Home Assignment 1

Monte Carlo and Empirical Methods for Stochastic Inference

Group 13

Adrian Murphy
ad5880mu-s@student.lu.se

Daniel Larsson
da6625la-s@student.lu.se

March 27, 2023

1 Random Number Generation

Per the instructions, we let X be a random variable on $\mathbb{R}$ where the density function f_X, the invertible distribution function F_X and the distribution inverse function F_X^{-1} are known. We moreover define the interval $I = (a, b)$ such that $\mathbb{P}(X \in I) > 0$.

Firstly, we want to find the conditional distribution function $F_{X|X\in I}(x) = \mathbb{P}(X \leq x \mid X \in I)$ and the density $f_{X|X\in I}(x)$. We start by expanding

$$F_{X|X\in I}(x) = \mathbb{P}(X \leq x \mid X \in I) = \mathbb{P}(X \leq x \mid a < X < b)$$

For any distribution function it holds that $F : \mathbb{R} \to [0, 1]$. Since we condition on $a < X < b$, we must by definition have that $F_{X|X\in I}(\hat{x}) = 0$ for any $\hat{x} \leq a$ and $F_{X|X\in I}(\hat{x}) = 1$ for any $\hat{x} \geq b$. We expand further to account for the case when $a < x < b$

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

We find $f_{X|X\in I}(x)$ by differentiating $F_{X|X\in I}(x)$, where of course $X \notin I \implies f_{X|X\in I}(x) = 0$ since we are differentiating a constant. As for $X \in I$, we find

$$\frac{d}{dx}\frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} = \frac{f_X(x)(F_X(b) - F_X(a)) - 0}{(F_X(b) - F_X(a))^2} = \frac{f_X(x)}{F_X(b) - F_X(a)}$$

1(a) Answer

$$F_{X|X\in I}(x) = \begin{cases} 0, & \text{if } x \leq a \\ 1, & \text{if } x \geq b \\ \frac{F_X(x)-F_X(a)}{F_X(b)-F_X(a)}, & \text{otherwise} \end{cases}$$

$$f_{X|X\in I}(x) = \begin{cases} \frac{f_X(x)}{F_X(b)-F_X(a)}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$$

We also want to find the inverse $F^{-1}_{X|X\in I}$. Per the definition of a function's inverse we have that

$$\begin{aligned}
&F_{X|X\in I} \circ F^{-1}_{X|X\in I}(x) = x \iff \\
&\frac{F_X \circ F^{-1}_{X|X\in I}(x) - F_X(a)}{F_X(b) - F_X(a)} = x \iff \\
&F_X \circ F^{-1}_{X|X\in I}(x) = x \cdot [F_X(b) - F_X(a)] + F_X(a) \iff \\
&F^{-1}_{X|X\in I}(x) = F_X^{-1}(x \cdot [F_X(b) - F_X(a)] + F_X(a))
\end{aligned}$$

1(b) Answer

The sought inverse, under the presumption that $a < F^{-1}_{X|X\in I}(x) < b$, is

$$F^{-1}_{X|X\in I}(x) = F_X^{-1}(x \cdot [F_X(b) - F_X(a)] + F_X(a))$$

As described in the second lecture[1], we can use this in the inversion sampling method to generate pseudo-random numbers from X conditionally on $X \in I$. We first define the general inverse as $F^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}\ \forall u \in [0, 1]$. If the distribution function is strictly monotone, we can use that $F^{\leftarrow} = F^{-1}$. A strictly monotone function is of course a one-to-one mapping, a bijection, which by definition only has one $x \in \mathbb{R}$ for which $F(x) \geq u$. The inverse of a bijection is also a bijection, which is why $F^{\leftarrow} = F^{-1}$ holds in that case.

We may use the following procedure to simulate X conditionally on $X \in I$

draw $U \sim \mathcal{U}(0, 1)$
set $X \leftarrow F^{\leftarrow}_{X|X\in I}(U)$
return X

---

2 Power Production of a Wind Turbine

We now wish to investigate the amount of power generated by a wind turbine, model V164 from Vestas with a rotor diameter of 164 meters and a tower height of 105 meters, given some data about how the wind, which we assume follows a Weibull distribution, varies throughout a year for a site in northern Europe. We are given a function P which maps wind speed onto output power in watts for the V164 turbine according to

$$P_{\text{tot}}(v) = \frac{1}{2}\rho\pi\frac{d^2}{4}v^3$$

where v is the wind speed in meters per second, ρ the air density (about 1.225 kilos per cubic meter at sea level) and d is the turbine's rotor diameter.

We start by constructing a 95% confidence interval for the produced power each month of the year using crude Monte Carlo. We accomplish this by drawing random $X_i \sim f$, where $f(v) = \frac{k}{\lambda}(\frac{v}{\lambda})^{k-1}\exp(-(\frac{v}{\lambda})^k)$ is the Weibull distribution function with λ and k corresponding to a certain month. As described in the first lecture[2] we have that, by the law of large numbers

$$\tau_N = \lim_{N\to\infty} \sum_{i=1}^{N} \phi(X_i) \to \mathbb{E}(\phi(X_i))$$

In terms of notation, it is convenient to let the objective function ϕ be P. Crude Monte Carlo for a given month can subsequently be implemented as

for $i = 1 \to N$ do
  draw $X_i \sim f$
end for
set $\tau_N \leftarrow \sum_{i=1}^{N} P(X_i)/N$
return τ_N

As described in the second lecture[3], a confidence interval, by implication of the central limit theorem, can be constructed as

$$\mathcal{I}_\alpha = (\tau_N - \lambda_{\alpha/2}\frac{\sigma(P)}{\sqrt{N}},\ \tau_N + \lambda_{\alpha/2}\frac{\sigma(P)}{\sqrt{N}})$$

We subsequently move on to, again, constructing a 95% confidence interval, this time only considering the winds within the interval 3.5 to 25 meters per second (the interval in which the turbine is operational). We now draw random $X_i | X_i \in (3.5, 25)$ according to the procedure described in problem 1(b). This procedure (inversion sampling) replaces the draw $X_i \sim f$ described in the pseudo-code for crude Monte Carlo above. We also have to adjust for the conditioning when computing the estimate by multiplying $P(X)$ with $F_X(b) - F_X(a)$.

