# Home Assignment 1: Random Number Generation, Power production of a wind turbine and Combined power production of two wind turbines

**FMSN50 - Monte Carlo and Empirical Methods for Stochastic Inference**

Victoria Lagerstedt & Kajsa Hansson Willis

February 10, 2025

# Contents

# 1 Introduction

Renewable energy sources are playing an increasingly important role in worldwide energy creation. Among them, wind power is a key contributor due to its sustainability and technological advancements. However, evaluating the feasibility and efficiency of a wind turbine at a given location requires a detailed statistical analysis of wind speed distributions and power output.

This report uses Monte Carlo methods and empirical statistical techniques to analyze the expected power production of a Vestas V164 9.5 MW wind turbine based on wind conditions from a site in northern Europe. Wind speeds are modeled using a Weibull distribution and are leveraged in an empirically determined power curve. Several variance reduction techniques—including importance sampling, control variates, and antithetic sampling—are implemented to improve the efficiency of Monte Carlo estimations. Further, bivariate distributions are considered.

# 2 Part 1: Random number generation

## 2.1 a) Conditional distribution function and density given an interval

In the first part of the assignment, the conditional distribution function

$$F_{X|X\in I}(x) = P(X \leq x|X \in I)$$

and the density function were determined. X is here a real random variable, and I=(a,b) is an interval.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

$$P(A \cap B) = \int_a^x f_X(z)\,dz = F_X(x) - F_X(a) \tag{2}$$

$$P(B) = \int_a^b f_X(z)\,dz = F_X(b) - F_X(a) \tag{3}$$

To calculate the conditional distribution function Equation 1, Equation 2 and Equation 3 were used. The conditional distribution density is therefore described by

$$\frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}$$

inside of the interval (a,b), zero if x is smaller than a and one if x is larger than b. This is of course because if x is smaller than a none of stochastically determined values can be smaller than that range, and similarly, if x is larger than b then none of the stochastically determined values can be smaller than x, since they are all within the interval.

To calculate the density function, the derivative of the conditional probability function was calculated with equation 4 for a<x<b and zero otherwise.

$$f_{X|X\in I}(x) = \frac{d}{dx}F_{X|X\in I}(x) = \frac{f_X(x)}{F_X(b) - F_X(a)} \tag{4}$$

## 2.2 b) Inverse function

The inverse conditional distribution function is calculated from the conditional distribution function with Equation 5, where I = (a,b) is an interval.

$$F_{X|X\in I}(x)^{-1} = F_{X|X\in I}(x) * (F_{X|X\in I}(b) - F_{X|X\in I}(a)) + F_{X|X\in I}(a) \tag{5}$$

This inverse conditional distribution can be used to simulate X conditionally on X in I. To simulate this, one can use inverse transform sampling. The first step is to draw random samples from a uniform distribution U in [0,1] and then find the function T that solves T(U)=X. Since uniform samples this can be done by following equation 6.

$$F_X(x) = P(X \le x) = P(T(U) \le x) = (U \le T^{-1}(x)) = T^{-1}(x) \tag{6}$$

Due to these equalities, T(x) is the same as the inverse conditional distribution of x. To simulate X, uniform random variables are used as input into the inverse conditional distribution.

# 3 Part 2: Power production of a wind turbine

Here the potential of a wind turbine at the specific location is evaluated. In each of the following parts, an answer was created for each month by running a loop.

## 3.1 a) Confidence interval drawing from the Weibull distribution

In this part of the assignment, the purpose was to create a 95% confidence interval of the expected amount of power generated by the wind turbine using draws from a Weibull distribution. Further, another confidence interval was supposed to be created based on the truncated version described in part 2.

The confidence interval can be determined according to Equation 7, where $\tau$ is the expected value $\sigma$ the standard deviation, N is the sample size, and p is the corresponding p-value from the normal distribution for the degree of confidence interval (1.96 for a two-sided 95% confidence interval).

$$\text{Confidence interval} = [\tau - \frac{\sigma * p}{\sqrt{N}}, \tau + \frac{\sigma * p}{\sqrt{N}}] \tag{7}$$

In the first step of the Matlab code, vectors were created to store the $\lambda$ and k values for each month, so that vectors could store each month's expected value and confidence interval. For each month, a random Weibull distribution was generated. These values were then inserted into the power curve and the mean was saved as the expected value. Additionally, the standard deviation was calculated for the power curve of the distribution. These values of $\tau$ and $\sigma$ were then inserted into Equation 7 where N was chosen to be 10 000 samples. The results from the standard Monte Carlo can be viewed in Table 1, where the average of each column has been added as a yearly estimate.

| Month: | Mean value (W): | Lower bound (W): | Upper bound (W): | Width: |
|---|---|---|---|---|
| January | 4612986 | 4565966 | 4710358 | 144392 |
| February | 4162590 | 4050305 | 4188583 | 138278 |
| March | 3862859 | 3711460 | 3846808 | 135348 |
| April | 2998917 | 2916247 | 3040740 | 124493 |
| May | 2896244 | 2763523 | 2885496 | 121973 |
| June | 3083294 | 3064023 | 3191759 | 127736 |
| July | 2872771 | 2826996 | 2949610 | 122614 |
| August | 3060296 | 3005860 | 3132857 | 126997 |
| September | 3765101 | 3748644 | 3884213 | 135569 |
| October | 4215729 | 4174607 | 4316852 | 142245 |
| November | 4611597 | 4579472 | 4723601 | 144129 |
| December | 4638101 | 4571896 | 4716416 | 144520 |
| Yearly | 3731700 | 3664916 | 3798941 | 134024 |

