Home Assignment 1

Felix E. Slothower
Theodora M. Gaiceanu

February 24, 2022

Task 1

a

Notations: X - a random variable, f_X - probability density function of X, F_X - cumulative distribution function of X, $F_{X|X\in I}(x)$ - conditional cumulative distribution function given that $X \in I$, $f_{X|X\in I}(x)$ - conditional probability density function given that $f_{X|X\in I}(x)$, where $I = (a, b)$.

$$F_{X|X\in I}(x) = \frac{P(X \leq x, X \in I)}{P(x \in I)} = \frac{\int_a^x f_X(x)\,dx}{\int_a^b f_X(x)\,dx} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}$$

$$f_{X|X\in I}(x) = F_{X|X\in I}(x)^{'} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}^{'} = f_X \frac{1}{F_X(b) - F_X(a)}$$

b

Notations: $F_{X|X\in I}(x)^{-1}$ - the inverse of the conditional distribution function.

It is known that $F^{\leftarrow} = inf\{x \in \mathbb{R} : F(x) \geq u\}$, $u \sim Unif(0, 1)$, where $F^{\leftarrow}$ is the general inverse and $Unif(0, 1)$ is the uniform distribution [3].

One should solve $F(F^{-1}(u)) = u \Rightarrow F^{-1}(F(u)) = u$.

$$F_{X|X\in I}(x) = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}$$

$$F_{X|X\in I}(x)(F^{-1}(u)) = u \Rightarrow$$

$$\frac{F_X(F^{-1}(u)) - F_X(a)}{F_X(b) - F_X(a)} = u \Rightarrow$$

$$F_X(F^{-1}(u)) - F_X(a) = u(F_X(b) - F_X(a)) \Rightarrow$$

$$F_X^{-1}(F_X(F_X^{-1}(u))) = F_X^{-1}(u(F_X(b) - F_X(a)) + F_X(a)) \Rightarrow$$

$$F_{X|X\in I}^{-1}(u) = F_X^{-1}(u(F_X(b) - F_X(a)) + F_X(a))$$

It is known the fact that if U is a uniform random variable on (0,1), then the inverse $X = F_X^{-1}$ has the distribution function F [3]. So, if one knows the distribution function F, the inverse method can be used to generate data from a truncated distribution. First, one needs to generate uniform variables on the interval [F(a), F(b)], where a and b are the bounds of the interval. Then one computes the inverse of the distribution function (according to the result above) and uses it to compute the expectation. The resulting value are from the F distribution truncated on [a,b] [6].

Task 2

a

For the standard Monte Carlo, one needs to do a Monte Carlo experiment for every month. Therefore, one should draw a sufficient number of independent random variables from the Weibull distribution, gives them as input to the wind power function P and compute the mean of the powers. This can be done in MATLAB using the *wblrnd* function with the corresponding λ and k to each month. Then the confidence interval is computed with the following equation:

$$I_\alpha = (\tau_N - \lambda_{\alpha\div 2}\frac{\sigma(\phi)}{\sqrt{N}}, \tau_N + \lambda_{\alpha\div 2}\frac{\sigma(\phi)}{\sqrt{N}}), \tag{1}$$

where τ_N is the expectation value from the Monte Carlo experiment, λ_p is the p-quantile of the standard normal distribution covering τ with probability $1-\alpha$, $\sigma(\phi)$ is the standard deviation of the objective function (the power of the wind turbine in our case) and N is the number of samples (10000 in our case).

For the truncated Monte Carlo version, we computed the Weibull distribution functions $F(3)$ and $F(30)$ using the MATLAB function *wblcdf* with the the parametes 3 and 30 respectively, also applying the corresponding λ and k for each month. Then, in a Monte Carlo experiment, we computed the inverse of the distribution function according to the result from b, $F_{X|X\in I}^{-1}(u) = F_X^{-1}(u(F_X(b) - F_X(a)) + F_X(a))$. Here, F_X^{-1} is the Weibull inverse and it can be computed with the MATLAB function *wblinv*. Then the result of the Weibull inverse is given as input to the wind power. The expectation is computed as the mean of the powers. The confidence interval in computed using Equation (1), but with the expectation from the truncated version of Monte Carlo.

The width of the confidence interval when using 10000 samples for the standard Monte Carlo can be seen in the second line of Table 1 and Table 2. The width of the confidence interval when using 10000 samples for the truncated Monte Carlo can be seen in the third line of the tables. The difference between the widths of the intervals can be seen in the last lines of Table 1 and Table 2 But this difference is not so big.

For a better understanding of the process, we plotted the Monte Carlo expectation for the standard and the truncated version for January. From Figure 1, it can be noticed that the standard Monte Carlo has a high variance in the beginning. Both expectations converge to almost the same value. The confidence intervals are very similar, fact that can be noticed from Table 1 and Table 2 also. 10000 samples were used for the confidence interval.

Jan	Feb	March	April	May	June
236096.8	238269.8	237267.0	234195.9	229696.7	233473.5
228897.7	228406.4	226858.7	215384.3	213205.7	216970.9
6649	9477	10166	16192	17000	17695

Table 1: Table with the width of the 95 % confidence intervals for the standard Monte Carlo and the truncated version - months January - June

July	Aug	Sep	Oct	Nov	Dec
231439.1	240223.4	235251.3	235629.4	232154.3	235890.8
212862.0	226946.8	230421.7	230421.7	226836.6	226983.1
17107	14733	10739	10077	8630	9448

Table 2: Table with the width of the 95 % confidence intervals for the standard Monte Carlo and the truncated version - months July - December