We use sample size $N = 10\ 000$ for all methods. A larger sample size would of course theoretically reduce variance and the confidence interval width. However, since this assignment is an exercise in implementation and feasibility of different Monte Carlo methods and not an exercise in achieving as exact values as possible, we opted for a lower sample size to enable faster calculations.

2(a) Answer

We use Matlab's **confplot** function to plot the expected power generation as a blue line and the 95% confidence interval as a surrounding gray area.

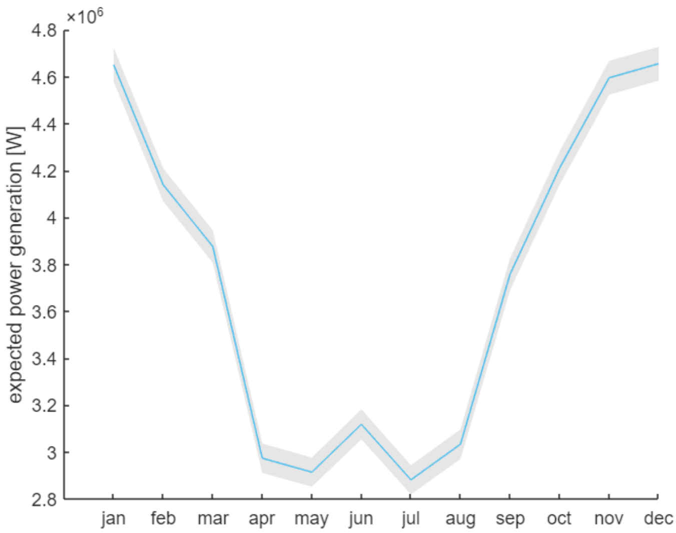

Figure 1: Expected power generation using crude Monte Carlo

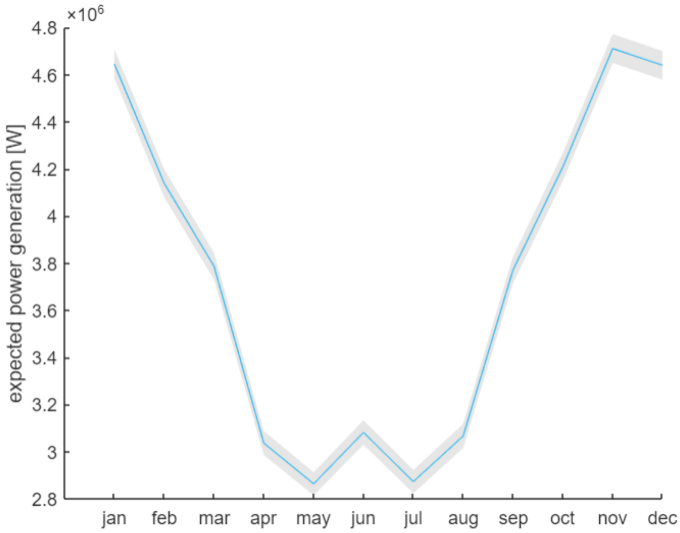

Figure 2: Expected power generation using truncated Monte Carlo

Crude Monte Carlo	134253.9785
Truncated Monte Carlo	110801.1131

Table 1: Mean confidence interval width

In table 1 we can see that we, on average, achieve an $\approx 17\%$ reduction in confidence interval width using the truncated method. This reduction owes to the fact that we do not draw wind speeds that yield zero power in the truncated method (i.e., wind speeds outside the operational interval), which leads to a lower variance. Since we adjust for the conditioning when computing the estimated power generation, the expected values are about the same for both methods. The full tables describing the confidence interval for both crude 4(a) and truncated 4(a) Monte Carlo for each month can be found in the appendix.

For the next part of the assignment, we used the wind V as a control variate to decrease the variance, constructing yet another 95% confidence interval for the power generation. If the power generation is a random variable G, we say that for some $\beta \in \mathbb{R}$

$$Z = G + \beta(V - \mathbb{E}(V))$$

We satisfy the two requirements that make the control variate method feasible in the sense that $\mathbb{E}(V)$ is known and can be computed analytically as $\lambda\Gamma(1 + 1/k)$. λ and k are the Weibull parameters for a certain month and Γ is simply the gamma function. Moreover, G and V can be simulated at the same complexity since $G = P(V)$. We now have that

$$\mathbb{E}(Z) = \mathbb{E}(G + \beta(V - \mathbb{E}(V))) = \mathbb{E}(G) + \beta(\mathbb{E}(V) - \mathbb{E}(V)) = \tau$$

In other words, the expected value is preserved. We moreover have that

$$\begin{aligned}&\mathbb{V}(Z) = \mathbb{V}(G + \beta V) = \mathbb{C}(G + \beta V,\ G) + \beta V) = \\ &\mathbb{V}(G) + 2\beta\mathbb{C}(G, V) + \beta^2\mathbb{V}(V)\end{aligned}$$

where $\mathbb{C}$ denotes the covariance. Optimizing with regards to β gives us a minimal variance

$$0 = 2\mathbb{C}(G,V) + 2\beta\mathbb{V}(V) \iff \beta = \beta^* = -\frac{\mathbb{C}(G,V)}{\mathbb{V}(V)}$$

