the stock price
- Stock price at node (m, j): $S_{m,j} = u^j d^{m-j} S_0$

CRR binomial tree

- Select u and d in the binomial model as follows (**Cox-Ross-Rubinstein binomial model**)

  $$u = e^{\sigma\sqrt{\delta}}, \quad d = e^{-\sigma\sqrt{\delta}}$$

  Denote the risk neutral probability of an up move by $p = \frac{e^{(r-q)\delta} - d}{u-d}$. Probability of a down move is $1 - p$
- As M gets large, the CRR model converges to the Black-Scholes-Merton model with initial stock price S_0, risk free interest rate r, continuous yield q, volatility σ

CRR binomial tree

- For a European style contract with payoff depending on S_T, use BSM directly
- For American vanilla options, use **backward induction** on a binomial tree. Easy to implement and still popular
- For path dependent contracts (Asian, lookback, etc.), direct implementation of the backward induction is computationally expensive, due to the need to differentiate paths
- **Simulating a binomial tree**: generate a uniform r.v. U on $[0, 1]$, multiply the stock price by u if $U \leq p$ and by d otherwise

With stratification

- Let N be the number of up moves in a path, N has a binomial distribution

$$\mathbb{P}(N = m) = \begin{pmatrix} M \\ m \end{pmatrix} p^m (1-p)^{M-m}, \;\; m = 0, \cdots, M$$

- Use **N as the stratification variable**, with strata

$$A_1 = \{0, \cdots, m_1\}, A_2 = \{m_1 + 1, \cdots, m_2\}, \cdots,$$

$$A_K = \{m_{K-1} + 1, \cdots, m_K = M\}$$

with probabilities $p_k = \mathbb{P}(N \in A_k)$

- Simulate N conditional on $N \in A_k$ similarly as in the Poisson case

Simulate a path conditional on N

- Conditional on $N = m \in A_k$ (m up moves, $M - m$ down moves)
- All paths that contain m up moves are equally likely. To generate such a path, at any time, generate a uniform r.v. U on $[0, 1]$, let the next move be an up move if

$$U \leq \frac{\text{number of remaining up moves}}{\text{number of remaining total moves}}$$

- Binomial tree provides an easy to implement method to price contracts such as Asian options

Variance reduction techniques

- Antithetic variates, control variates, importance sampling, stratified sampling, etc.
- Determine which method to use case by case, explore problem structures
- Effectiveness and complexity (roughly)