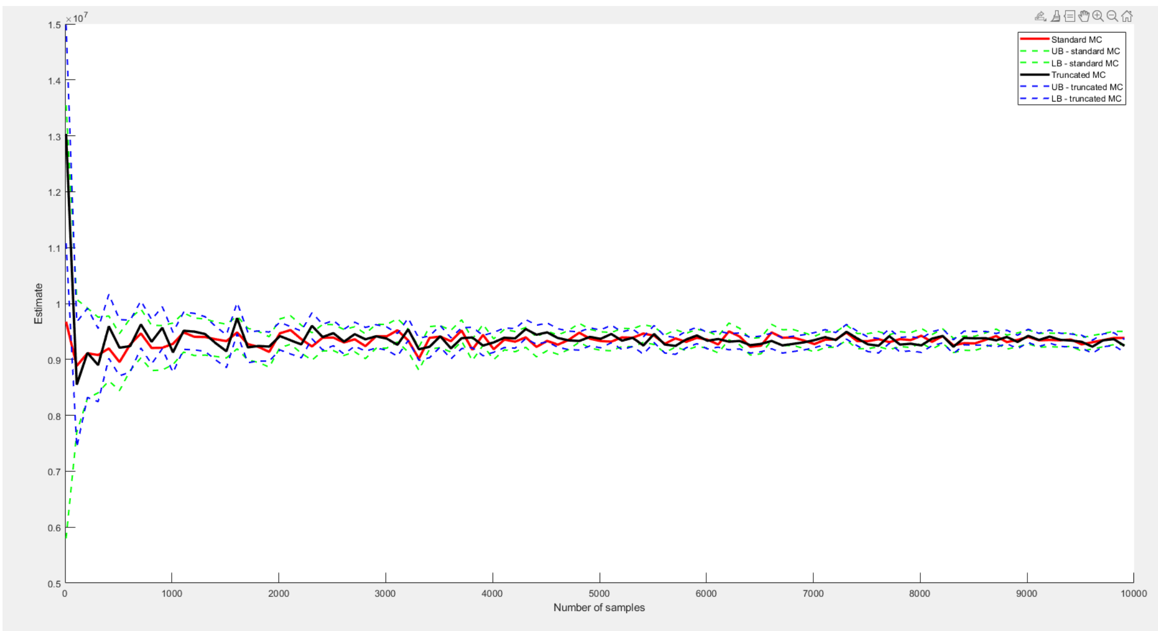

Figure 1: The standard Monte Carlo estimation (red line), the upper bound and the lower bound of the 95% confidence interval for the standard Monte Carlo (green lines). the truncated Monte Carlo estimation (black line), and the upper bound and the lower bound of the 95% confidence interval for the truncated Monte Carlo (blue lines); the estimation is the power generated in MW (y axis)

b

Let Y be a random variable that is a control variate with a known mean m [4]. That means that, for a parameter β, we can form:

$$Z = \phi(X) + \beta(Y - m), \tag{2}$$

where X is a random variable, $\phi(X)$ is the objective function.

In this case, the expectation τ becomes:

$$\tau = \mathbb{E}(Z) = \mathbb{E}(\phi(X) + \beta(Y - m)) \tag{3}$$

We can find an optimal β^* in terms of variance by computing it with the following expression:

$$\beta^* = -\frac{Covariance(\phi(X), Y)}{Variance(Y)} \tag{4}$$

Therefore, in order to do the Monte Carlo simulation with V as a control variate, one should draw some random samples from a Weibull distribution with the corresponding λ and k to each month. Then the control variate V is represented as a Weibull random variable with the suitable λ and k. The expectation (mean) of this control variate is known and it can be computed as in Equation (5):

$$\mathbb{E}[V] = \Gamma(1 + \frac{1}{k})\lambda^m \tag{5}$$

Equation (5) can be easily computed in MATLAB using the *gamma* function with the parameters $1 + \frac{1}{k}$ and the corresponding λ for the current month. Then the optimal β is computed according to Equation (4), where the objective function is the wind power of the random variable X, and Y is the control variate V. Then Z is computed and the expectation τ is calculated as the mean of all Zs. The confidence interval is calculated according to Equation (1).

Jan	Feb	March	April	May	June
9286575.5	8474717.8	7859834.4	6631946.3	6378520.6	6781424.2
9521844.1	8712102.6	8097605.0	6865172.1	6608770.8	7016317.5

Table 3: Table lower bound and the upper bound of the 95 % confidence intervals for the Monte Carlo with control variate V (wind) - months January - June

July	Aug	Sep	Oct	Nov	Dec
6392117.7	6661939.5	7909589.7	8522437.3	9225419.7	9298302.3
6623529.0	6895587.6	8145699.7	8760954.8	9461259.9	9533781.1

Table 4: Table lower bound and the upper bound of the 95 % confidence intervals for the Monte Carlo with control variate V (wind) - months July-December

The bound of the 95% confidence interval for the control variate version of Monte Carlo can be analysed in Table 3 and Table 4. The second line represents the lower bound and the third line represents the upper bound. The confidence interval for the control variate version is slightly narrower than the confidence interval for the standard Monte Carlo and the truncated version. 10000 samples were used for the confidence interval. This can be seen also in Figure 2. But for less than 1000 samples, the variance of estimate for the control variate version is still very high. For more than 5000 samples, the variance is reduced. But what happens for less than 1000 samples suggests that the wind V is not a very good control variate.

c

For importance sampling, one needs to choose an instrumental density g on domain X that satisfies the fact $g(x) = 0 \Rightarrow \phi(x)f(x) = 0$ [5]. The integral for the expectation can be rewritten as:

$$\tau = \mathbb{E}_f(\phi(X)) = \int_X \phi(x)f(x)\,dx = \int_{|\phi(x)|f(x)>0} |\phi(x)f(x)|\,dx = \int_{g(x)>0} \phi(x)\frac{f(x)}{g(x)}g(x)dx =$$

$$= \mathbb{E}_g(\phi(X)\frac{f(X)}{g(X)}) = \mathbb{E}_g(\phi(X)\omega(X)), \tag{6}$$

where ω is the importance weight function, $\omega : x \in X : g(x) > 0x \mapsto \frac{f(x)}{g(x)}$. X is a random variable, *f(x)* is the probability density of X, $\phi(x)$ is the objective function.