Table 1: Estimation of power generation with the expected value, lower and upper bounds of the 95% confidence interval including width for each month and the average as a yearly estimate using standard Monte Carlo

After this, the truncated version was calculated. For each month N random uniform samples were drawn and used in the inverse conditional probability function to create random Weibull samples. This instead of using Matlab's Weibull random generator. The conditional probability function of a Weibull distribution can be seen in equation 8. The inverse probability function of a Weibull distribution was calculated from equations in part 2, the result is seen in equation 9 where k(i) means the k-value for the month with the index i.

$$F_x = 1 - \exp\left(-\left(\frac{x}{\lambda(i)}\right)^{k(i)}\right) \tag{8}$$

$$F_{X|X\in I}(x)^{-1} = \lambda(i) * (-ln(e^{-(3.5/\lambda(i))^{k(i)}} - (e^{-(3.5/\lambda(i))^{k(i)}}) - e^{-(25/\lambda(i))^{k(i)}}) * F_{X|X\in I}(x))^{1/k(i)} \tag{9}$$

Since the interval, I, represented in the truncated version was [3.5,25] m/s, these were the values used in Equation 9. Inserting these generated samples into the power curve and taking the mean generated the expected value. Further, the standard deviation was extracted and the values were, as previously, inserted into Equation 7 to create the confidence interval. The results using the truncated Monte Carlo can be observed in Table 2.

| Month: | Mean value (W): | Lower bound (W): | Upper bound (W): | Width: |
|---|---|---|---|---|
| January | 5196525 | 5128171 | 5264878 | 136707 |
| February | 4727662 | 4660548 | 4794777 | 134229 |
| March | 4523411 | 4457397 | 4589424 | 132027 |
| April | 3707373 | 3645433 | 3769313 | 123880 |
| May | 3566214 | 3505037 | 3627392 | 122355 |
| June | 3776242 | 3713595 | 3838889 | 125294 |
| July | 3557027 | 3495922 | 3618132 | 122210 |
| August | 3800015 | 3737420 | 3862610 | 125190 |
| September | 4380146 | 4314765 | 4445526 | 130761 |
| October | 4865207 | 4796693 | 4933721 | 137028 |
| November | 5259859 | 5192126 | 5327593 | 135467 |
| December | 5236940 | 5168575 | 5305305 | 136730 |
| Yearly | 4383051 | 4317973 | 4448130 | 130156 |

Table 2: Estimation of power generation with the expected value, lower and upper bounds of the 95% confidence interval including width for each month and the average as a yearly estimate using truncated Monte Carlo

Naturally, the expected power generated by the truncated version is larger, as it is conditional on the turbine actually generating power. However, the width of the confidence intervals can also be compared. When observering the yearly estimates it is clear that the truncated version has a slightly shorter interval. It is 130 156 W compared to 134 024 W for the standard Monte Carlo. This is not an extremely significant difference, but a difference nevertheless. Further, the interval of the standard Monte Carlo is wider for nearly every month, except in May.

## 3.2 b) Using wind as a control variate

One method that decreases the variance of a Monte Carlo simulation is to use control variates. In this case, wind was used as the control variate. To do this equation 10 was used, where $\phi$(x) is the power of the simulated wind, Y is simulated wind and m is the mean of Y. Beta is the covariance between $\phi$(x) and Y divided by the variance of Y, times -1. The 95% confidence interval was calculated using equation 7.

$$Z = \phi(X) + \beta(Y - m) \tag{10}$$

The results of the estimation with wind as a control variate can be observed in Table 3. The width of the confidence intervals has decreased significantly which shows that using the wind as a control variate is variance reducing. The reason for this is that information about errors in the wind distribution is used to decrease the errors in the power estimation.

| Month: | Mean value (W): | Lower bound (W): | Upper bound (W): | Width: |
|---|---|---|---|---|
| January | 4638162 | 4603998 | 4672326 | 68328 |
| February | 4119444 | 4090632 | 4148256 | 57624 |
| March | 3779134 | 3750158 | 3808111 | 57953 |
| April | 2978494 | 2947984 | 3009004 | 61020 |
| May | 2824509 | 2793336 | 2855683 | 62347 |
| June | 3127891 | 3098467 | 3157316 | 58849 |
| July | 2888303 | 2857091 | 2919515 | 62424 |
| August | 3069358 | 3039393 | 3099324 | 59931 |
| September | 3816429 | 3787489 | 3845368 | 57879 |
| October | 4245730 | 4213831 | 4277628 | 63797 |
| November | 4651536 | 4618326 | 4684746 | 66420 |
| December | 4644156 | 4610391 | 4677920 | 67529 |
| Yearly | 3731928 | 3700924 | 3762933 | 62008 |

Table 3: Estimation of power generation with the expected value, lower and upper bounds of the 95% confidence interval including width for each month and the average as a yearly estimate with wind as a control variate

## 3.3 c) Using importance sampling

In this part, the purpose was to use importance sampling to construct a 95% confidence interval of the power of the wind turbine. To do so, an importance sampling function, g, must first be determined from an instrumental distribution so that Equation 11 can be applied.

$$\mu = \int \phi(x) f(x) dx = \int \phi(x) \frac{f(x)}{g(x)} g(x) dx \tag{11}$$

To do so, the function $\phi(x)f(x)$, where $\phi(x) = P(v)$, was plotted and then compared to other distributions that could be appropriate instrumental distributions, as can be seen in Figure 1. The function was also evaluated analytically.

