1 Motivation

1. **derivatives** financial assets that derives value from the performance of an underlying, e.x. asset, index, interest rate

2. **options** the buyer has the option to sell/buy an underlying asset; the seller has the obligation to fullfill such request. The buyer pays the seller a premium for this right

3. **Black–Scholes–Merton model** gives theoretical estimates of European-style options and shows that the option has a unique price regardless of the risk of the security and its expected return

2 Financial Theory

1. **Uncertainty** probability distribution of a random variable, i.e. price of stock, captures the uncertainty. Capture subjective uncertainty by updating the posterior. Binomial model are building blocks for discrete models; Wiener process with differential equation are buildingblocks for continuous models

2. **Financial assets**

   (a) **bond** firms/administration issues bond to fund their activities. Bond does not imply ownership of a firm. The buyer of bonds lend issuer some money. At maturity, issuer pay the bond owner an ammount called par value and periodically as coupons

   (b) **stocks** entitle owner to a share of the issuing firm. stocks are limited liability assets, i.e. buyer not responsible for undesirable conduct of the firm. Stocks havce no predefined maturity and entitle to owner to some stream of payments by nature of dividents, which are stochastic, i.e. depend on how the firm is faring.

   (c) **derivatives** financial contracts, dependent on value of some underlying variable, i.e. stock price, interest rate, index, non-financial asset.

4 Numerical Integration: Deterministic and Monte Carlo Methods

4.1 Determinstic Quadrature

1. **Goal** approximate the value of a definite integral

$$I(f) = \int_a^b f(x)dx$$

2. **Idea** Approximate the integral by a weighted sum of integrand evaluated at a set of integration points

$$Q(f) = \sum_j w_j f(x_j)$$

for $a = x_0 < x_1 < \cdots < x_N = b$

4.2 Monte Carlo Integration

1. **Idea** Cast determinstic evaluation of integral to evaluation of expected value of a random variable. To evaluate

$$I = \int_0^1 g(x)dx \qquad \longrightarrow \qquad I = \mathbb{E}\left\{g(U)\right\} \quad U \sim Unif(0,1)$$

We can estimate the expected value by generating a sequence $\{U\}_i$ independent random samples from uniform distribution and evaluate

$$\hat{I}_m = \frac{1}{m}\sum_{i=1}^m g(U_i)$$

where $\lim_{m\to\infty} \hat{I}_m = I$ with probability 1 by all of large numbers

2. **Experiment** Lowered variance for larger sample sizes. Would want to know how many samples needed in order to attain a given precision

4.3 Generating Pseudorandom Variates

1. **Linear Congruential Generators (LCG)** Deterministic sequence of samples, whose value depends on the value of the seed. Generate according to

$$Z_i = (aZ_{i-1} + c) \mod m$$

where m controls periodicity and Z_0 is the seed. The choice of parameters controls if the samples are **uniform** and **statistically independent**. LCG exhibit lattice structure that makes it suboptimal for a random number generator

2. **Accept-Reject Method**

   (a) **Goal** Want to generate random variates according to a density $f(x)$ but cannot invert f easily

   (b) **Assume** exists $t(x)$ such that

   $$t(x) \ge f(x) \quad \forall x \in I \qquad\qquad r(x) = \frac{1}{c}t(x) \quad c = \int_I t(x)dx$$

   in other words, $r(x)$ is a probability density computed by normalizing $t(x)$. We assume we can simulate $r(x)$ easily, i.e. generating samples that follows $r(x)$

   (c) **Procedures**

       i. Generate $Y \sim r$

       ii. Generate $U \sim U(0,1)$ independently of Y

       iii. If $U \le f(Y)/t(Y)$, then return $X = Y$; otherwise repeat the procedure

If I is bounded, then we can pick $r(x)$ as uniform, i.e.

$$t(x) = \max_{x \in I} f(x)$$

the Y variable is uniformly generated, For any Y, we compare values of $f(Y)$ and $t(Y)$

   (d) **Constraint** Usually requires support of f be bunded so we can use uniform as the upper bound function

3. **Computing Normal Random Variates** with polar rejection methods

4.4 Setting the Number of Replications

Definition. *q-quantile are values that partition a finite set of values into q subsets of (nearly) equal sizes. There are q - 1 of the q-quantiles, for each of $0 < k < q$. When cdf of a random variable is known, the q-quantiles are the application of quantile function (inverse function of the cdf) to the values $\{1/q, 2/q, \cdots, (q-1)/q\}$*

$$k - quantile = F^{-1}\left(\frac{k}{q}\right)$$

for all $k \in \{1, \cdots, q-1\}$

1. **Idea** In Monte Carlo simulation, the larger the number of samples, or replications, the better the quality of estimates. Idea is we compute sample mean/variance, which are unbiased estimators of true mean and true variance

$$\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}(n))^2$$

where we quantify **quality of the estimator** by expected values of squared error of estimate

$$\mathbb{E}\left\{(\overline{X}(n) - \mu)^2\right\} = var\left\{\overline{X}(n)\right\} = \frac{\sigma^2}{n}$$