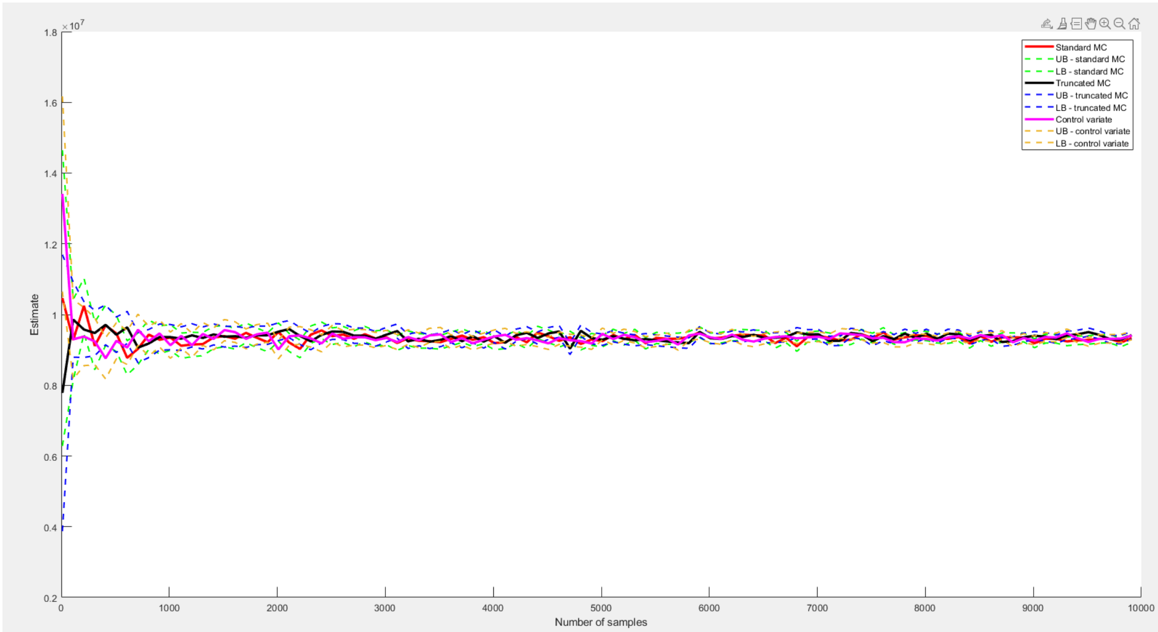

Figure 2: The standard Monte Carlo estimation (red line), the upper bound and the lower bound of the 95% confidence interval for the standard Monte Carlo (green lines). the truncated Monte Carlo estimation (black line), the upper bound and the lower bound of the 95% confidence interval for the truncated Monte Carlo (blue lines), the control variate estimate (magenta line), the upper and lower bounds for the control variates (yellow lines); the estimation is the power generated in MW (y axis)

The instrumental density g needs to be a standard distribution. Therefore, one good choice for g may be the normal density function, defined in MATLAB as *normpdf*. But one needs to choose a proper mean and standard deviation. In our case, f is the Weibull density function for some Weibull random variables with a λ and k from the table in the assignment. ϕ is the wind power of the random variables and $\omega = \frac{f}{g}$. After several experiments, a good value for the mean seems to be 11, while a good value for the standard deviation seems to be 7.

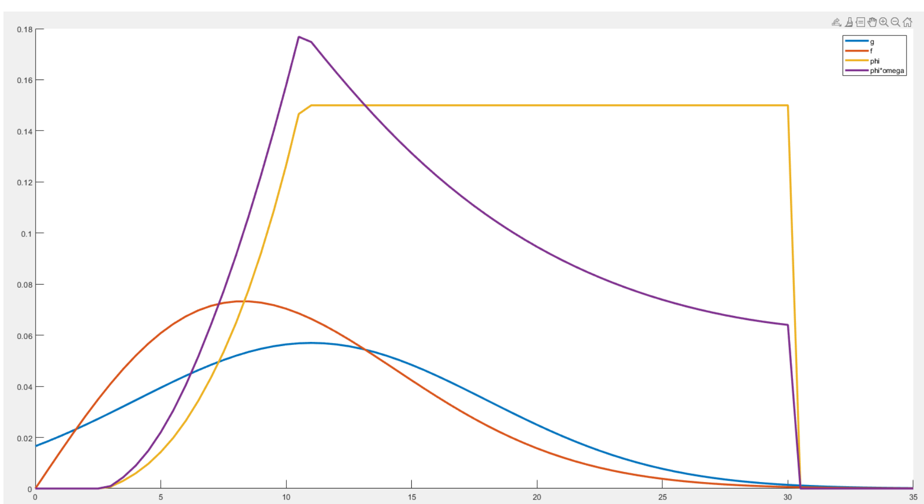

Figure 3: Probability densities for variables inside the interval [0,35] : instrumental density g with mean 11 and standard deviation 7 (blue line), f - probability density of X (red line), ϕ - objective function, or the wind power of X (yellow line), $\phi * \omega$ (purple line); the estimation is the power generated in MW (y axis)

From Figure 3, it can be seen that this particular choice for g makes the function $x \mapsto \phi(x) * \omega(x)$ close to constant in the support of g. Therefore, the mean and the variance have good values. We can expect for a reduction in the variance of the estimate.