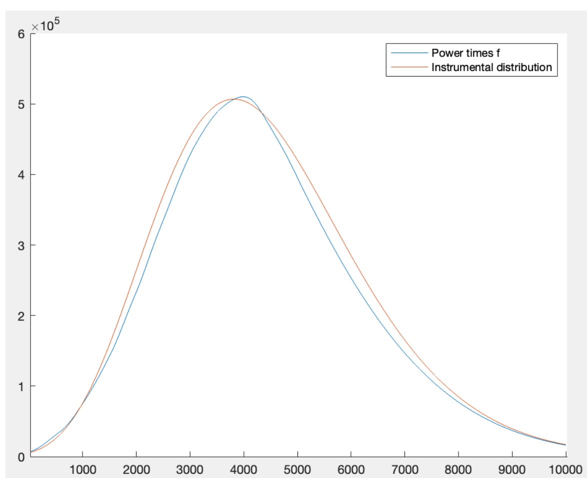


Figure 1: The instrumental distribution and $f(x)\phi(x)$ plotted against each other for January

Further, the ratio between $\phi(x)f(x)$ and g(x), Equation 12, was plotted in Figure 2. If the instrumental distribution is proper, this ratio should be as constant as possible in order to reduce variance as much as possible and shrink the confidence interval.

$$r = \frac{\phi(x)f(x)}{g(x)} \tag{12}$$

The importance sampling that is plotted and was deemed satisfactory is g = 5000000*gampdf(x,10,1.3), which is a Matlab function of the gamma distribution where x is the winds used as input that follow

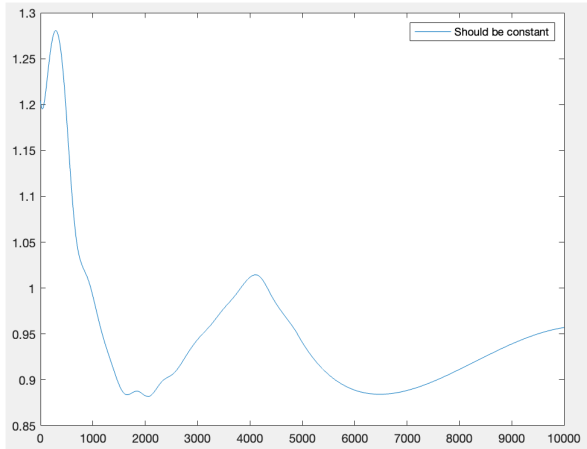


Figure 2: Plot of the ratio described in Equation 12 for January

the Weibull distribution and 10 and 1.3 are the shape and scale parameters. The scalar was chosen to enable visualization of the plots, but it was not necessary for further calculations as it would cancel itself out.

The gamma distribution was deemed as an option because of its similar structure to the power curve, and with proper scaling and choice of parameters this g function was chosen as it appeared very similar to the power curve and the ratio appeared to be quite constant, and it satisfied the requirement such that $g(x) = 0 => \phi(x)f(x) = 0$. The parameters for the gamma distribution that were the best fit were 10 as the shape parameter and 1.3 as the scale parameter.

When the importance sampling had been determined, a loop was created so that the following process was done for each month separately with the appropriate lambda and k values in the Weibull process. These values were gathered into a vector and displayed in Table 4.

First, Weibull probability density functions were constructed for the month and samples of the gamma distribution were extracted. The target function was then constructed with the determined values and by taking the mean, the expected value was extracted. Then the standard deviation was calculated and the confidence interval was determined according to Equation 7.

| Month | Mean (W) | Lower bound (W) | Upper bound (W) | Width of confidence interval |
|---|---|---|---|---|
| January | 4658503 | 4649132 | 4667873 | 18 741 |
| February | 4137374 | 4120199 | 4154549 | 34 350 |
| March | 3836589 | 3814116 | 3859063 | 44 947 |
| April | 3025950 | 2996370 | 3055530 | 59 160 |
| May | 2854193 | 2823257 | 2885130 | 61 873 |
| June | 3072233 | 3043609 | 3100856 | 57 247 |
| July | 2853416 | 2822449 | 2884382 | 50 933 |
| August | 3090459 | 3062005 | 3118913 | 56 908 |
| September | 3745987 | 3722510 | 3769464 | 46 954 |
| October | 4215118 | 4203637 | 4226599 | 22 962 |
| November | 4655079 | 4645239 | 4664919 | 19 680 |
| December | 4662216 | 4652935 | 4671497 | 18 562 |
| Yearly | 3733926 | 3712954 | 3754897 | 41 943 |

Table 4: Estimation of the power with the expected value, lower and upper bounds of the 95% confidence interval including width for each month and the average as a yearly estimate using importance sampling

When comparing the width of the confidence intervals with the previous estimation methods, it is apparent that the average width is significantly lower than with any other method. With the importance sampling the average width is only about 42 000 W, compared to values between 62 000 and 134 000 for the other methods. This is a very large variance reduction, so it is clearly a better variance reduction technique than truncation or using control variates in this situation. It is, however, interesting to note that the width of the confidence intervals for the importance sampling method varies more between the months than the other methods do. The width of the confidence interval for the method using wind as a control variate is also significantly lower than for the standard Monte Carlo or the truncated version, which shows that that method also can accomplish significant variance reduction.