We make the following modifications to our algorithm, again using $N = 10\ 000$, and apply it to every month

```
for i = 1 → N do
    draw V_i ~ f
    set G_i ← P(V_i)
end for

set m ← λΓ(1 + 1/k)
set β ← −C(G,V)/V(V)
set τ_N ← Σ_{i=1}^{N} (G_i + β(V_i − m))/N

return τ_N
```

2(b) Answer

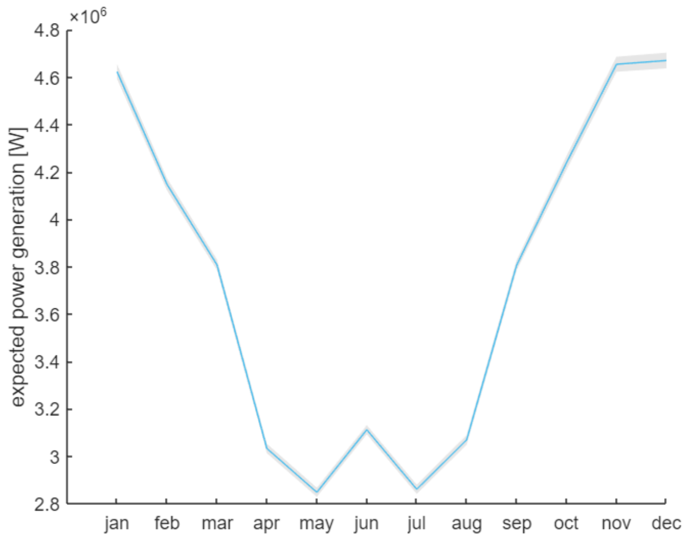

Figure 3: Expected power generation using V as control variate

We can already see in figure 3 that the variance and confidence interval width have been reduced. The mean interval width is 48498.285, which corresponds to a 64% decrease compared to the crude Monte Carlo. As described in lecture 4[4], we can expect a large reduction in variance if the correlation between G and V is close to ± 1. In our case, the power production is a function of the wind, so intuitively it is reasonable to believe that the two variables are highly correlated. This could of course be confirmed by calculations, but with respect to the already large reduction in confidence interval width we deemed this unnecessary. The full confidence interval table 4(a) can be found in the appendix.

We move on to implementing importance sampling. As described in the third lecture[5], importance sampling requires us to formulate an instrumental density g on X such that

$$g(x) = 0 \implies \phi(x)f(x) = 0$$

The idea is that the instrumental distribution should be simpler to sample from and that we somehow can compensate for the bias that is generated when sampling from another distribution. To show why importance sampling works, we can rewrite the integral from crude Monte Carlo as

$$\begin{aligned}\tau = \mathbb{E}_f(\phi(X)) &= \int_X \phi(x)f(x)\mathrm{d}x = \int_{f(x)>0} \phi(x)f(x)\mathrm{d}x \\ &= \int_{g(x)>0} \phi(x)\frac{f(x)}{g(x)}g(x)\mathrm{d}x = \mathbb{E}_g(\phi(X)\frac{f(X)}{g(X)})\end{aligned}$$

From this, we can see that it is a good idea if g is proportional to ϕf. To find a suitable g, we simulate $\phi(x)f(x)$ for every month of the year, where we let ϕ be P and f the Weibull probability distribution, and plot the results. From figure 4, a Gaussian distribution seems like a reasonable instrumental distribution g.

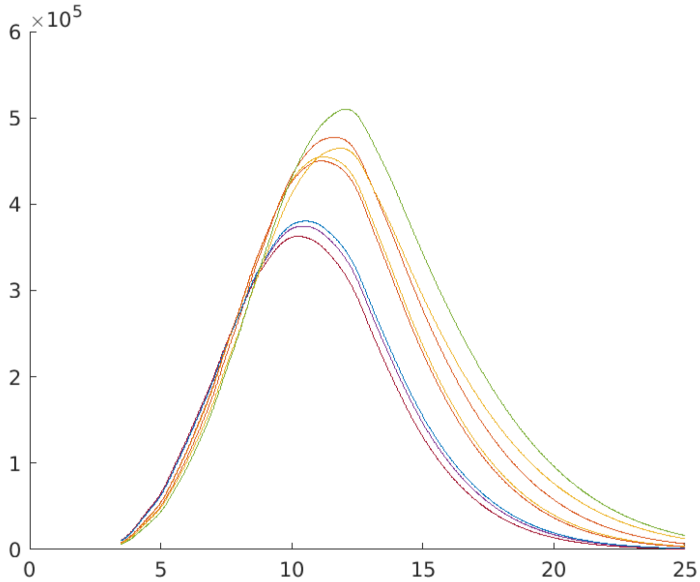

Figure 4: Simulated $\phi(x)f(x)$ for all 12 months

To find a mean μ for each month, we numerically choose the argument that maximizes $\phi(x)f(x)$, typically somewhere between 10 and 12. We choose the standard deviation σ by doing a simulation for the month of January, testing values between 1 and 10 (looking at figure 4, the standard deviation is surely within these bounds) with an incrementation of 0.5. The result can be found in figure 5 below, where we plot the ultimate confidence interval width given a certain σ, from which we deduced that a standard deviation of 4.5 would be reasonable for our Gaussian instrumental distribution.