Once the mean and the standard deviation are found for the normal distribution, the simulation can start. First, one draws some normal distribution numbers of the mean 11 and standard variance 7 using the MATLAB function *randn*. Then the probability distribution f is computed as the Weibull probability distribution of the normal variables (of course, with the parameters λ and k corresponding to the current month). g is computed as the normal distribution using the MATLAB function *normpdf*. Next we compute the weight function $\omega = \frac{f}{g}$. The objective function ϕ is now computed as $P(X) * \omega$, where $P(X)$ is the wind power of the normal variables. The estimate is the mean of all values of the objective function ϕ multiplied with the weight function; $\tau = \mathbb{E}_g(\phi(X)\omega(X))$. The confidence interval is computed using Equation (1).

Jan	Feb	March	April	May	June
9284743.3	8440345.9	7926456.5	6637955.7	6378155.5	6733227.2
9510229.0	8663561.2	8149619.0	6838976.7	6576208.8	6936234.0

Table 5: Table lower bound and the upper bound of the 95 % confidence intervals for the Monte Carlo with importance sampling - months January - June

July	Aug	Sep	Oct	Nov	Dec
6381398.1	6618461.8	7887707.2	8530504.0	9229079.4	9133616.9
6581298.3	6821471.3	8108095.2	8741520.4	9455154.0	9360734.3

Table 6: Table lower bound and the upper bound of the 95 % confidence intervals for the Monte Carlo with importance sampling - months July-December

The bound of the 95% confidence interval for the importance sampling version of Monte Carlo can be analysed in Table 5 and Table 6. The second line represents the lower bound and the third line represents the upper bound. The confidence interval for the importance sampling version is similar to the confidence interval of the control variate version. So, it is narrower than the confidence interval for the standard and truncated Monte Carlo. 10000 samples were used for the confidence interval.

The estimation along with the confidence interval can be observed in Figure 4 that illustrates the estimations for January. First of all, the variance of the estimate is clearly reduced. Even for less than 1000 samples, the importance sampling estimation looks quite smooth. Again, it can be observed that the confidence intervals for the importance sampling and the control variate versions are narrower than the confidence intervals for the standard and the truncated Monte Carlo. Though, after 1000 samples, the difference is not so big.

d

In antithetic sampling, we have a variable V that is equal to the objective function $\phi(X)$ and it holds that $\tau = \mathbb{E}(V)$. One can generate a variable $\tilde{V}$ that has the estimation τ, has the same variance as V and can be simulated at the same complexity [4]. Then one can have that for:

$$W = \frac{V + \tilde{V}}{2}$$

it is true that $\mathbb{E}(W) = \tau$. Also, if the antithetic variables V and $\tilde{V}$ are negatively correlated, we could gain computational work [4].

In order to do the experiment with the antithetic sampling, one should draw some samples from the standard uniform distribution. Here, as we will have two variables, we should use only half of the

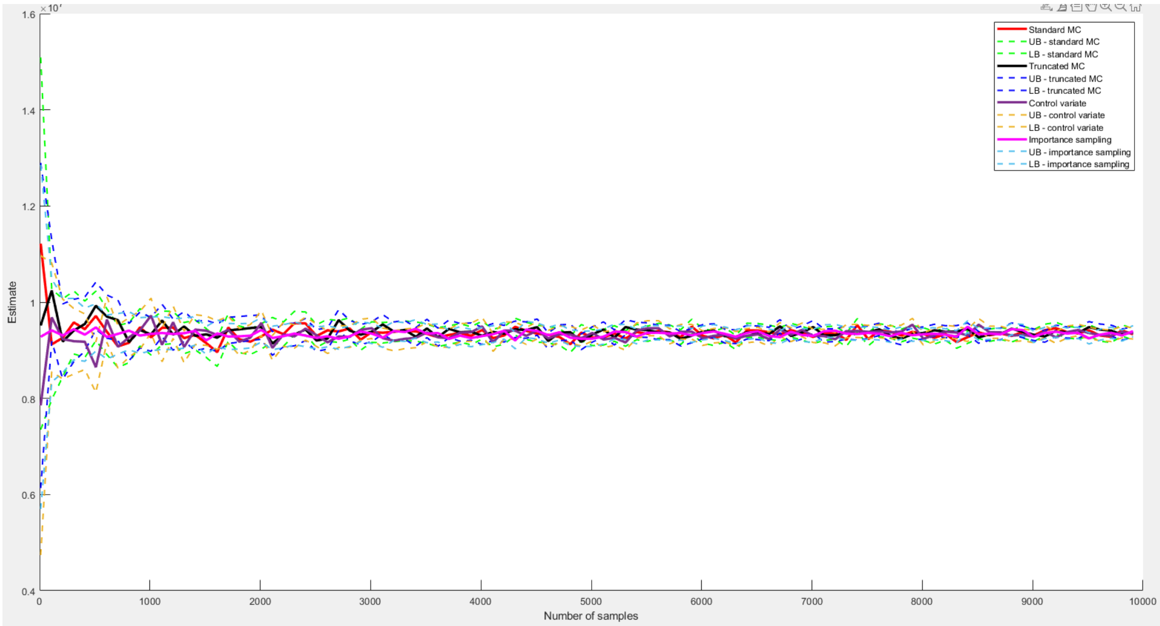

Figure 4: The standard Monte Carlo estimation (red line), the upper bound and the lower bound of the 95% confidence interval for the standard Monte Carlo (green lines). the truncated Monte Carlo estimation (black line), the upper bound and the lower bound of the 95% confidence interval for the truncated Monte Carlo (blue lines), the control variate estimate (purple line), the upper and lower bounds for the control variates (yellow lines), the importance sampling estimate (magenta line), the upper and lower bounds of the 95% confidence interval for the importance sampling (light blue lines)

normal distribution samples. As the samples should be only inside the interval [3,30], one should use the inverse method from Problem 1 in order to make the V and $\tilde{\text{V}}$ variables. One should also check if the covariance of the V and $\tilde{\text{V}}$ variables is negative. And, after checking, their covariance is negative indeed. Then the variable W is computed as $\frac{P(V)+P(\tilde{V})}{2}$. The estimation is computed as the mean of all W variables. The confidence interval is computed according to Equation (1).