It is also interesting to observe the expected values between the methods, which, excluding the truncated version that is expected to have a higher value, are very similar to each other, especially the standard Monte Carlo and the control variate method. This gives credibility to the fact that importance sampling is a very good variance reduction method that has very good mean precision and a compact confidence interval.

### 3.4 d) Using antithetic sampling

Another variance reducing method is antithetic sampling. Antithetic sampling involves generating two sets of samples with negative correlation. The power curve is divided into two parts, with one being monotonously increasing and the other being constant. Due to this, it was concluded to do two different antithetic samplings. For the monotonously increasing part over the interval (3.5, 14), the wind was drawn from a Weibull distribution. First, half of the samples were drawn from a Weibull distribution with the scale and shape corresponding to the month. The other half of the samples for the first part were generated by equation 13. The correlation between the two halves of samples for the first part was negative. For the constant part over the interval (14, 25), half of the samples were drawn from a uniform distribution and the other samples were generated by taking 1 minus the uniform samples.

$$X_2 = \max(X_1) - X_1 \tag{13}$$

The mean of each of the parts were calculated by taking the mean of the power of the first half of the samples plus the power of the other half samples divided by two. Then the mean of the two part's mean was calculated. The variance of the mean was calculated using equation 14 and a confidence interval of 95% was created using equation 7.

$$\text{var} = \frac{1}{4} * (\text{var}[P(X_1)] + \text{var}[P(X_2)]) + \frac{1}{2} * \text{cov}(P(X_1), P(X_2)) \tag{14}$$

The expected values as well as the confidence interval for the expected power generated by the wind turbine using antithetic sampling can be observed in Table 5. It can be noted that the yearly expected value is considerably lower than for the previous estimation methods, about 800 000 W lower. This indicates that there is something dysfunctional with the antithetic sampling, as it is expected to have the same expected value as the other estimation methods excluding the truncated one. Additionally, the confidence intervals are much slimmer than the previous ones. Although the antithetic sampling can reduce variance, it is not possible to conclude whether it can be done to this extent or if there is something else at play. Therefore, there could possibly be an error in the code. Despite extensive trouble-shooting, the potential error could not be corrected so this leaves the question unanswered. This creates difficulty in comparing the values from the antithetic sampling with the other estimation methods. However, the discrepancy in the mean does indicate that there is at least something affecting that. In general, though, the method would be expected to reduce the variance and thus slim the confidence interval, especially compared to the standard Monte Carlo.

| Months: | Expected value (W): | Lower bound (W): | Upper bound (W): | Width: |
|---|---|---|---|---|
| January | 3004588 | 3004074 | 3005102 | 1028 |
| February | 3240518 | 3240070 | 3240967 | 897 |
| March | 3054766 | 3054318 | 3055214 | 896 |
| April | 3014854 | 3014461 | 3015248 | 787 |
| May | 2944902 | 2944506 | 2945299 | 793 |
| June | 3017191 | 3016792 | 3017589 | 797 |
| July | 2837342 | 2836942 | 2837743 | 801 |
| August | 2909630 | 2909217 | 2910043 | 826 |
| September | 3169425 | 3168992 | 3169858 | 866 |
| October | 2675179 | 2674644 | 2675714 | 1070 |
| November | 2918371 | 2917848 | 2918893 | 1045 |
| December | 2239015 | 2238470 | 2239560 | 1090 |
| Yearly | 2918815 | 2918361 | 2919269 | 908 |

Table 5: Estimation of the power with the expected value, lower and upper bounds of the 95% confidence interval including width for each month and the average as a yearly estimate using antithetic sampling

### 3.5 e) The probability of the turbine delivering power

In this part of the assignment, the purpose was to estimate the probability that the turbine delivers power $P(P(V) > 0$. As it had been noted that the wind turbine only provides power when the wind speeds are between 3.5 and 25 m/s, the relationship that can be seen in Equation 15 was leveraged where F(x) is described by Equation 8.

$$P(P(V) > 0) = P(3.5 \leq V \leq 25) = F(25) - F(3.5) \tag{15}$$

This relationship was applied to each month and the probabilities of the turbine generating power for each month can be observed in Table 6. It can be observed that the likelihood is larger for the winter months.

| Month | Probability of generating power (%) |
|---|---|
| January | 89.3 |
| February | 87.7 |
| March | 86.5 |
| April | 81.2 |
| May | 80.4 |
| June | 81.6 |
| July | 80.4 |
| August | 81.6 |
| September | 86.2 |
| October | 86.8 |
| November | 89.3 |
| December | 89.3 |