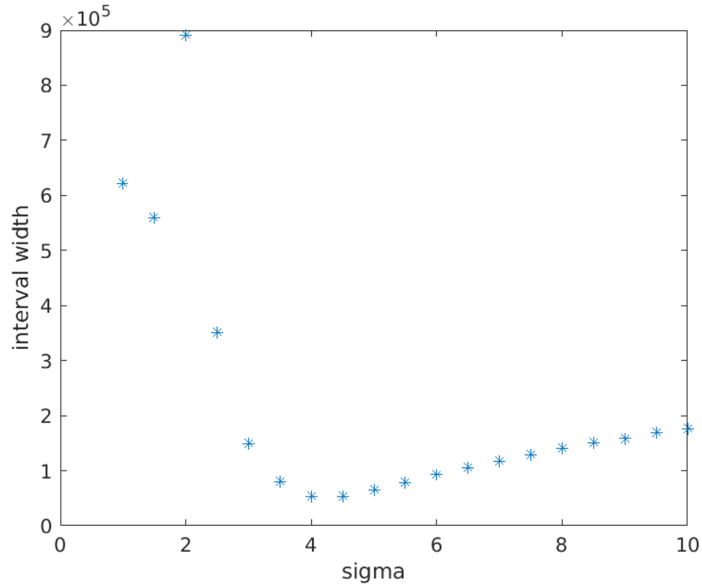

Figure 5: Interval width for different standard deviations σ when simulating the month of January with a Gaussian instrumental distribution

We now implement the following algorithm

for $i = 1 \to N$ **do**
  draw $X_i \sim g = \mathcal{N}(\mu, 4.5^2)$
end for
set $\tau_N \leftarrow \sum_{i=1}^{N}(P(X_i)\frac{f(X_i)}{g(X_i)})/N$
return τ_N

2(c) Answer

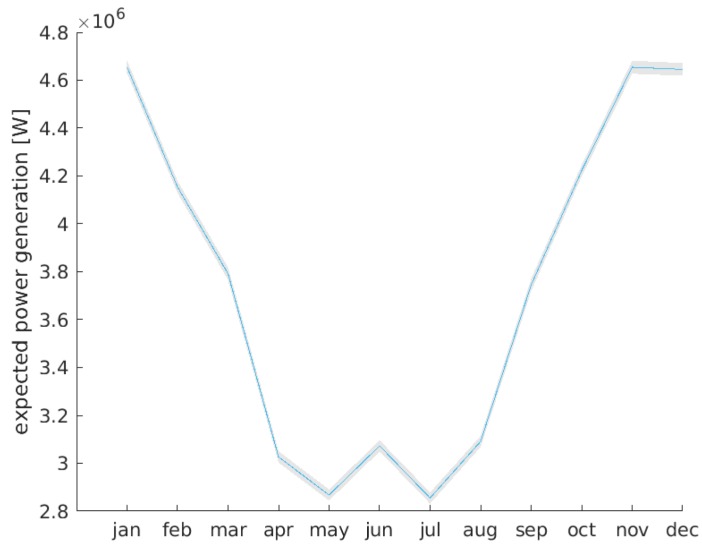

Figure 6: Expected power generation using importance sampling

With importance sampling, we achieved an average interval width of 49780.6836, which is on par with the control variate method and a 63% decrease compared to crude Monte Carlo. Full confidence interval table can be found in the appendix 4(a).

Lecture 4 also describes the antithetic sampling method[6]. In this method, we define $V = \phi(x)$ which in turn means that $\tau = \mathbb{E}(V)$. We now make an assumption that we can generate a variable $\tilde{V}$ where:

i) $\mathbb{E}(\tilde{V}) = \tau$

ii) $\mathbb{V}(V) = \mathbb{V}(\tilde{V})$ $(= \sigma^2(\phi))$

iii) $\tilde{V}$ can be simulated at the same complexity as V

Then, defining $W = \frac{V+\tilde{V}}{2}$ we have $\mathbb{E}(W) = \tau$ and

$$\mathbb{V}(W) = \mathbb{V}(\frac{V+\tilde{V}}{2}) = \frac{1}{4}(\mathbb{V}(V) + 2\mathbb{C}(V,\tilde{V}) + \mathbb{V}(\tilde{V})) = \frac{1}{2}(\mathbb{V}(V) + \mathbb{C}(V,\tilde{V}))$$

To reason why we should use W instead of V we remember that to estimate a τ with an error lower than any ϵ we need that the number of generated samples N_V of V's must satisfy

$$N_V = \lambda^2_{\alpha/2}\frac{\mathbb{V}(V)}{\epsilon^2}$$

and for W:

$$N_W = \lambda^2_{\alpha/2}\frac{\mathbb{V}(W)}{\epsilon^2}$$

With these, we can reason that choosing W is better than V if

$$\begin{aligned} 2\lambda^2_{\alpha/2}\frac{\mathbb{V}(W)}{\epsilon^2} < \lambda^2_{\alpha/2}\frac{\mathbb{V}(V)}{\epsilon^2} &\iff \\ \mathbb{V}(V) + \mathbb{C}(V,\tilde{V}) < \mathbb{V}(V) &\iff \\ \mathbb{C}(V,\tilde{V}) &< 0 \end{aligned}$$

Meaning that if V and $\tilde{V}$ are negatively correlated then we will gain computational work. To prove that this is the case, we use the following theorem referenced by Wiktorsson

Theorem 1 *Let $V = \rho(U)$, where $\rho : \mathbb{R} \to \mathbb{R}$ is a monotone function. Moreover, assume that there exists a non-increasing transform $T : \mathbb{R} \to \mathbb{R}$ such that $U \overset{d.}{=} T(U)$. Then $V = \rho(U)$ and $\tilde{V} = \rho(T(U))$ are identically distributed and*

$$\mathbb{C}(V,\tilde{V}) = \mathbb{C}(\rho(U), \rho(T(U))) \leq 0$$

Using the theorem above, we define F as the distribution function and P as the monotone function. We then continue with $U \sim U(0,1)$, $T(u) = 1-u$, and $\rho(u) = P(F^{-1}(u))$. Finally, letting $V = P(F^{-1}(U))$ and $\tilde{V} = P(F^{-1}(1-U))$ we have

$$V \overset{d.}{=} \tilde{V}$$

and

$$\mathbb{C}(V,\tilde{V}) \leq 0$$

We implement the following algorithm

```
for i = 1 → N/2 do
    draw X_i ~ U(0,1)
    set V_i ← P(F^{-1}(X_i))
    set Ṽ_i ← P(F^{-1}(1 − X_i))
    set W_i ← (V_i + Ṽ_i)/2
end for

set τ_N ← Σ_{i=1}^{N/2} 2W_i/N
return τ_N
```