The bound of the 95% confidence interval for the antithetic sampling version of Monte Carlo can be analysed in Table 7 and Table 8. The second line represents the lower bound and the third line represents the upper bound. The confidence interval for the antithetic sampling version is similar to the confidence interval of the control variate and importance sampling versions. So, it is narrower than the confidence interval for the standard and truncated Monte Carlo. 10000 samples were used for the confidence interval. Actually, this confidence interval is the narrowest of all intervals.

The estimation along with the confidence interval can be observed in Figure 5 that illustrates the estimations for January. The variance of the estimate is reduced compared to the standard, the truncated and the control variate version of Monte Carlo. But, there is still some variance when there are less than 500 samples. The importance sampling version reduces better the variance. Anyway, when using between 5000 and 10000 samples, the variance is reduced very much.

Jan	Feb	March	April	May	June
9338200.6	8470709.1	7859137.1	6176970.8	5884309.7	6323426.9
9388501.1	8494082.9	7881018.6	6231631.0	5944125.7	6375842.7

Table 7: Table lower bound and the upper bound of the 95 % confidence intervals for the Monte Carlo with antithetic sampling - months January - June

July	Aug	Sep	Oct	Nov	Dec
5881413.7	6324876.0	7751134.3	8531476.7	9324298.0	9350626.5
5940534.6	6377464.6	7774825.4	8556732.9	9374720.1	9401185.4

Table 8: Table lower bound and the upper bound of the 95 % confidence intervals for the Monte Carlo with antithetic sampling - months July-December

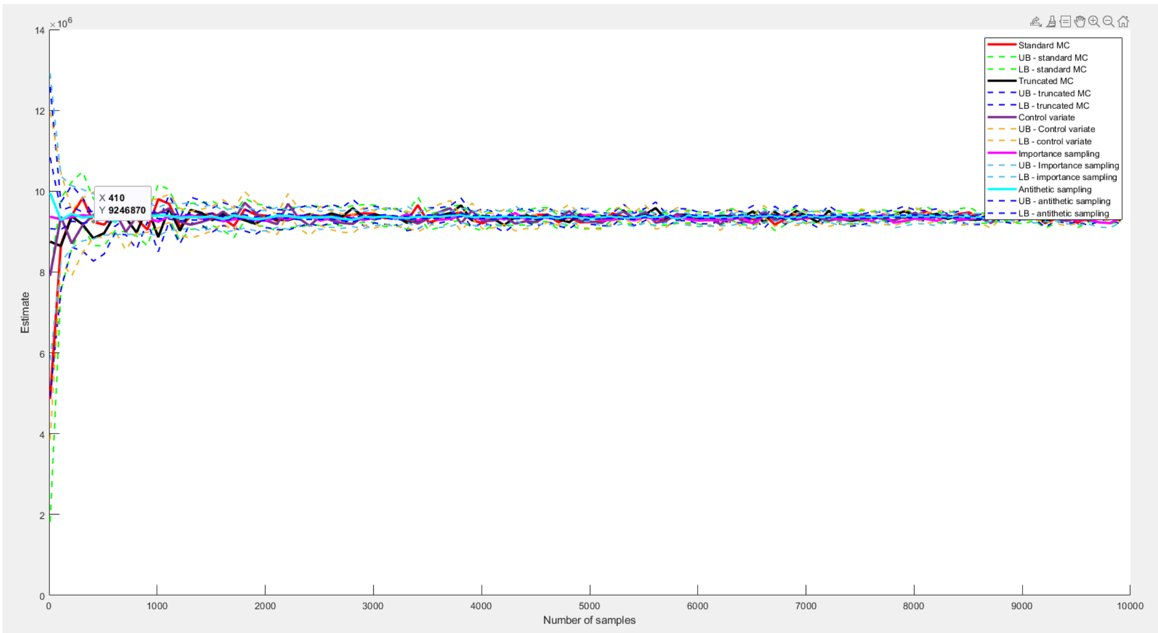

Figure 5: The standard Monte Carlo estimation (red line), the upper bound and the lower bound of the 95% confidence interval for the standard Monte Carlo (light green lines). the truncated Monte Carlo estimation (black line), the upper bound and the lower bound of the 95% confidence interval for the truncated Monte Carlo (blue lines), the control variate estimate (purple line), the upper and lower bounds for the control variates (yellow lines), the importance sampling estimate version (magenta), the upper and lower bounds of the confidence interval for the importance sampling version (light blue lines), the antithetic sampling estimate version (cyan line), the upper and lower bounds of the confidence interval for the antithetic sampling (blue lines); the estimation is the power generated in MW (y axis)

e

First, some random Weibull numbers are generated using the MATLAB function *wblrnd* with the correct parameters λ and k for the current month. Then the wind power is calculated for each variable. The positive powers are counted using the MATLAB functions *length* and *find*. The probability is computed as the number of positive powers divided by the total number of samples.

The results have been illustrated in Figure 6. 10000 samples were used. It can be observed that all months, except June, July, August, and December, have a probability above 90% of having positive wind powers. Anyway, even for the rest of the months, the probability is still high,being over 86%.

The results can be observer numerically in Table 9 and Table 10. The results are obtained using 10000 samples.