Table 6: Estimation of the probability that the wind turbine generates power for each month

### 3.6 f) The average ratio of the actual wind turbine output and the total wind power

This part of the assignment had the objective to create a 95% confidence interval for the average ratio of actual wind turbine output to total wind power, which is calculated with Equation 16. $EP_{\text{tot}}(V)$ is calculated from Equation 17.

$$\frac{EP(V)}{EP_{\text{tot}}(V)} \tag{16}$$

$$P_{\text{tot}}(v) = \frac{1}{2}\rho\pi\frac{d^2}{4}v^3 \quad (17)$$

In Matlab, the $EP_{\text{tot}}(V)$ was first calculated using Equation 18 to determine the expected wind speeds for each month and inserting this into Equation 17.

$$E[V^m] = \Gamma\left(1 + \frac{m}{k}\right)\lambda^m \quad (18)$$

| Month: | Mean ratio(%): | Lower bound(%): | Upper bound(%): |
|---|---|---|---|
| January | 22.64 | 22.29 | 22.99 |
| February | 26.24 | 25.80 | 26.68 |
| March | 28.22 | 27.71 | 28.72 |
| April | 31.95 | 31.28 | 32.62 |
| May | 32.69 | 31.98 | 33.39 |
| June | 32.32 | 31.66 | 32.98 |
| July | 33.43 | 32.72 | 34.13 |
| August | 31.72 | 31.06 | 32.37 |
| September | 29.45 | 28.92 | 29.97 |
| October | 24.03 | 23.63 | 24.43 |
| November | 22.71 | 22.36 | 23.06 |
| December | 22.67 | 22.32 | 23.02 |

Table 7: The ratio of average actual power generated and total wind power with lower and upper bounds of a 95% confidence interval

The $EP(V)$ was calculated by taking the mean of the power of the Weibull-simulated values for each month. The ratio was then calculated according to Equation 16. The confidence interval was, as previously, determined by the standard deviation and the p-value for the normal distribution of a 95% confidence interval according to Equation 7 and the standard deviation was calculated according to Equation 19.

$$\sigma = \frac{\sigma_{EP(V)}}{\sqrt{N} * P_{\text{tot}}(v)} \quad (19)$$

The ratios can be observed in Table 7. It can be noted that there is a moderate variability between the months in the ratio, where the summer months typically have a larger ratio.

## 3.7 g) The capacity factor and availability factor

In this part, the capacity factor and availability factor were calculated. The capacity factor is the ratio of the actual output over a time-period and the maximum possible output during the time-period, which is 9.5 MW per time-period for the Vestas V164. Typically, this is between 20% and 40% for wind turbines. The availability factor is the amount of time electricity is produced during a given time-period divided by the length of the period. This is typically above 90% for wind turbines. This ratio is very similar to the probability that was calculated in section 3.5, but it spans over a larger time-period.

In this task, the time-period was chosen to be 12 months. For the capacity factor, the vectors of the mean from the previous tasks were leveraged. The elements of the vectors were summed together and then divided by the time-period multiplied with the maximal output. This was done for the mean power output of the standard Monte Carlo and the truncated Monte Carlo, the mean using wind as a control variate, the mean using importance sampling, and the mean using antithetic sampling. The results can be observed in Table 8.

For the availability factor, the vector that stored the probability values from part 3.5 was utilized. In order to determine the amount of time power was produced during the whole year these values were summed together. This was then divided by the time-period. The result can be viewed in Table 9.

The value of 85% for the availability factor is quite reasonable, but on the low side as the value is typically above 90% for a wind turbine. This indicates that the site may not be optimal for a wind turbine, as one would prefer a larger availability factor.

| Monte Carlo estimation method: | Capacity factor (%): |
|---|---|
| Standard Monte Carlo (2a) | 39.3 |
| Truncated Monte Carlo (2a) | 46.0 |
| Control Variate Monte Carlo (2b) | 39.3 |
| Importance sampling Monte Carlo (2c) | 39.3 |
| Antithetic sampling Monte Carlo (2d) | 32.1 |

Table 8: Annual capacity factor of the different Monte Carlo estimation methods

| Availability factor: | 85.0% |
|---|---|

Table 9: Annual availability factor of the wind turbine

The capacity factor varies considerably between the estimation methods, but they are within the typical interval if not above it. For the truncated Monte Carlo, the capacity factor exceeds the typical 20-40% quite substantially with a factor of 46%. The average over all the estimation methods is 39.2% which is very high in the typical range.

Bearing this in mind, one can conclude that this site is not optimal for wind turbines, but it can however be satisfactory if there are no better options, as its capacity factor is very good and the availability factor is close to acceptable.

# 4 Part 3: Combined power production of two wind turbines

In this section of the assignment, an additional identical wind turbine was introduced to explore how the turbines affect each other.

## 4.1 a) The expected amount of combined power generated by both turbines

In this part, the problem was extended to determining the expected amount of combined power generated by both turbines together, E(P(V1) + P(V2)). In this case, it is possible to exploit the fact that they have the same distribution and thus have the same expected value. This implies that we can rewrite the problem as 2*E(P(V)) and that the problem is hence simplified to a one-dimensional plane.

