Introduction to simulations and Monte Carlo methods
Project 2

Weronika Szota & Mateusz Szczepański

1. Description

The main goal of this project is to estimate following formula:

$$I = e^{-r}\mathbb{E}[A_n - K]_+, \tag{I}$$

using different approaches and variance reduction techniques. It is related to *options*, with discounted payoff at time 1 which can be interpreted as 1 year. In real we estimate only the part $\mathbb{E}[(A_n - K)_+]$ and add the term e^{-r} at the end.
We define A_n as:

$$A_n = \frac{1}{n}\sum_{i=1}^{n} S\left(\frac{i}{n}\right),$$

where $S(t)$ is Geometric Brownian Motion (with parameters μ, σ) given as:

$$S(t) = S(0) \cdot \exp\left[\left(r - \frac{\sigma^2}{2}\right)t + \sigma B(t)\right],$$

with $B(t)$ defined as standard Brownian Motion. We can rewrite it using $\mu^* = r - \sigma^2/2$, then we have:

$$S(t) = S(0) \cdot \exp\left[\mu^* t + \sigma B(t)\right],$$

In case $n = 1$ we know the exact value of I which is Black-Scholes formula given as:

$$\mathbb{E}[S(1) - K]_+ = S(0)\Phi(d_1) - Ke^{-r}\Phi(d_2), \tag{BS}$$

where Φ is cumulative distribution function of standard normal distribution and d_1 and d_2 are following:

$$d_1 = \frac{1}{\sigma}\left[\ln\left(\frac{S(0)}{K}\right) + r + \frac{\sigma^2}{2}\right], \qquad d_2 = d_1 - \sigma.$$

In this project we fix the constants: $r = 0.05, \sigma = 0.25, \mu^* = 0.01875, S(0) = 100, K = 100$.
With these parameters we get that the exact value of I is:

$$I = 12.335998930368717.$$

In all calculations we use Python's `numpy` package and `numpy.random.seed()` has been set to $465726236011 \bmod (2^{32} - 1)$ for every calling of estimator.

2. Estimating

2.1 Crude Monte Carlo estimator (European)

Crude Monte Carlo estimator is the simplest estimator one given as:

$$\hat{Y}_R^{cmce} = \frac{e^{-r}}{R} \sum_{i=1}^{R} Y_i.$$

We will think about Y_i as like $(A_n - K)_+$ and will just add term e^{-r} at the end. In case of CMC $A_n = (S(1) - K)_+$. The variance is given by:

$$Var[\hat{Y}_R^{cmce}] = \frac{e^{-2r}}{R} Var[Y_1],$$

theorethical value of this parameter can be computed as below:

$$\begin{aligned}
\frac{e^{-2r}}{R} Var[Y_1] &= \frac{e^{-2r}}{R} Var[(S(1) - K)_+] = \\
&= \frac{e^{-2r}}{R} \Big(\mathbb{E}[(S(1) - K)_+^2] - \mathbb{E}^2[(S(1) - K)_+] \Big) = \\
&= \frac{e^{-2r}}{R} \Big(\mathbb{E}[(S(1) - K)_+^2] - (I \cdot e^r)^2 \Big) = \\
&= \frac{e^{-2r}}{R} \Big(\int_{\mathbb{R}} \Big(S(0) e^{\mu^* + \sigma x} - K \Big)_+^2 \cdot e^{\frac{-x^2}{2}} \, dx - (I \cdot e^r)^2 \Big) = \\
&= \frac{e^{-2r}}{R} \Big(\int_{\frac{\sigma}{2} - \frac{r}{\sigma}}^{\infty} \Big(S(0) e^{\mu^* + \sigma x} - K \Big)^2 \cdot e^{\frac{-x^2}{2}} \, dx - (I \cdot e^r)^2 \Big) = \\
&= \frac{e^{-2r}}{R} \Big(\int_{-0.075}^{\infty} \Big(100 e^{0.01875 + 0.25x} - 100 \Big)^2 \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \, dx - (I \cdot e^r)^2 \Big) = \\
&\approx \frac{e^{-2r}}{R} \Big(546.681 - 168.181 \Big) = \frac{342.481}{R}
\end{aligned}$$