Jan	Feb	March	April	May	June
0.93	0.92	0.91	0.87	0.87	0.87

Table 9: Probability of the wind power to be greater than 0 - months January - June

July	Aug	Sep	Oct	Nov	Dec
0.86	0.88	0.91	0.91	0.93	0.93

Table 10: Probability of the wind power to be greater than 0 - months July-December

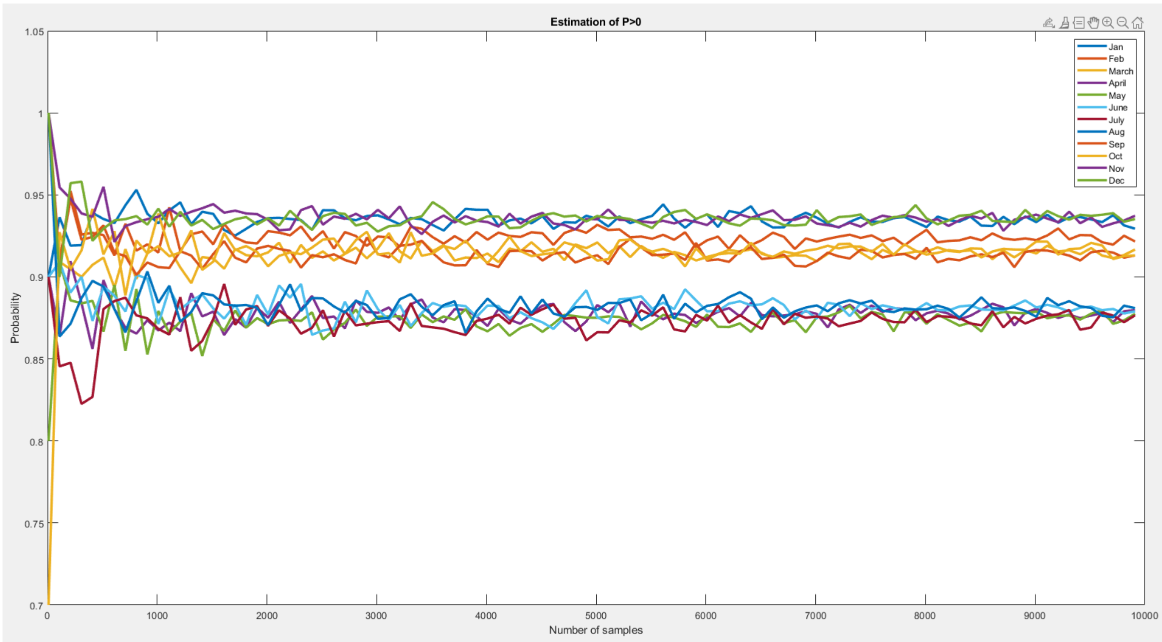

Figure 6: Estimation for each month of the probability that the wind power is greater than 0

f

The numerator $\mathbb{E}(P(V))$ is the expectation computed earlier with Monte Carlo methods. On the other hand, the denominator should be calculated in another way.

In general, we have that:

$$\mathbb{E}(\phi(X)) = \int_A \phi(x) f(x)\, dx \Rightarrow \mathbb{E}P_tot(V) = \int_A P_tot(v) f(v)\, dv = \int_A (\frac{1}{2}\rho\pi\frac{d^2}{4}v^3)(\frac{k}{\lambda}(\frac{v}{\lambda})^{k-1} exp(-(\frac{v}{k})^k))\, dv =$$

$$= \frac{1}{2}\rho\pi\frac{d^2}{4}\int_A v^3 f(v)\, dv = \frac{1}{2}\rho\pi\frac{d^2}{4}\mathbb{E}[V^3] \tag{7}$$

We know that $\mathbb{E}[V^m] = \Gamma(1 + \frac{m}{k})\lambda^m$, $m > -k$.

We also know that $\Gamma(r) = \int_0^\infty t^{r-1} exp(-t)\, dt$, $r > 0$ from [2].

In our case $r = 1 + \frac{m}{k}$. From the expression of the Γ integral and from Equation (7), we have that:

$$\mathbb{E}(P_{tot}(V)) = \frac{1}{2}\rho\pi\frac{d^2}{4}\int_0^\infty t^{\frac{m}{k}} exp(-t)\lambda^m\, dt \tag{8}$$

We can compute the result of Equation (8) in MATLAB using the *gamma* function. In our case, m equals to 3. In MATLAB, we will have 12 values (one for each month because λ varies each month) for the $P_{tot}(V)$. Then, after having the values for $P_{tot}(V)$, we can compute the ratio. For $\mathbb{E}(P(V))$ we have chosen to use the importance sampling estimate, as it is the most accurate (has minimal variance). The values of the $\frac{\mathbb{E}(P(V))}{\mathbb{E}(P_{tot}(V))}$ can be seen in the Table 11 and Table 12. Also, Figure 7 illustrates the values of the average power coefficients for every month for different number of samples.

Jan	Feb	March	April	May	June
0.1638	0.1963	0.2169	0.2618	0.2687	0.2574

Table 11: Average power coefficient - months January - June

July	Aug	Sep	Oct	Nov	Dec
0.2688	0.2537	0.2248	0.1763	0.1637	0.1648

Table 12: Average power coefficient - months July-December

The confidence interval for the average power coefficient can be seen in Table 13. The second line represents the lower bound, while the third line represents the upper bound.

Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
0.1614	0.1936	0.2169	0.2539	0.2645	0.2494	0.2656	0.2492	0.2189	0.1726	0.1616	1.1601
0.1655	0.1991	0.2234	0.2630	0.2741	0.2582	0.2752	0.2580	0.2256	0.1776	0.1657	0.1643