Increasing sample size n improves the estimates, we compute a reasonable size n by giving out certain guarantees. We can control the absolute error / relative error, s.t. with probability $(1 - \alpha)$,

$$|\overline{X}(n) - \mu| \leq \beta \qquad \frac{|\overline{X}(n) - \mu|}{\mu} \leq \gamma$$

we can then derive the appropriate bound to achieve such guarantees

2. **Intuition** The rate of decrease of error as we increase sample size is about $O(1/\sqrt{n})$. We can overcome this problem by reducing the variance σ^2 of the samples, i.e. adopting quasi-Monte Carlo

4.5 Variance Reduction Techniques

1. **Antithetic sampling**

   (a) **Idea** is to generate a sequence of paired replications $(X_1^{(i)}, X_2^{(i)})$ for $i = 1, \cdots, n$, where the samples are horizontally independent, i.e. $X_j^{(i_1)}$ and $X_k^{(i_2)}$ are independent however we choose $j, k = 1, 2$ provided $i_1 \neq i_2$. And the samples are vertically negatively correlated, i.e. for a fixed i, $cov\{X_1^{(i)}, X_2^{(i)}\} < 0$. To induce negative correlation, we generate $\{U_k\}$ for the first replication in each pair, and then $\{1 - U_k\}$ in the second one. So the input streams are negatively correlated and hope that the output stream is negatively correlated. We now show that variance of samples generated with antithetic sampling is reduced

   $$
   \begin{aligned}
   var\left\{\overline{X}(n)\right\} &= \frac{var\left\{X^{(i)}\right\}}{n} \\
   &= \frac{var\left\{X_1^{(i)}\right\} + var\left\{X_2^{(i)}\right\} + 2cov\{X_1^{(i)}, X_2^{(i)}\}}{4n} \\
   &= \frac{var\left\{X\right\}}{2n}(1 + \rho(X_1, X_2))
   \end{aligned}
   $$

   (b) **Limitations** The sampling scheme sometimes produce positive correlation such that the variance increases. This happens when the integrand is non-monotonic. In this case, it maybe that even though the input streams are negatively correlated, but the output streams are positively correlated. For example, applying antithetic sampling to integrate e^x over $(0, 1)$ is Ok since exponential is monotonic

   (c) **Gaussian Samples** The input pair $Z_i \sim N(0, 1)$ and $-Z_i$ are negatively correlated

2. **Importance Sampling**

   (a) **Idea** The idea behind importance sampling is that certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these important values are emphasized by sampling more frequently, then the estimator variance can be reduced. It is often used to simulate rare events or sampling from the tail of a distribution. In order to sample from the citical region that is rare but important, we distort the probability measure while still obtain an unbiased estimate

   (b) **Method** Want to estimate

   $$
   \theta = \mathbb{E}_f\left\{h(\mathbf{X})\right\} = \int h(\mathbf{x})f(\mathbf{x})d\mathbf{x}
   $$

where $\mathbf{X}$ is a random vector with joint density $f(\mathbf{x})$. If we know another density g such that $f(\mathbf{x}) = 0$ whenver $g(\mathbf{x}) = 0$, then we can write

$$\theta = \int \frac{h(\mathbf{x})f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x})d\mathbf{x} = \mathbb{E}_g \left\{ \frac{h(\mathbf{X})f(\mathbf{X})}{g(\mathbf{X})} \right\}$$

a change of measure

(c) **Choice of biased distribution** g is important in variance reduction. We can write

$$\mathbb{E}_f \{h(\mathbf{X})\} = \mathbb{E}_g \{h^*(\mathbf{X})\} \qquad \text{where} \qquad h^*(\mathbf{X}) = \frac{h(\mathbf{X})f(\mathbf{X})}{g(\mathbf{X})}$$

assuming $g(\mathbf{x}) = 0$ if and only if $h(\mathbf{x}) = 0$, i.e. integrating only on the support. The two estimators have the same expectation, but their variances differ

$$var_f \{h(\mathbf{X})\} = \int h^2(\mathbf{x})f(\mathbf{x})d\mathbf{x} - \theta^2$$

$$var_g \{h^*(\mathbf{X})\} = \int h^2(\mathbf{x})\frac{f(\mathbf{x})}{g(\mathbf{x})}f(\mathbf{x})d\mathbf{x} - \theta^2$$

In general we want to find g^* such that

$$g^* = \arg\min_g \ var_g \{h^*(\mathbf{X})\}$$

The choice of $g(\mathbf{x}) = \frac{h(\mathbf{x})f(\mathbf{x})}{\theta}$ would make variance vanish

$$var_g \{h^*(\mathbf{X})\} = \theta \int h(\mathbf{x})f(\mathbf{x})d\mathbf{x} - \theta^2 = \theta\theta - \theta^2 = 0$$

but this requires knowledge of θ, the parameter we want to estimate in the first place. To ensure a reduction in variance, we select g such that

$$\begin{cases} g(\mathbf{x}) > f(\mathbf{x}) & \text{when the term } h^2(\mathbf{x})f(\mathbf{x}) \text{ is large} \\ g(\mathbf{x}) < f(\mathbf{x}) & \text{when the term } h^2(\mathbf{x})f(\mathbf{x}) \text{ is small} \end{cases}$$