We used exact value of (I) to compute this parameter, we could use $\hat{I}_R^{cmc}$ in case if we would not know the exact value of (I).
In our case the estimator with $n = 1$ takes the form:

$$\hat{I}_R^{cmce} = \frac{e^{-r}}{R} \sum_{i=1}^{R} (S_i(1) - K)_+ = \frac{e^{-r}}{R} \sum_{i=1}^{R} \max(S(0) \cdot \exp\left[\mu^* + \sigma B_i(1)\right] - K, 0), \tag{CMC\_eu}$$

Let's consider $R \in \{100, 10\,000, 100\,000\}$. Results are following:

R	estimator	relative error	theoretical variance
100	15.4827	25.5079%	3.42481
10000	12.3277	0.0675%	0.0342481
1000000	12.3377	0.0139%	0.000342481

Firstly we can notice that estimation for $R = 100$ is not quite good. Relative error is high and definitely this is too small number of observations to estimate our *option* result. For $R = 10^4$ final score is much better, difference is on the 2nd decimal place. It's even better for $R = 10^6$ because we have the difference on the 3rd decimal place. CMC seems very good in computations if we only consider R bigger or equal than $\sim 10^4$.

2.2 Crude Monte Carlo estimator (Asian)

Now we consider Crude Monte Carlo but in general for $n \in \mathbb{N}_+$.
In computations we will consider $n \in \{2, 3, 17\}$ and again $R \in \{100, 10\,000, 100\,000\}$.
Our estimator is now:

$$\hat{I}^{cmca}_{R,n} = \frac{e^{-r}}{R} \sum_{i=1}^{R} \max(A_n - K, 0), \tag{CMC\_as}$$

where $A_n = \frac{1}{n} \sum_{i=1}^{n} S(\frac{i}{n}) = \frac{1}{n} \sum_{i=1}^{n} S(0) \cdot \exp\left[\mu^* \frac{i}{n} + \sigma B(\frac{i}{n})\right]$.
Table shows actual results of Asian CMC estimator:

n	R	estimator
2	100	8.0902
	10000	7.2314
	1000000	7.1921
3	100	7.0333
	10000	5.5454
	1000000	5.5738
17	100	2.7803
	10000	2.6738
	1000000	2.6816

We see that for bigger n our estimator decreases. In case $n = 1$ we had value about 12.33 although now we observe values like ~ 8 or even ~ 2.7. The key thing here is that bigger R does not lead to decrease the estimator. It's clearly visible in table where $\hat{I}^{cmca}_{10^4,17} > \hat{I}^{cmca}_{10^6,17}$. It's hard to determine which approximation is better because we don't have method to compute analitical value for $n \geq 2$. But for example we can notice that if values oscilate near 2.7 in $n = 17$ we can expect this is clearly good estimation.

2.3 Stratified estimator

Let $A^1, A^2, ..., A^m$ be disjoint sets, such that $P(Y \in \bigcup_{j=1}^{m} A^j) = 1$. This split defines strata $S^j = \{\omega : Y(\omega)^j\}$. Denote:

- $p_j = P(Y^j)$ a probablity of choosing j-th stratum
- $I^j = E[Y|Y \in A^j]$, estimator of I in j-th stratum $j = 1, 2, ..., m$

From the formula for the total probability we have:

$$I = p_1 I^1 + p_2 I^2 + ... + p_m I^m$$

We split number of total replications R into R_j replications Y^j in j-th stratum. Let $Y^j_1, Y^j_2, ..., Y^j_{R_j}$ be i.i.d. replications of Y^j in the j-th stratum. We also assume independence between replications within indyvidual strata. Of course $R_1 + R_2 + ... + R_m = R$. We define

$$\hat{Y}^j_{R_j} = \frac{1}{R_j} \sum_{i=1}^{R_j} Y^j_i$$

as an estimator of I^j. The stratfied estimator is defined as below

$$\hat{Y}^{str}_R = p_1 \hat{Y}^1_{R_1} + \hat{Y}^2_{R_2} ... + \hat{Y}^m_{R_m} \tag{STR}$$