2(d) Answer

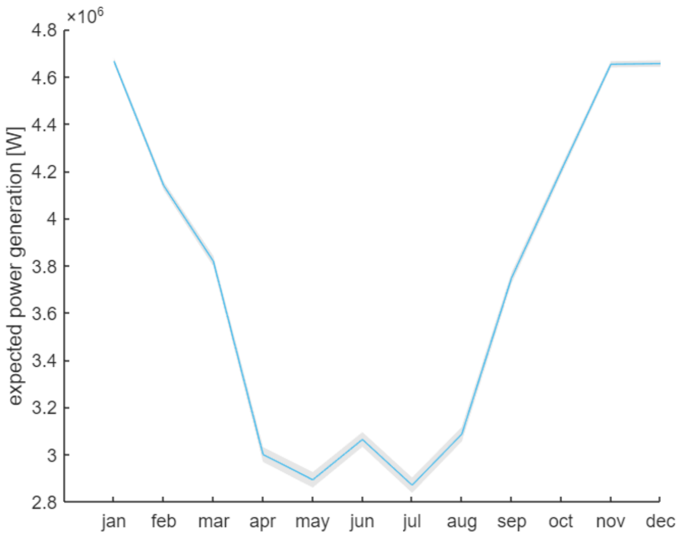

Figure 7: Expected power generation using antithetic sampling

The results using antithetic sampling are shown above in figure 5. One can observe that the confidence interval is quite narrow in this image, and indeed with antithetic sampling we have achieved the best results yet. The mean interval width for this trial was 46366.5236, which is only slightly better than using a control variate or importance sampling but significantly better than the other methods. Full confidence interval table can be found in the appendix 4(a).

An interesting question to ask is what probability there is for each month that our turbine actually produces electricity, i.e. $\mathbb{P}(P(V) > 0)$. This can be calculated analytically for each month by using the wind turbine's operational interval. Power is generated if $V \sim \text{Weib}(\lambda, k) \in (3.5, 25)$. Thus,

$$\mathbb{P}(3.5 < V < 25) = F(25) - F(3.5) = (1 - \exp(-(\frac{25}{\lambda})^k)) - (1 - \exp(-(\frac{3.5}{\lambda})^k)) =$$
$$= \exp(-(\frac{3.5}{\lambda})^k)) - \exp(-(\frac{25}{\lambda})^k))$$

2(e) Answer

Standard sampling	
Month	$\mathbb{P}(3.5 < V < 25)$
Jan	0.89287
Feb	0.87662
Mar	0.86463
Apr	0.81212
May	0.8039
Jun	0.81603
Jul	0.8039
Aug	0.81603
Sep	0.86196
Oct	0.86751
Nov	0.89287
Dec	0.89287

Table 2: Probabilities of generating power for every month with standard sampling from the Weibull distribution

---

We can also take a look at the percentage of power from the wind that the turbine can convert to electrical power, i.e. the average power coefficient. This is calculated as

$$\frac{\mathbb{E}(P(V))}{\mathbb{E}(P_{tot}(V))}$$

where $P_{tot}(V)$ is the total amount of power in the wind that passes through the turbine. $P_{tot}(V)$ has been given as

$$P_{tot}(V) = \frac{1}{2}\rho\pi\frac{d^2}{4}V^3$$

which means that

$$\mathbb{E}(P_{tot}(V)) = \mathbb{E}(\frac{1}{2}\rho\pi\frac{d^2}{4}V^3) = \frac{1}{2}\rho\pi\frac{d^2}{4}\mathbb{E}(V^3)$$

One should also note that $\mathbb{E}(V^m) = \Gamma(1 + m/k)\lambda^m$, which in our case means that

$$\mathbb{E}(P_{tot}(V)) = \frac{1}{2}\rho\pi\frac{d^2}{4}\Gamma(1 + \frac{3}{k})\lambda^3$$

for each k and λ of every month, which means that we won't have to calculate the denominator through simulation.

2(f) Answer

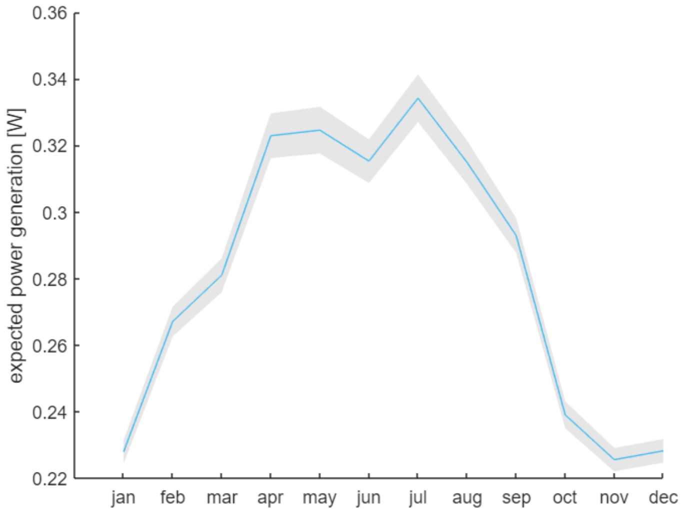

Figure 8: Average power coefficient (crude MC)

Figure 8 shows the average power coefficient for the months. While the results may not be that impressive in general, the months in the middle of the year do at least produce around 10 percentage points more than the earlier and later months of the year. This can be contributed to the lower scale parameter values during these middle months. The full confidence interval table can be found in the appendix 4(a).