This means that it is enough to estimate 2*E(P(V)) using importance sampling to solve this problem. It is possible to leverage the results from part 3.3 when determining this value. Since the parameter values for the Weibull distribution had been shifted from monthly values to yearly values, the correct probability density function was calculated. The new values were $k = 1.96$ and $\lambda = 9.13$ Then the power function was adjusted to represent two wind turbines and the importance sampling function was doubled as well to account for this change. These were then plotted together to ensure that they resembled each other, which can be observed in Figure 3. The figure shows that the distributions are quite similar, although they do not follow each other perfectly. Further, the ratio between them was plotted in Figure 4. Optimally, the ratio should be constant. In this case, the ratio fluctuated moderately, but the interval was still very small, between about 1.6 and 0.2. Thus, the instrumental distribution was deemed satisfactory.

Further, the confidence interval was calculated identically to in part 3.3, with the adjustment that $\phi(x) = 2 * P(x)$. The expected value and confidence interval can be observed in Table 10.

| Expected value (W): | Lower bound (W): | Upper bound (W): | Width: |
|---|---|---|---|
| 7134365 | 7084823 | 7183907 | 99 084 |

Table 10: The expected power generated by two wind turbines estimated with importance sampling

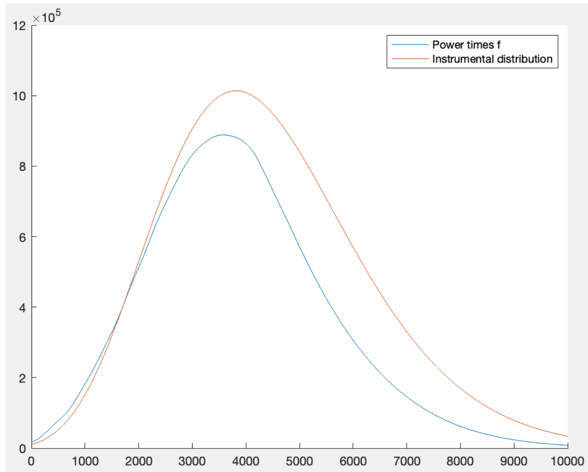


Figure 3: The instrumental distribution and $f(x) * \phi(x)$ in the bivariate case plotted against each other for January

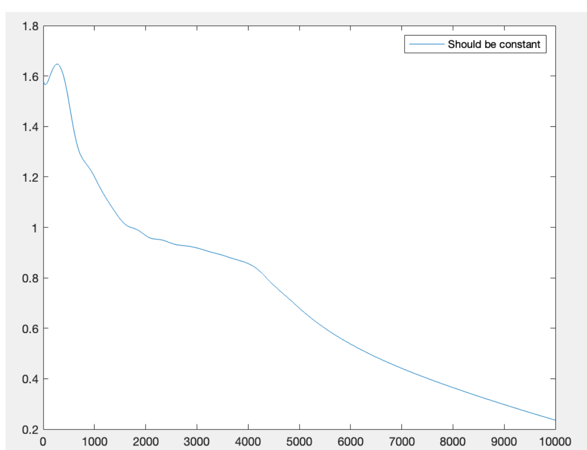


Figure 4: Plot of the ratio described by Equation 12 in the bivariate case for January

## 4.2 b) The covariance between the produced power in two identical wind power plants

The objective in this part of the assignment was to determine the covariance C(P(V1), P(V2)) between the produced power in two identical wind turbines receiving dependent winds. In order to calculate the covariance it had to be rewritten according to Equation 20 first.

$$C(P(V1), P(V2)) = E(P(V1)P(V2)) - E(P(V1))E(P(V2)) \tag{20}$$

When utilizing importance sampling the first term on the right-hand term can be rewritten according to Equation 21.

$$E(P(V1)P(V2)) = E_g(P(V1)P(V2)f(V1, V2)/g(V1, V2)) \tag{21}$$

In order to determine the covariance, the values of E(P(V1)) and E(P(V2)) had to be calculated. As they are identically distributed, the expected values are the same. This was determined in the same way as part 3.3 but with the appropriate values of $\lambda$ and k. Additionally, the variance was determined which will be useful for part 4.3.

The next step was to determine the instrumental distribution to be able to perform importance sampling. In a 3D plot the bivariate Weibull distribution multiplied with P(V1)*P(V2) was plotted against different instrumental distributions, see Figure 5. It was difficult to find an instrumental distribution that followed the other curve well in 3D but a reasonable function was determined, see Matlab Equation 22. Then the mean was calculated using Equation 12, with P(V1)*P(V2) as $\phi$, the bivariate Weibull distribution as f and the instrumental distribution as g.

$$g(V_1, V_2) = \text{gampdf}(V_1, 10, 1.3)' * \text{gampdf}(V_2, 25, 0.9) \tag{22}$$

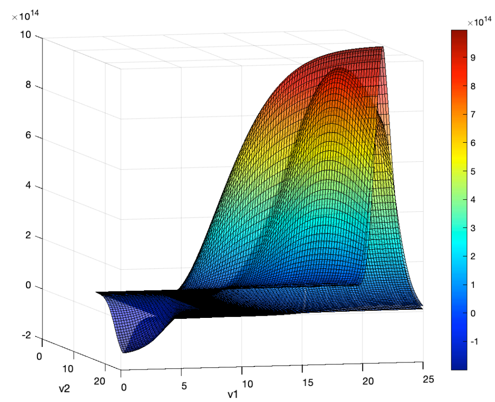


Figure 5: The instrumental distribution against bivariate Weibull distribution times the power of V1 times the power of V2

The covariance between the produced power from two identical wind power plants receiving dependent winds was estimated to approximately 55.9*e24 using importance sampling.