Table 13: Confidence interval for the average power coefficient

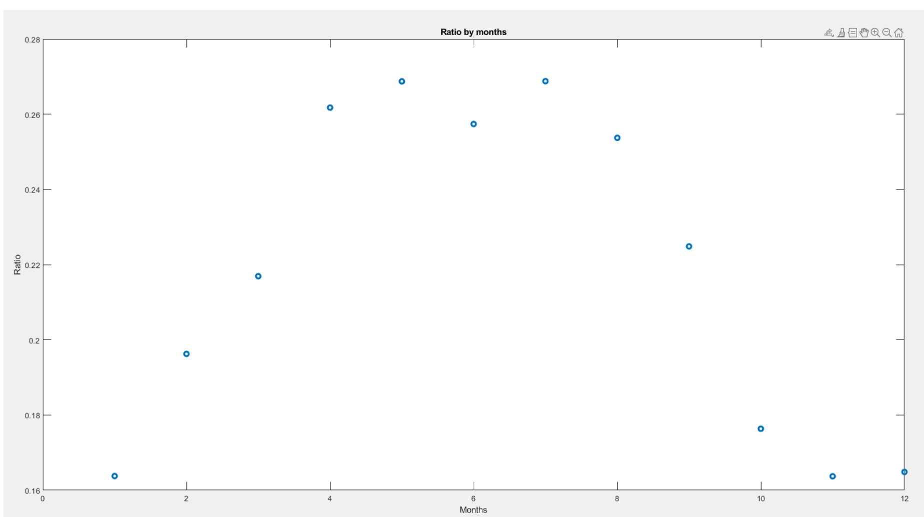

Figure 7: Average power coefficient for each month

g

The capacity factor can be computed as the mean of all values of the ratio $\frac{\mathbf{E}P(V)}{15MW}$, this gives a value of 0.5246. The availability factor can also be computed as the mean of all probabilities from *e*. This gives a value of 0.9051. As the capacity factor is 52.46% and the availability factor is 90.51%, one can conclude that this is a good site to build a wind turbine.

Task 3

a

To model the combined power generation of two turbines we must use a bivariate Weibull distribution with joint probability distribution function $f(v_1, v_2)$ given below [1].

$$f\left(v_1, v_2\right) = f\left(v_1\right) f\left(v_2\right) \left[1 + \alpha \left(1 - F\left(v_1\right)^p\right)^{q-1} \left(1 - F\left(v_2\right)^p\right)^{q-1} \left(F\left(v_1\right)^p \left(1 + pq\right) - 1\right) \left(F\left(v_2\right)^p \left(1 + pq\right) - 1\right)\right]$$

where $\alpha = 0.638$, $p = 3$, $q = 1.5$. Also the parameters for our univariate Weibull probability distribution functions $f(v_1)$ and $f(v_2)$ have the same parameters $k = 1.95$ and $\lambda = 10.05$.

The importance sampling Monte Carlo method will be conducted in a similar was as in Part however we will now update $\omega()$ to instead include the joint distribution functions f and g;

$$\omega() = \frac{f(v_1, v_2)}{g(v_1, v_2)}$$

We must also update our instrumental distribution g to now be a multivariate normal distribution function. Using the *mvnpdf()* command in MATLAB we can find values from it's joint probability distribution. We must also update the mean and variance of this multivariate normal distribution such that $\phi()\omega()$ is roughly constant. After some manual tuning we ended up with the following;

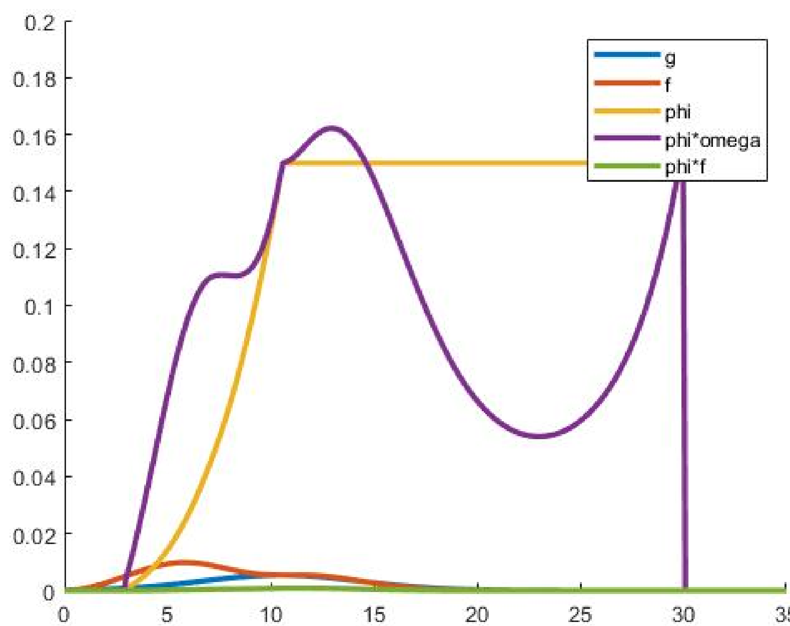

Figure 8: Plot of joint distribution f, instrumental density g and the power function $\phi()$. Through the selection of g we were able to find a function $\phi\omega$ which is close to constant.

Where the means and variances of our multivariate normal distribution were both, $\mu = 10.5$ and $\sigma^2 = 29.5$, respectively.

The expected combined power generation;

$$\mathbb{E}(P(V_1) + P(V_2))$$

can be reduced to a one dimensional problem since V_1 and V_2 are both Weibull distributed with equivalent parameters λ and k. Thus their expected values should be equal;

$$\mathbb{E}(V_1) = \mathbb{E}(V_2)$$

As well as the associated power output as a function of wind speed.

$$\mathbb{E}(P(V_1)) = \mathbb{E}(P(V_2))$$

Since the expected value of the sum of several distributions is equal to the sum of the expected values of each distribution, we have;

$$\mathbb{E}(P(V_1) + P(V_2))$$
$$\mathbb{E}(P(V_1) + P(V_1))$$
$$\mathbb{E}(2P(V_1))$$