We now wish to investigate whether the proposed location for the wind turbine is suitable by analyzing the capacity factor and the availability factor. We can calculate the capacity factor as the ratio between the de facto power generation and the maximum power generation, where we use the numbers from the crude Monte Carlo simulation for the de facto power generation. The maximum power generation, according to the specifications of the turbine, is 9.5 MW. The availability factor is simply the average of the percentages in table 2.

2(g) Answer

Capacity factor	Availability factor
0.39303	0.85011

Table 3: Capacity and availability factor

Per the instructions, we can expect a capacity factor between 20-40% and an availability factor over 90% for a typical wind turbine. Our capacity factor is in the upper end of the expectation, while the availability factor is slightly lower. In fact, the availability is never over 90% for a given month according to the figures in table 2. If the power plant is to be installed in an electrical grid with other sources that do not depend on weather conditions, such as nuclear energy, then availability might be of less importance. However, if the grid relies almost solely on wind energy, then a high availability is of utmost importance. In that case, we should perhaps consider another location. If the grid has other sources of energy, the relatively high capacity factor may be a good enough reason to construct the turbine in the proposed location.

3 Combined Power Production of Two Wind Turbines

In the final part of the assignment, we investigate the power production of two turbines placed in the vicinity of each other while being exposed to similar winds V_1 and V_2.

We first want to calculate the expected power production of both turbines, given by

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

Since V_1 and V_2 are both described by a Weibull distribution with the same parameters λ and k, the problem is reduced to one dimension by

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

We wish to find this expection by means of importance sampling. We simulate $P(x)f(x)$ to find a suitable instrumental distribution.

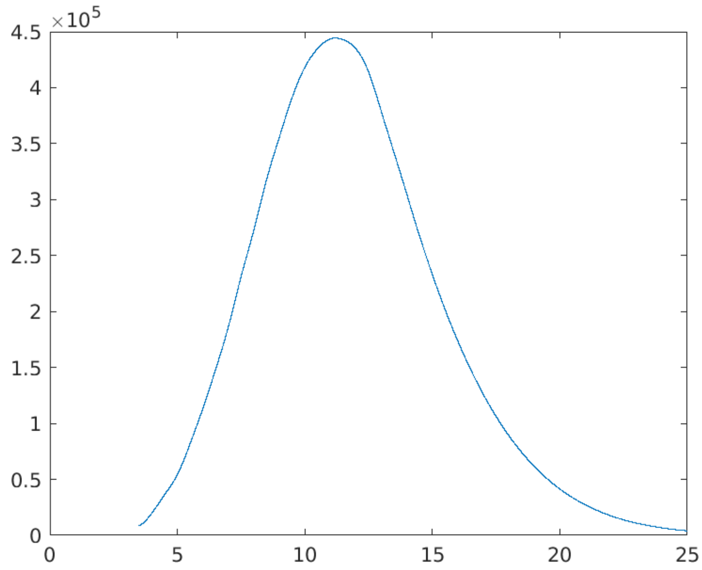

Figure 9: Simulated $P(x)f(x)$

A Gaussian distribution once again seems like an appropriate choice. We choose μ by finding the parameter that maximizes $P(x)f(x)$, in this case 11.187. We choose σ by the same method as in exercise 2(c), where we obtain the following plot

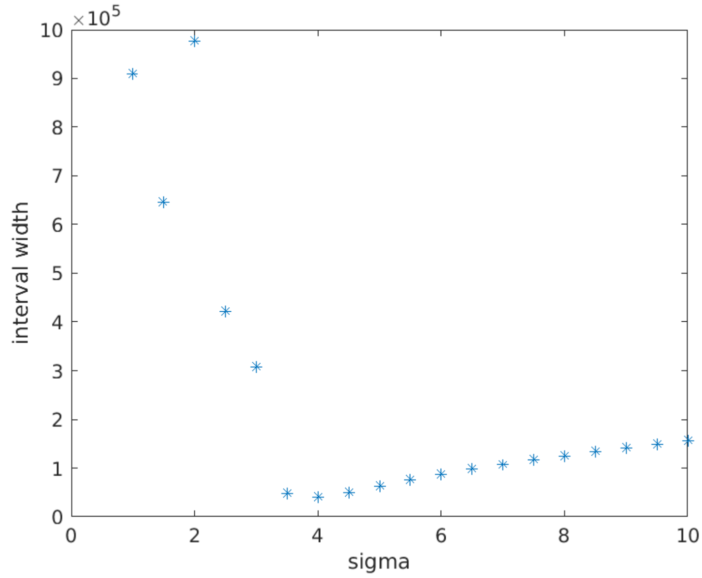

Figure 10: Interval width for different standard deviations σ when simulating with a Gaussian instrumental distribution

Based on figure 10, we choose $\sigma = 4$. We now implement the following algorithm

for $i = 1 \to N$ **do**
  draw $X_i \sim g = \mathcal{N}(11.187, 4^2)$
end for
set $\tau_N \leftarrow \sum_{i=1}^{N}(P(X_i)\frac{f(X_i)}{g(X_i)})/N$
return τ_N

3(a) Answer

We obtain $2\mathbb{E}(V_1) = 7.5176$ MW.