We will consider two options regarding choosing number of replications in each strata:

- Proportional allocation - take $R_i = p_i R, i = 1, 2, ..., m$
- Optimal allocation - take $R_i = \frac{p_i \tilde{s}_i}{\sum_{k=1}^{m} p_k \tilde{s}_k} R$

Using this method variance of Y can be decomposed into within- and between-stratum components and is minimal for R_j chosen with formula given above (optimal allocation).
In order to perfom optimal allocation method firstly, we have to estimate variances $\tilde{s}^2_j$ with pilot R' simulations using proportional allocation $R_j = p_j R'$. For cases where dimension $n \geq 2$ we will apply stratified sampling of a Brownian motion which is based on stratified sampling of multivariate normal $N(0, \Sigma)$ variable.

2.3.1 Stratified estimator (European)

Let us consider stratified estimator for $n = 1$. Table below shows results obtained for some specific parameters R and m with proportional allocation. Recall that the exact value is 12.335998930368717.

m	R	estimator	relative error
5	10000	12.4329	0.7857%
10	10000	12.3091	0.2175%
5	1000000	12.3252	0.0871%
10	1000000	12.3308	0.0415%
50	1000000	12.3373	0.0110%

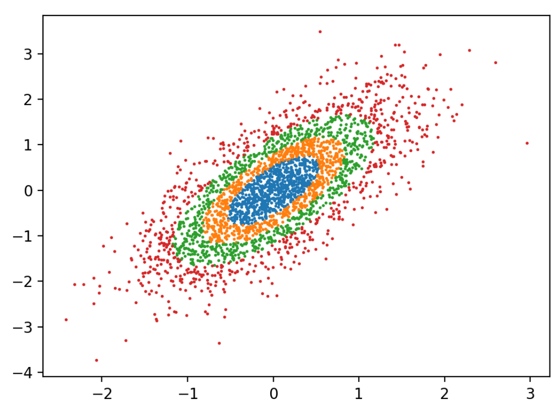
We obtain the best results for $m = 50$ and $R = 1000000$, but all the results are very close to the exact value. We can see that values of estimator vary depending on number of stratum.
Let us also examine results obtained with optimal allocation, R' pilot replications, m stratum and R replications.

m	R	R'	estimator	relative error
5	10000	100	12.4221	0.6984%
10	10000	100	12.3150	0.1702%
5	10000	1000	12.3488	0.1038%
10	10000	1000	12.2977	0.3098%
5	1000000	100	12.3238	0.0985%
10	1000000	100	12.3345	0.0116%
50	1000000	100	12.3334	0.0202%
5	1000000	1000	12.3443	0.0676%
10	1000000	1000	12.3398	0.0312%
50	1000000	1000	12.3364	0.0038%

We can notice that all the values are close to the exact one. There are some differences depending on m, R, R'. We can see that for $m = 5$ and 10, $R = 10000$ results obtained with optimal allocation $R' = 100$ are better than ones obtained with proportional allocation. But it is not the rule (for optimal allocation with $R' = 1000$ results are not better). We have to keep in mind that those results are obtained for some specific parameters and seed.

2.3.2 Stratified estimator (Asian)

Let us examine stratified estimator for $n \geq 2$. However, we do not have the exact value with which we can compare obtained results. It is worth seeing, how stratified sampling from normal distribution with specified covariance matrix Σ with difeerent allocations schemes looks like. This is shown in a figure below for dimension $n = 2$ and $n = 3$ using proportional allocation.