Intuitively, pick $g(\mathbf{x})$ such that we reduce value of the integrand if its large to begin with; and to account for AUC for pdf is 1, pick $g(\mathbf{x})$ in other cases to increase value of integrand that is small to begin with

(d) **Estimating Pi** We can estimate π by the following formula

$$\pi = \int_0^1 \sqrt{1-x^2}dx = \frac{\pi}{4} = \int_0^1 h(x)dx = \mathbb{E}_f \{h\}$$

where $h(x) = \sqrt{1-x^2}$ and $f(x) = Unif(0,1)$. To get a better estimator with importance sampling, we pick $\tilde{g}(x)$ as an estimator of the optimal bias distribution

$$\tilde{g}(x) = \frac{h(x)f(x)}{\tilde{\theta}} = \frac{h(x)L}{\sum_k h(s_k)} \qquad \text{where} \qquad \tilde{\theta} = \frac{\sum_k h(s_k)}{L}$$

where we divide $[0,1]$ into L equally spaced subintervals, and $s_k = \frac{k-1}{L} + \frac{1}{2L}$, for $k = 1, \cdots, L$.

Intuitively, $x \sim g$ samples more frequently for large values of $h(x)$, i.e. values that have a bigger impact on the true value of the integral. The importance sampling in this case distort the original probability measure of
$$f(x) = Unif(0,1) \text{ to be } h(x)$$

Note $\tilde{g}(x)$ may not be a density integrating to 1 and sampling from $\tilde{g}(x)$ is not that simple. An alternative idea that deals with the problem is to define a discrete density $q(x)$ representing the probability of sampling from a subinterval and use a uniform density within each subinterval, i.e.

$$q_k = \frac{h(s_k)}{\sum_j^L h(s_j)} \qquad k = 1, \cdots, L$$

So we can define $g(x)$ to be a piecewise constant density over the L subintervals, where multiplication by L needed to make it a valid density

$$g(x) = Lq_k \qquad \frac{k-1}{L} \le x \le \frac{k}{L}$$

We can now estimate π with importance sampling

$$\pi = \int_0^1 h(x)dx = \mathbb{E}_{x \sim f}\{h(x)\} = \mathbb{E}_{x \sim g}\left\{\frac{h(x)f(x)}{g(x)}\right\}$$

The algorithm is as follows

i. Sample $k \sim q$, where k represent the k-th subinterval.

```
cs = cumsum(hvals);
k = sum(rand*cs(L) > cs) + 1;
```

ii. Sample x uniformly from k-th subinterval

```
x = (k - 1)/L + rand/L;
```

iii. Compute the estimate $\frac{h(x)f(x)}{g(x)} = \frac{\sqrt{1-x^2}\cdot 1}{Lq_k}$

```
q = hvals(k) / cs(L);
est(j) = sqrt(1 - x.^2) / (q*L);
```

iv. Repeated sampling and average the result

(e) **Estimating Rare Events** Importance sampling used when small probabilities involved. Suppose $\mathbf{X}$ has joint density f. We want to estimate

$$\theta = \mathbb{E}\left\{h(\mathbf{X})|\mathbf{X} \in \mathcal{A}\right\}$$

where $\{\mathbf{X} \in \mathcal{A}\}$ are rare events with a small unknown probability $p\{\mathbf{X} \in \mathcal{A}\}$. We want to reformulate the equation to get rid of the conditional term. The conditional density is given by

$$f(\mathbf{x}|\mathbf{X} \in \mathcal{A}) = \frac{f(\mathbf{x})}{p(\mathbf{X} \in \mathcal{A})}$$

for $\mathbf{x} \in \mathcal{A}$. Let $\mathbb{1}_{\mathbf{X} \in \mathcal{A}}(\mathbf{X})$ be an indicator representing if $\mathbf{X}$ is in $\mathcal{A}$. We can reformulate the problem by

$$\theta = \frac{\int_{\mathbf{x} \in \mathcal{A}} h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}}{p(\mathbf{X} \in \mathcal{A})} = \frac{\mathbb{E}\left\{h(\mathbf{X}) \mathbb{1}_{\mathcal{A}}(\mathbf{X})\right\}}{\mathbb{E}\left\{\mathbb{1}_{\mathcal{A}}(\mathbf{X})\right\}}$$

Now assume there is a density g such that the event is more likely under the corresponding probability measure, then we can generate samples $\mathbf{X}_i \sim g$ and estimate

$$\hat{\theta} = \frac{\sum_i h(\mathbf{X}_i) \mathbb{1}_{\mathcal{A}}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}}{\sum_i \mathbb{1}_{\mathcal{A}}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}}$$