We would also like to calculate the covariance of the power generated from the two turbines, $\mathbb{C}(P(V_1), P(V_2))$, which can be calculated as

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

From the last exercise, we know $\mathbb{E}(V_1)$ and, since they are the same, $\mathbb{E}(V_2)$. We therefore seek $\mathbb{E}(P(V_1)P(V_2))$. From Bairamov et. al.[7], we know that we can write our joint density function $f(v_1, v_2)$ as

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

where $f(v)$ is the density function of a univariate Weibull distribution with parameters $\lambda = 9.13$ and $k = 1.96$. We also have that $\alpha = 0.638$, $p = 3$ and $q = 1.5$. We define $\phi(V_1, V_2) = P(V_1)P(V_2)$. We now define a function ω to be used in the importance sampling

$$\omega = \frac{f(V_1, V_2)}{g(V_1, V_2)}$$

where we sample from a bivariate Gaussian distribution g, analogous to the one-dimensional case from before. As explained in the lecture notes for STAT 505 by Pennsylvania State University [8], we know that a bivariate Gaussian distribution is defined by a vector $\vec{\mu}$ and the covariance matrix Σ as such

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right]$$

We can use that $\mu_1 = \mu_2 = 11.187$ and $\sigma_1 = \sigma_2 = 4$ from the previous exercise. Since the turbines are placed closely together, we conjecture that V_1 and V_2 are strongly correlated. If one wind is strong, the other one will likely be too. When using the joint density function above, the maximum admissible ρ is 0.5021 for $p = 3$ and $q = 1.496$ [7], which are very close to our parameters. For $\rho = 0.5021$, we obtain

$$\vec{\mu} = \begin{pmatrix} 11.187 \\ 11.187 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 16 & 8.0336 \\ 8.0336 & 16 \end{pmatrix}$$

We sample from this bivariate Gaussian distribution and implement importance sampling to calculate $\mathbb{E}(P(V_1)P(V_2))$.

3(b) Answer

$\mathbb{C}(P(V_1), P(V_2)) = 6.575 \cdot 10^{12}$

Having calculated the covariance $\mathbb{C}(P(V_1), P(V_2))$, we can now calculate the variance $\mathbb{V}(P(V_1) + P(V_2))$ using the equation below

$$\mathbb{V}(P(V_1) + P(V_2)) = \mathbb{V}(P(V_1)) + \mathbb{V}(P(V_2)) + 2\mathbb{C}(P(V_1), P(V_2))$$

where our two missing expressions $\mathbb{V}(P(V_1))$ and $\mathbb{V}(P(V_2))$ can be calculated individually by sampling from the Weibull distribution and using crude Monte Carlo.

With the variance calculated, we can also obtain the standard deviation $\mathbb{D}(P(V_1) + P(V_2))$ from

$$\mathbb{D}(P(V_1) + P(V_2)) = \sqrt{\mathbb{V}(P(V_1) + P(V_2))}$$

3(c) Answer

$\mathbb{V}(P(V_1) + P(V_2)) = 3.7149 \cdot 10^{13}$
$\mathbb{D}(P(V_1) + P(V_2)) = 6.0950 \cdot 10^{6}$

Finally, we want to approximate a 95% confidence interval for the probability $\mathbb{P}(P(V_1) + P(V_2) > 9.5 \cdot 10^6)$ and $\mathbb{P}(P(V_1) + P(V_2) < 9.5 \cdot 10^6)$. We define two vectors ϕ_1, containing ones where the total power production is greater than 9.5 MW and zeros otherwise, and ϕ_2, containing ones where the total production is less than 9.5 MW and zeros otherwise.

We construct ϕ_1 by sampling from the bivariate Gaussian distribution in exercise 3(b). For ϕ_2 however, we find that this instrumental distribution makes ω tend to infinity in some cases. To reduce this effect, we increase the standard deviations σ_1 and σ_2 to 5, in hopes that g will decay slower and ω will no longer tend to infinity.

After having sampled from both of these instrumental distribytions and constructed ϕ_1 and ϕ_2, we implement important sampling and obtain the estimated probabilities according to

$$\hat{p}_i = \frac{\phi_i \cdot \omega}{N}$$

We construct 95% confidence intervals for these estimations according to

$$\mathcal{I}_\alpha = (\hat{p}_i - \lambda_{\alpha/2}\frac{\sigma(\phi_i)}{\sqrt{N}},\ \hat{p}_i + \lambda_{\alpha/2}\frac{\sigma(\phi_i)}{\sqrt{N}})$$

3(d) Answer

Probability	Estimated Probability	Confidence interval
$\mathbb{P}(P(V_1) + P(V_2) > 9.5)$	0.3732	0.3732 ± 0.0091
$\mathbb{P}(P(V_1) + P(V_2) > 9.5)$	0.6108	0.6108 ± 0.0090

Table 4: Probabilities and confidence intervals

The probabilities do not sum to 1 exactly, but rather 0.984. We believe there is some bias when sampling from the instrumental distribution that is not fully corrected in the importance sampling, thus skewing the data a bit.