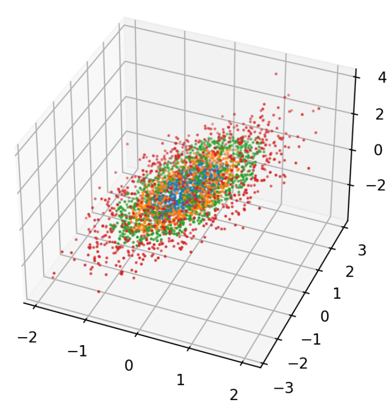


Figure 1: 3000 points from 2-dimensional (left) and 3-dimensional (right) normal distribution with covariance matrix $\Sigma = min(i, j)/n$. Each strata represented in a different colour.

Now let us have a look at value of stratified estimator using proportional allocation. We will examine results for 3 different dimensions and different parameters m, R.

n	m	R	estimator
2	4	10000	9.4495
3	4	10000	8.7109
17	4	10000	7.0560
2	10	10000	9.5430
3	10	10000	8.9687
17	10	10000	7.1969

n	m	R	estimator
2	4	1000000	9.5892
3	4	1000000	8.6857
17	4	1000000	7.1783
2	10	1000000	9.5970
3	10	1000000	8.6850
17	10	1000000	7.1608

Let us see how results of stratified sampling from normal distribution with specified covariance matrix Σ with optimal allocation scheme look like. This is shown in a figure below for dimension $n = 2$ and $n = 3$.

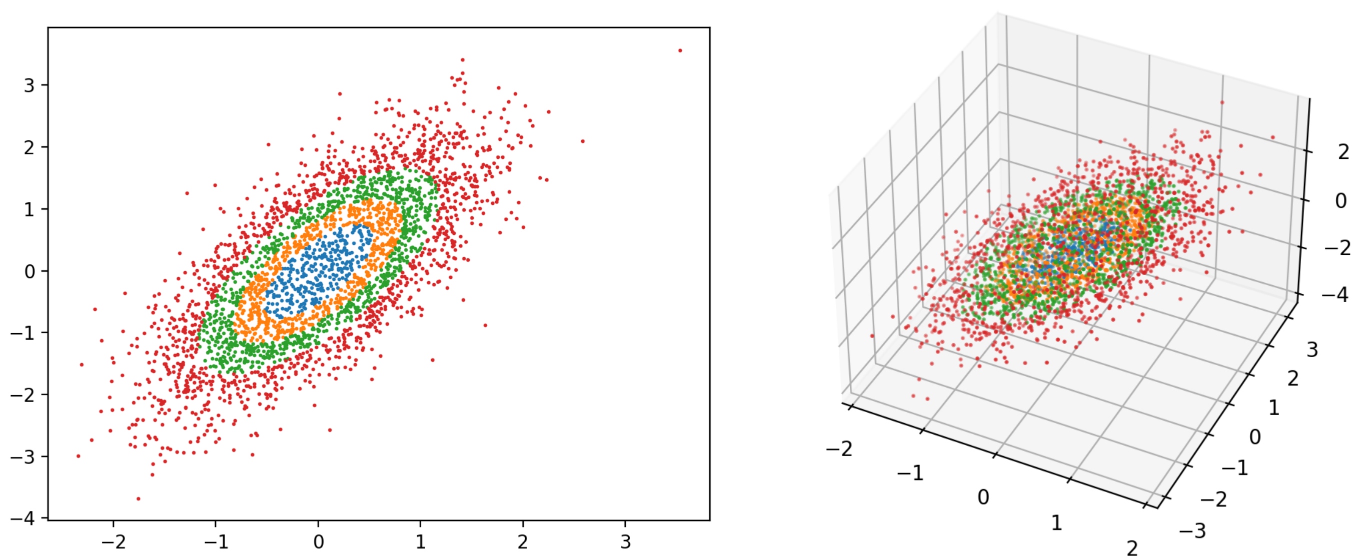


Figure 2: 3000 points from 2-dimensional (left) and 3-dimensional (right) normal distribution with covariance matrix $\Sigma = min(i, j)/n$. Each strata represented in a different colour. Using optimal allocation with pilot $R' = 100$ replications.