By using the importance sampling Monte Carlo method as detailed above and further in question 2 Part c, and by sampling N samples from our distribution twice we found the expected amount of combined power generated to be;

$$\mathbb{E}(P(V_1) + P(V_2)) = 15.9MW$$

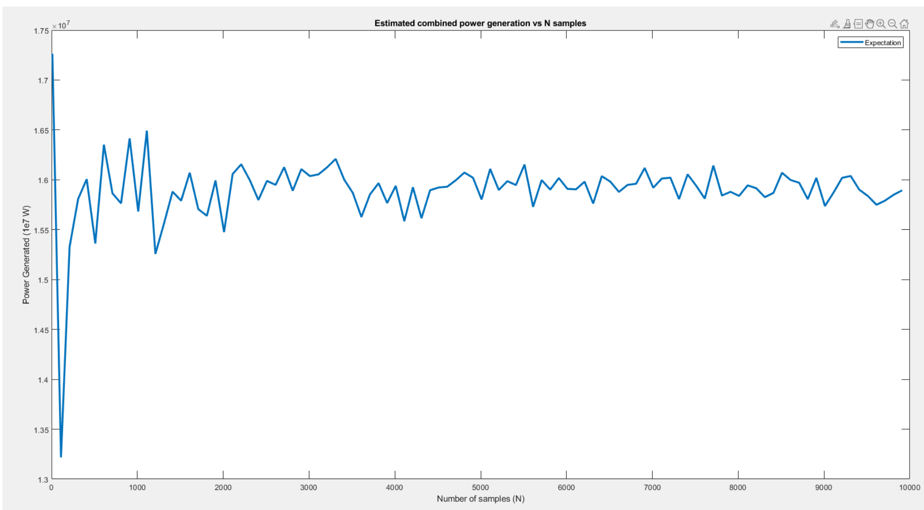

Figure 9: Plot of our estimate τ_N as we increase the number of samples N from 100 to 10,000. This is meant to illustrate the decrease in estimate variance as we increase N.

b

To calculate the covariance between the power generated from either turbine we refer to the definition of covariance;

$$\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

And in our case, would look like;

$$\mathbb{C}(P(V_1), P(V_2)) = \mathbb{E}(P(V_1)P(V_2)) - \mathbb{E}(P(V_1))\mathbb{E}(P(V_2))$$

This means we must find the expectation of power generated from each of the wind turbines as well as the expectation of the multiplied power generation, each time using the importance sampling monte carlo method to find the estimate.

Also since each wind speed alone is marginally Weibull distributed, we use the univariate Weibull distribution when calculating $\mathbb{E}(P(V_1)$ and $\mathbb{E}(P(V_2))$;

$$\mathbb{C}(P(V_1), P(V_2) = 1.88\text{e}13$$

c

The definition of variance is described below;

$$\mathbb{V}(X) = \frac{\sum(x_i + \bar{x})^2}{n-1}$$

In our case X is a random variable which models the combined power generation of our two wind turbines, $P(V_1) + P(V_2)$. We can use the power generation estimate found in Part A in place of $\bar{x}$. This gives us the expression;

$$\mathbb{V}(P(V_1) + P(V_2)) = \frac{(P(V_1) + P(V_2)) - \tau_N}{n-1}$$

This gives;

$$\mathbb{V}(P(V_1) + P(V_2)) = 5.48\text{e}13$$

The standard deviation is defined as the square root of the variance, therefore we can find;

$$\mathbb{D}(P(V_1) + P(V_2)) = \sqrt{5.48\text{e}13} = 7.40\text{e}6$$

d

To find the probabilities that our combined turbine power generation was either above or below 15MW we simply modified our Monte Carlo estimator to record samples where the sum of power generated was greater than 15 MW;

$$H \leftarrow P(V1) + P(V2) > 15MW$$

Then take the dot product between this and our importance sampling weight vector ω finally dividing by the number of samples to find our estimated probability τ_N;

$$\tau_N = \frac{H \cdot \omega}{N}$$

Where our weight vector is defined as the fraction of the multivariate Weibull distribution f to our instrumental distribution g;

$$\omega = \frac{f(v_1, v_2)}{g(v_1, v_2)}$$

Because our weights are unbounded in the 'above 15MW' case, as can be seen by the exponentially increasing purple line in Figure 8, the variance of g was increased to 50.5 so that it would decay at a slower rate than the multivariate Weibull distribution for large wind speeds. Similar variance and mean were used in the 'below 15MW' case as from Part A. Confidence intervals were calculated using equation 1. All information about the estimated probabilities and confidence intervals can be found in the table below;

$\mathbb{P}$	τ_N	CI
$\mathbb{P}(P(V1) + P(V2) > 15MW)$	0.5061	[0.5052,0.5070]
$\mathbb{P}(P(V1) + P(V2) < 15MW)$	0.4819	[0.4813,0.4825]

The sum of our two estimates τ_N is not equal to 1. This is because of the way we are estimating our parameters. The bias caused by the Weibull distribution is not perfectly corrected for with our instrumental distribution during importance sampling meaning it will still be present and skew our final estimation.

References

[1] I. Bairamov, S. Kotz, and M. Bekçi. "New generalized Farlie-Gumbel-Morgenstern distributions and concomitants of order statistics". In: *Journal of Applied Statistics* 28.5 (July 2001), pp. 521–536. DOI: 10.1080/02664760120047861. URL: https://doi.org/10.1080/02664760120047861.

[2] Hoeting J. A. Givens G. H. *Computational Statistics*. John Wiley Sons, 2013. Chap. 1.

[3] Wiktorsson M. *Monte Carlo and Empirical Methods for Stochastic Inference*. Lund University, 2022. Chap. Lecture 2.

[4] Wiktorsson M. *Monte Carlo and Empirical Methods for Stochastic Inference*. Lund University, 2022. Chap. Lecture 4.

[5] Wiktorsson M. *Monte Carlo and Empirical Methods for Stochastic Inference*. Lund University, 2022. Chap. Lecture 3.

[6] Wicklin R. *The inverse CDF method for simulating from a distribution*. URL: https://blogs.sas.com/content/iml/2013/07/22/the-inverse-cdf-method.html. (accessed: 05.02.2022).