During the calculations of the covariance it became clear that the values did not always seem reasonable. Large amounts of trouble-shooting were conducted but to no avail. In order to understand the range of values that was reasonable a standard Monte Carlo simulation of the covariance was conducted based on Equation 20. In this process, wind values were simulated from the Weibull distribtion and inserted into the power curve. The expected values of E(P(V1)), E(P(V2)), and E(P(V1)*P(V2)) were calculated and then inserted into the equation. The estimated covariance from the standard Monte Carlo was 144 921 476 960, which was vastly different from the obtained value using importance sampling, and indicates an error in the calculation. However, this error could not be determined.

## 4.3 c) The variability in the amount of combined power

The purpose of this part was to estimate the variability V(P(V1)+P(V2)) in the amount of combined power generated by both wind turbines, as well as the standard deviation of the same case.

Here, it is possible to utilize the relationship between the variances and covariance described in Equation 23. Further, it can be noted that the relationship between the variance and standard deviation is described by Equation 24.

$$V(P(V1) + P(V2)) = V(P(V1)) + V(P(V2)) + 2 * C(P(V1), P(V2)) \tag{23}$$

$$D(P(V1) + P(V2)) = \sqrt{V(P(V1) + P(V2))} \tag{24}$$

As the variance for the wind turbines had already been calculated in part 4.2, this value could be reused as V(P(V1)) and (P(V2)) as they are identically distributed. Additionally, the covariance was calculated in part 4.2 as well. This means that the previously determined values can be inserted into Equation 23 and 24 respectively. The obtained values can be observed in Table 11. It is apparent that these values are quite unreasonable, which indicates that something has gone wrong in the calculations. Despite intense trouble-shooting, the error could not be found.

In order to find a reasonable value, a standard Monte Carlo approach to the problem was implemented. The covariance that had been found in part 4.2 was inserted into Equation 23 as well. The standard devation and variance frmo this operation are found in Table 12. It is apparent that the latter variance and standard deviation are more reasonable, indicating some type of miscalculation along the way.

| | |
|---|---|
| Variance: | 111811077341903747398238208 |
| Standard deviation: | 10574075720454W |

Table 11: The variance and standard deviation of the combined power production using importance sampling

| | |
|---|---|
| Variance: | 2599189602318 |
| Standard deviation: | 1612200W |

Table 12: The variance and standard deviation of the combined power production using standard Monte Carlo

## 4.4 d) The probability that the combined power generated by the two turbines is greater than half of their installed capacity

In this part the probability that two power turbines in the same area produce more than 9.5MW power together and the probability that they produce less than 9.5MW were calculated. The first probability is

$$P(P(V_1) + P(V_2) > 9.5MW)$$

. Since the two wind turbines are located in the same area, it is assumed that they experience the same wind at the same time. Therefore, the problem can be simplified to

$$P(2P(V) > 9.5MW)$$

. This expression can be rewritten to

$$P(V > P^{-1}(9.5/2MW))$$

. To find what V value solves the equation

$$V = P^{-1}(9.5/2MW)$$

, the power function was plotted, see figure 7. After carefully observing the graph and testing different inputs to P, the value V = 9.2156 was found to solve the equation. For the power generated by both power turbines to be greater than 9.5 MW, the wind has to be between 9.2156 m/s and 25 m/s. Therefore, the probability can be written as

$$P(V < 25m/s) - P(V < 9.2156m/s)$$

. Then an importance sampling on this probability was done by using equation 11, with the probability as $\phi(x)$, and finding a convenient instrumental distribution. A convenient instrumental distribution, equation 25, was found by plotting different gamma functions against $\phi(V)$*f(V), see figure 6. When an appropriate function had been created the ratio in Equation 12 was plotted and was deemed enough constant to continue. A 95% confidence interval was created using Equation 7.

$$g(V) = \text{gampdf}(V, 4.1, 2) \tag{25}$$

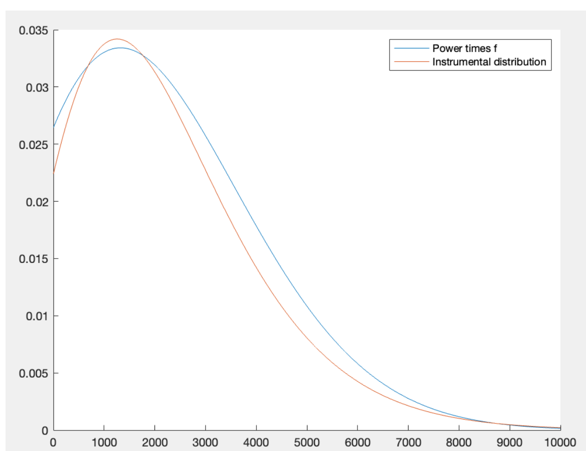


Figure 6: Plot of the instrumental distribution and $f(x) * \phi(x)$ in the bivariate case for exercise 3d)

For the opposite probability, the same procedure was performed, but with the probability

$$P(V > 25m/s) + P(V < 9.2156m/s)$$

. The same instrumental distribution as for the previous probability was used since it followed the curve well, see Figure 9 and 10.