We can see a difference in comparison to previous figure, for example it is clearly visible that number of replications in blue strata is smaller than in the previous case. That is because variances obtained with R' pilot replications in our case are getting bigger: $[0.3386658, 0.66316108, 0.93791292, 1.36581994]$, so we have the smallest number of replications in the first strata (the blue one) and the biggest number of replications in the last strata (the red one).

Now let us have a look at values of the estimator obtained for different parameters R, R', m.

n	m	R	R'	estimator
2	4	10000	100	9.6251
3	4	10000	100	8.7878
17	4	10000	100	7.1576
2	10	10000	100	9.6308
3	10	10000	100	8.6772
17	10	10000	100	7.1056

n	m	R	R'	estimator
2	4	1000000	1000	9.6023
3	4	1000000	1000	8.6766
17	4	1000000	1000	7.1989
2	10	1000000	1000	9.6148
3	10	1000000	1000	8.6977
17	10	1000000	1000	7.1565

There are small diferences in values obtained using two different allocation propositions.

2.4 Antithetic estimator (European)

Consider antithetic estimator using random variables $(Y_i, Y_{i+R/2})$ with $Y_{i+R/2} = -Y_i$, where $Y_i \sim \mathcal{N}(0,1)$ for $i = 1, ..., R/2$. We can think of it as generating $R/2$ samples and then copying them symmetrically relative to zero. Then estimator will be given as:

$$\hat{Y}_R^{ant} = \frac{e^{-r}}{R} \sum_{i=1}^{R} Y_i.$$

Variance in this case is:

$$\begin{aligned}
Var[\hat{Y}_R^{ant}] &= \frac{e^{-2r}}{R} Var\left[1 + Corr[Y_1, Y_2]\right] \\
&= \frac{e^{-2r}}{R} Var\left[1 + \frac{Cov[Y_1, Y_2]}{\sqrt{Var[Y_1]}\sqrt{Var[Y_2]}}\right] \\
&= \frac{e^{-2r}}{R} Var\left[1 + \frac{Cov[Y_1, -Y_1]}{Var[Y_1]}\right] \\
&= \frac{e^{-2r}}{R} Var\left[1 - \frac{Cov[Y_1, Y_1]}{Var[Y_1]}\right] \\
&= \frac{e^{-2r}}{R} Var\left[1 - \frac{Var[Y_1]}{Var[Y_1]}\right] \\
&= 0.
\end{aligned}$$

Theoretical variance is 0 so we can expect that this estimator should give very good results. This value is due to specific choose of samples where $Y_{i+R/2} = -Y_i$ what leads to correlation equal to -1. Our estimator will get the same form as in (CMC_eu).
Results below:

R	estimator	relative error
100	12.2205	0.9363%
10000	12.1181	1.7668%
1000000	12.3319	0.0331%

We can see that estimators are quiet good but the strange thing at first glance is that $R = 100$ gives better estimation than $R = 10000$.

2.5 Control variate estimator (European)

At the end we consider control variate estimator with $n = 1$, where our control variate wil be $X = B(1)$, where $B(t)$ is defined as earlier as standard Brownian Motion. In this case our estimator will be build from scratches in form:

$$\hat{Y}_i^{cv} = Y_i + c(X_i - \mathbb{E}X_i),$$

where $c = -\frac{Cov[Y_i, X_i]}{Var[Y_i]}$. Then we will combine all $\hat{Y}_i^{cv}$ using CMC method described in section 2.1.
Notice that $\mathbb{E}[B(1)] = 0$ then $\hat{Y}_i^{cv} = Y_i + cX_i$. Additionally we can notice that in every R-iteration we just need to add term cX_i to our computations. Finally estimator takes form:

$$\hat{I}_R^{cv} = \frac{e^{-r}}{R} \sum_{i=1}^{R} \Big[(S_i(1) - K)_+ - \frac{Cov[(S(1) - K)_+, B(1)]}{Var[(S(1) - K)_+]} \cdot B_i(1)\Big]. \qquad \text{(CV\_eu)}$$