4 Appendix

4(a) Confidence intervals

Crude MC 2(a)			
Month	Left side of interval	Right side of interval	τ_n
Jan	4.5791e+06	4.7227e+06	4.6509e+06
Feb	4.0692e+06	4.2092e+06	4.1392e+06
Mar	3.8078e+06	3.9447e+06	3.8762e+06
Apr	2.9111e+06	3.0359e+06	2.9735e+06
May	2.8522e+06	2.9759e+06	2.9141e+06
Jun	3.0546e+06	3.1817e+06	3.1181e+06
Jul	2.8196e+06	2.9433e+06	2.8814e+06
Aug	2.9703e+06	3.0961e+06	3.0332e+06
Sep	3.6898e+06	3.8250e+06	3.7574e+06
Oct	4.1389e+06	4.2815e+06	4.2102e+06
Nov	4.5233e+06	4.6673e+06	4.5953e+06
Dec	4.5836e+06	4.7273e+06	4.6555e+06

Truncated MC 2(a))			
Month	Left side of interval	Right side of interval	τ_n
Jan	4.5845e+06	4.7067e+06	4.6456e+06
Feb	4.0803e+06	4.1981e+06	4.1392e+06
Mar	3.7302e+06	3.8435e+06	3.7868e+06
Apr	2.9857e+06	3.0867e+06	3.0362e+06
May	2.8144e+06	2.9120e+06	2.8632e+06
Jun	3.0302e+06	3.1330e+06	3.0816e+06
Jul	2.8239e+06	2.9214e+06	2.8727e+06
Aug	3.0151e+06	3.1174e+06	3.0663e+06
Sep	3.7145e+06	3.8274e+06	3.7710e+06
Oct	4.1495e+06	4.2675e+06	4.2085e+06
Nov	4.6496e+06	4.7720e+06	4.7108e+06
Dec	4.5787e+06	4.7007e+06	4.6397e+06

Control variate 2(b)			
Month	Left side of interval	Right side of interval	τ_n
Jan	4.5922e+06	4.6544e+06	4.6233e+06
Feb	4.1223e+06	4.1730e+06	4.1477e+06
Mar	3.7859e+06	3.8303e+06	3.8081e+06
Apr	3.0128e+06	3.0522e+06	3.0325e+06
May	2.8285e+06	2.8665e+06	2.8475e+06
Jun	3.0926e+06	3.1305e+06	3.1116e+06
Jul	2.8410e+06	2.8797e+06	2.8603e+06
Aug	3.0492e+06	3.0888e+06	3.0690e+06
Sep	3.7843e+06	3.8276e+06	3.8060e+06
Oct	4.2095e+06	4.2681e+06	4.2388e+06
Nov	4.6229e+06	4.6860e+06	4.6544e+06
Dec	4.6373e+06	4.7032e+06	4.6702e+06

Importance sampling 2(c)			
Month	Left side of interval	Right side of interval	τ_n
Jan	4.6296e+06	4.6965e+06	4.6631e+06
Feb	4.1179e+06	4.1855e+06	4.1517e+06
Mar	3.8059e+06	3.8731e+06	3.8395e+06
Apr	2.9763e+06	3.0363e+06	3.0063e+06
May	2.8233e+06	2.8824e+06	2.8529e+06
Jun	3.0309e+06	3.0911e+06	3.0610e+06
Jul	2.8606e+06	2.9188e+06	2.8897e+06
Aug	3.0392e+06	3.0992e+06	3.0692e+06
Sep	3.7193e+06	3.7871e+06	3.7532e+06
Oct	4.1883e+06	4.2503e+06	4.2193e+06
Nov	4.6344e+06	4.7014e+06	4.6679e+06
Dec	4.6332e+06	4.6999e+06	4.6665e+06

Antithetic sampling 2(d)			
Month	Left side of interval	Right side of interval	τ_n
Jan	4.5839e+06	4.6856e+06	4.6348e+06
Feb	4.0904e+06	4.1962e+06	4.1433e+06
Mar	3.8248e+06	3.9634e+06	3.8941e+06
Apr	2.8774e+06	3.0748e+06	2.9761e+06
May	2.7497e+06	2.9571e+06	2.8534e+06
Jun	2.9609e+06	3.1575e+06	3.0592e+06
Jul	2.7393e+06	2.9428e+06	2.8410e+06
Aug	2.9358e+06	3.1239e+06	3.0298e+06
Sep	3.6898e+06	3.8311e+06	3.7605e+06
Oct	4.1843e+06	4.2879e+06	4.2361e+06
Nov	4.6188e+06	4.7061e+06	4.6624e+06
Dec	4.5966e+06	4.6906e+06	4.6436e+06

Average power coefficient 2(f)			
Month	Left side of interval	Right side of interval	τ_n
Jan	0.2224	0.2294	0.2259
Feb	0.2545	0.2633	0.2589
Mar	0.2795	0.2897	0.2846
Apr	0.3183	0.3318	0.3251
May	0.3280	0.3423	0.3352
Jun	0.3116	0.3246	0.3181
Jul	0.3247	0.3390	0.3319
Aug	0.3144	0.3276	0.3210
Sep	0.2830	0.2935	0.2883
Oct	0.2347	0.2427	0.2387
Nov	0.2227	0.2297	0.2262
Dec	0.2245	0.2316	0.2281

References

[1] M. Wiktorsson, *The inversion method*, January 2023, lecture 2, slide 17

[2] M. Wiktorsson, *The Monte Carlo (MC) method in a nutshell (Ch. 6.1)*, January 2023, lecture 1, slide 8

[3] M. Wiktorsson, *Confidence bounds*, January 2023, lecture 2, slide 6

[4] M. Wiktorsson, *Control variates (cont.)*, January 2023, lecture 4, slide 15

[5] M. Wiktorsson, *Importance sampling (IS, Ch.6.4.1)*, January 2023, lecture 3, slide 15

[6] M. Wiktorsson, *Antithetic sampling*, January 2023, lecture 4, slide 19 - 25

[7] Bairamov, I., Kotz, S., and Bekci, M. (2001), *New generalized Farlie-Gumbel-Morgenstern distributions and concomitants of order statistics*, Journal of Applied Statistics, 28(5), 521-536, Available at `http://ludwig.lub.lu.se/login?url=http://www.tandfonline.com/doi/pdf/10.1080/02664760120047861?needAccess=true`

[8] The Pennsylvania State University, *4.2 - Bivariate Normal Distribution*, source gathered January 2023, Available at `https://online.stat.psu.edu/stat505/lesson/4/4.2`