Since the Weibull input is sampled, the probabilities will not always be the same each time the code is run. Therefore, the two different probabilities will not always equal to one, because of the variance in the sampling. Further, the sum of these two probabilities misses the probability that it is exactly 9.5MW since it is either larger or smaller than that.

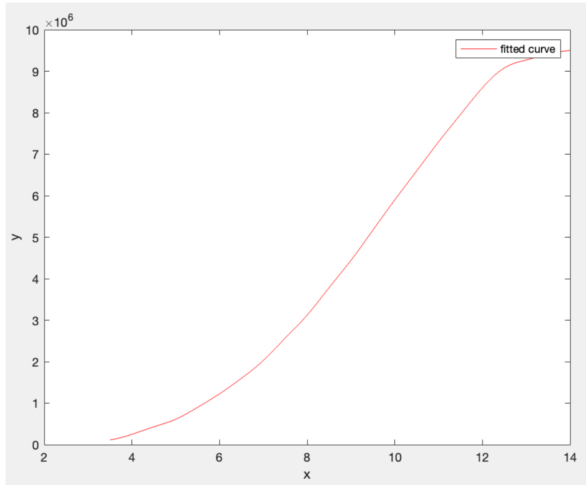


Figure 7: Fitted curve

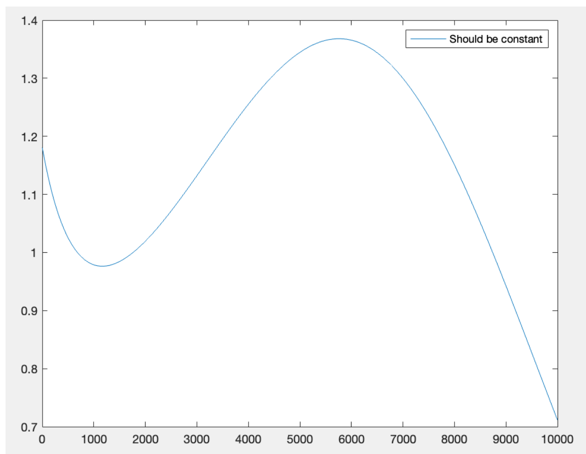


Figure 8: Ratio described by Equation 12 for the bivariate case in exercise 3d)

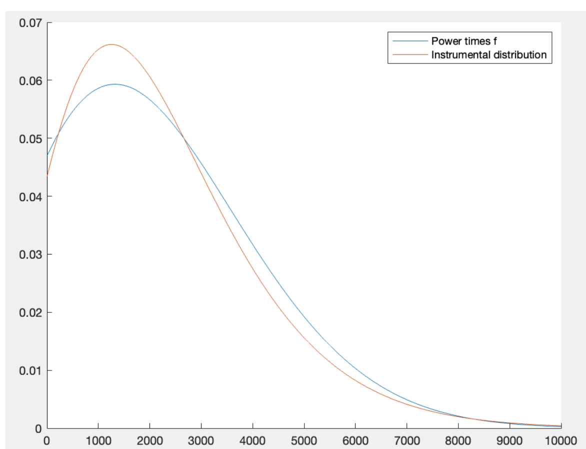


Figure 9: The opposite probability

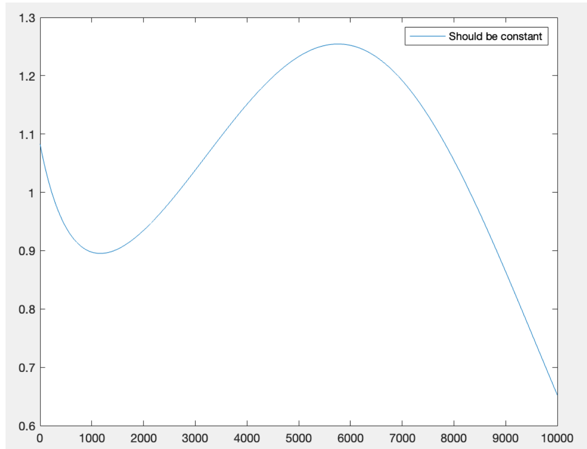


Figure 10: The opposite probability 3

# 5 Final words

In this assignment, various statistical techniques to estimate the expected power production of a wind turbine were explored, using Monte Carlo simulation, importance sampling, control variates, and antithetic sampling. By leveraging different variance reduction methods, we were able to enhance the precision of our estimates and compare the effectiveness of these approaches.

This assignment highlights the importance of advanced statistical methods in fields like renewable energy modeling, demonstrating how computational techniques can improve decision-making.

All MATLAB implementations, including the main script (proj1.m) and supporting files, have been submitted via email, and this report has been uploaded in PDF format to CANVAS.