We can notice that $Cov[(S(1) - K)_+, B(1)]$ an $Var[(S(1) - K)_+]$ are just a constant. Covariance is equal:

$$
\begin{aligned}
Cov[(S(1)-K)_+, B(1)] &= \mathbb{E}[(S(1)-K)_+ \cdot B(1)] - \mathbb{E}[(S(1)-K)_+] \cdot \mathbb{E}[B(1)] \\
&= \mathbb{E}[(S(1)-K)_+ \cdot B(1)] \\
&= \int_{\mathbb{R}} \left(S(0)e^{\mu^*+\sigma x} - K\right)_+ x \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx \\
&= \int_{\frac{\sigma}{2}-\frac{r}{\sigma}}^{\infty} \left(S(0)e^{\mu^*+\sigma x} - K\right) x \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx \\
&= \int_{-0.075}^{\infty} \left(100e^{0.01875+0.25x} - 100\right) x \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx \\
&\approx 16.4894,
\end{aligned}
$$

while variance is equal like in section 2.1:

$$
Var[(S(1)-K)_+] = 378.5
$$

Variance is:

$$
Var[\hat{Y}_i^{cv}] = (1 - Corr^2[Y_i, X_i])Var[Y_i],
$$

where correlation is given as $Corr[Y_i, X_i] = \dfrac{Cov[Y_i, X_i]}{\sqrt{Var[X_i]Var[Y_i]}}$, then we can notice that variance will be reduced according to CMC in case of non-zero covariance. Theoretical value of variance is:

$$
\begin{aligned}
Var[\hat{Y}_i^{cv}] &= (1 - Corr^2[Y_i, X_i])Var[Y_i] \\
&= \left(1 - \left(\frac{Cov[Y_i, X_i]}{\sqrt{Var[X_i]Var[Y_i]}}\right)^2\right) \frac{342.481}{R} \\
&= \left(1 - \left(\frac{16.4894}{\sqrt{1 \cdot 378.5}}\right)^2\right) \frac{342.481}{R} \\
&= \frac{96.4554}{R}.
\end{aligned}
$$

Results:

R	estimator	relative error	theoretical variance
100	15.4791	25.4795%	0.964554
10000	12.3273	0.0701%	0.00964554
1000000	12.3377	0.0139%	0.0000964554

We can notice that in our case relative error decreases as we increase number of replications R. Same as in CMC we need more than $R = 100$ replications - for this R estimation is not too good.

3. Summary

Now let us briefly summarize observations and conclusions based on our results presented in the previous section:

- Crude Monte Carlo estimator is one of the simplest method to estimate value I, but there are some techniques which lead to better results and what is worth noticing - variance reduction. This estimator need at least $\sim 10^4$ replications to give proper results.
- Based on the results obtained from using Control Variate estimator we can conclude that the bigger number of replications R, the better results we get, but unfortunately it is not the rule for the rest of estimators.
- We had a chance to observe that results obtained for different allocation methods for stratified estimator gives us different results, but there are still some doubts regarding choosing the number of pilot replications or number of strata. The proper choice of those parameters can nicely improve obtained results. And of course as we know from theory variance of estimator obtained with optimal allocation is reduced in comparison to proportional allocation.
- Theoretical variance of Control variate estimator is much lower than Crude Monte Carlo, but the estimators are very similar.
- Bigger R in asian options does not mean lower estimator. It starts oscilating near the value which we can expect is the estimated exact value.
- Antithetic estimator has variance equal 0 but still it does not provide the best results but we can see that it leads to situation when $R = 100$ can give better results than $R = 10000$. There is not too small chance that even $R = 10^{12}$ will give not too good result becase variance does not depend on R. Zero variance is good in case if we totally don't know how many replications we need - in this case even small number like $R = 100$ can give very good result.

References

[1] http://www.math.uni.wroc.pl/~rolski/Zajecia/sym.pdf