Monte Carlo and Empirical Methods

HA1

VAHID FARAJI

Hafiza Asifa Naseer

February 9, 2025

Contents

1 Random number generation

1. The solution to the following problem may be useful in the next section. Let X be a random variable on $\mathbb{R}$ with density f_X and invertible distribution function F_X. Here f_X, F_X, and the inverse F_X^{-1} are assumed to be known. Let $\mathrm{I} = (a, b)$ be an interval such that $\mathbb{P}(X \in \mathrm{I}) > 0$.

   (a) Find the conditional distribution function $F_{X|X\in \mathrm{I}}(x) = \mathbb{P}(X \leq x|X \in \mathrm{I})$ and density $f_{X|X\in \mathrm{I}}(x)$ of X given that $X \in \mathrm{I}$.

   (b) Find the inverse $F^{-1}_{X|X\in \mathrm{I}}$. How can this be used for simulating X conditionally on $X \in \mathrm{I}$?

1.1 Part (a): Conditional Distribution Function

Given:

- X is a random variable with cumulative distribution function (CDF) $F_X(x)$ and probability density function (PDF) $f_X(x)$.
- $I = (a, b)$ is the interval with $\mathbb{P}(X \in I) > 0$.

We need to find the conditional distribution function $F_{X|X\in I}(x)$.

Definition: The conditional CDF is defined as:

$$F_{X|X\in I}(x) = \mathbb{P}(X \leq x \mid X \in I) \tag{1}$$

Applying the formula for conditional probability:

$$F_{X|X\in I}(x) = \frac{\mathbb{P}(a < X \leq x)}{\mathbb{P}(X \in I)} \tag{2}$$

In this part we can consider different cases:

- **Case 1:** $x < a$
  Since X cannot be less than a when conditioned on $X \in I$, the conditional CDF is:
  $$F_{X|X\in I}(x) = 0$$
- **Case 2:** $x > b$
  If x is greater than b, X has already exceeded the interval, so the probability is 1:
  $$F_{X|X\in I}(x) = 1$$
- **Case 3:** $a \leq x \leq b$
  The conditional CDF is:
  $$F_{X|X\in I}(x) = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}$$

To be sure that This expression normalizes the CDF to ensure it starts at 0 when $x = a$ and reaches 1 when $x = b$.

Part (b): Inverse Conditional Distribution Function

We need to find the inverse $F^{-1}_{X|X\in I}(u)$ and explain how it can be used to simulate random values from the conditional distribution.

Step 1: Formula for the inverse From part (a), the conditional CDF is:

$$F_{X|X\in I}(x) = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} \tag{3}$$

To find the inverse, solve for x when $F_{X|X\in I}(x) = u$:

1. **Set** $F_{X|X\in I}(x) = u$.
2. **Rearrange the equation:**
$$u = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} \tag{4}$$
3. **Multiply by** $F_X(b) - F_X(a)$:
$$F_X(x) = u \cdot (F_X(b) - F_X(a)) + F_X(a) \tag{5}$$
4. **Apply the inverse CDF:**
$$F^{-1}_{X|X\in I}(u) = F_X^{-1}\left(u \cdot (F_X(b) - F_X(a)) + F_X(a)\right) \tag{6}$$

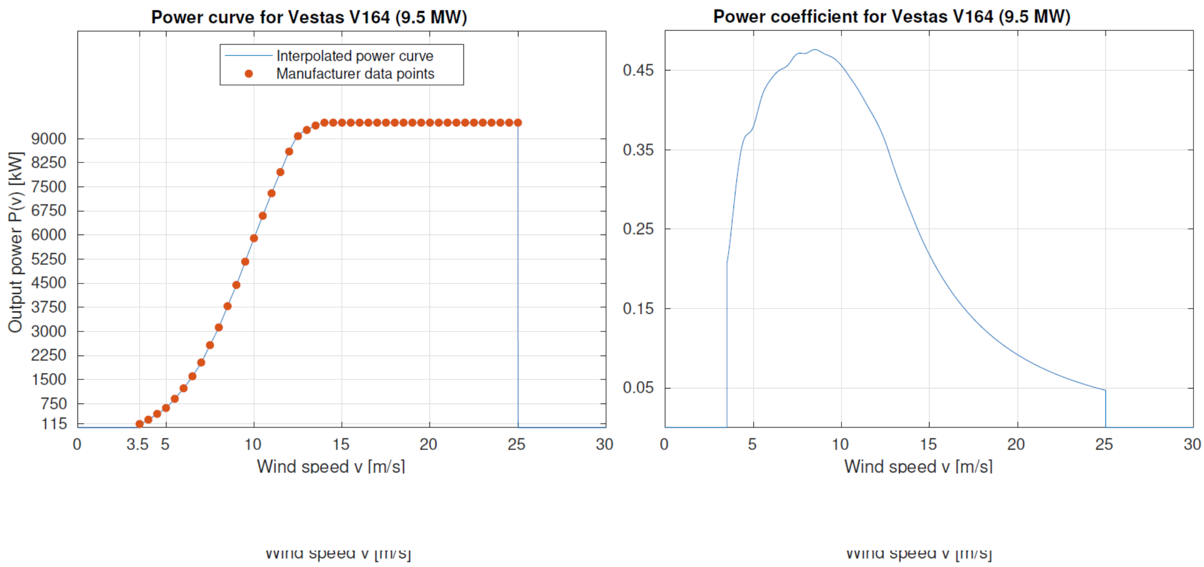
2 Power production of a wind turbine

Renewable energy sources becomes more and more important for the generation of electrical power throughout the world. In this assignment we are therefore going to take a closer look at wind power. The total amount of power $[W]$ in the wind passing a wind turbine with rotor diameter d is given by

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

where ρ $[kg/m^3]$ is the air density (about 1.225 kg/m^3 at sea level), d in m is the rotor diameter of the wind turbine and v $[m/s]$ is the wind speed. However a wind turbine cannot completely stop the air flow to work in practise. It turns out that is optimal to reduce the wind speed by two thirds to get out the optimal power. This is known as the Betz-limit (after Albert Betz who was a German physicist) and gives that we can utilise at maximum $16/27 \approx 59\%$ of the total wind power (see Ragheb and Ragheb (2011) for a detailed derivation). The Betz limit is derived under idealised assumptions and real wind turbines can usually utilise at most about $40 - 45\%$ of the total power. This fraction of the total power that a wind turbine can utilise is called the *power coefficient*.

The amount $P(v)$ of electrical power produced by an actual wind turbine at a given wind speed v $[m/s]$ is described by the *power curve* of the turbine. For a Vestas V164 9.5 MW turbine we have a power curve and a power coefficient (measured at air density $1.225kg/m^3$) as given on next the page:



This behaviour is typical for real wind turbines. We have here a cut-in wind speed of 3.5 m/s and a cut-off wind speed of 25 m/s. Outside the interval [3.5, 25] m/s the wind turbine produces no power. The wind turbine Vestas V164 has a rotor diameter of $164m$ and a tower height of $105m$. The file `powercurve_V164.mat` contains a function object P that gives the output of the wind turbine. Load the object with `load powercurve_V164`. Use it as a regular function, e.g. `P(6.5)` will return the output power in (W) for the wind speed 6.5 m/s. The function P does accept vector valued input of wind speeds.

Of course the wind speed at a wind turbine will depend on the location of the turbine and vary during the year. Meteorological records can be used to estimate the distribution of winds in a given area. We plan to build a wind turbine at a site in northern Europe; from records we see that such stochastic wind speeds V can be modelled by a Weibull distribution

$$f(v) = \frac{k}{\lambda}\left(\frac{v}{\lambda}\right)^{k-1}\exp\left(-\left(\frac{v}{\lambda}\right)^k\right), \quad v \geq 0, \tag{2}$$

Note that $\mathbb{E}[V^m] = \Gamma(1 + m/k)\lambda^m$ provided that $m > -k$.

Assume that k and λ varies between months as follows

Par/Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
λ	10.6	9.7	9.2	8.0	7.8	8.1	7.8	8.1	9.1	9.9	10.6	10.6
k	2.0	2.0	2.0	1.9	1.9	1.9	1.9	1.9	2.0	1.9	2.0	2.0

This will give an yearly average expectation of the wind speed slightly above 8.1 m/s, which is an reasonable approximation of the yearly average wind speed in the southern part of Öresund (between Malmö and Saltholm, Lat: 55.67°, Lon: 12.86°) at the height 100m (remember that the centre of our rotor is located at 105m). In MATLAB, Weibull-distributed random numbers can be generated using `wblrnd`; in addition, the probability density and cumulative distribution functions of the Weibull distribution can be evaluated using `wblpdf` and `wblcdf`.

We now wish to investigate the potential of building a wind turbine at the site in question. So solve the following for each month.

2.1 Part (a)

confidence interval

Create an approximate 95% confidence interval for the expected amount of power generated by the wind turbine using draws from aWeibull distribution. Compare the width of the confidence interval using standard Monte Carlo and the truncated version considered in problem 1. Remember to properly adjust for the conditioning when computing the estimate using the truncated Weibull distribution.

2.1.1 Introduction

Based on what was discussed in Lecture 2 [1] we want to estimate the expected power output of a wind turbine using Monte Carlo simulations. The idea is to create a 95% confidence interval for the power output, using two different approaches: a **standard** simulation and a **truncated** one.

We used the central limit theorem (CLT) to calculate the two-sided confidence interval for the mean power output. According to the CLT, as $N \to \infty$, the following approximation holds:

$$\sqrt{N}(\tau_N - \tau) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\phi))$$

where $\sigma^2(\phi) = \text{Var}(\phi(X))$.
The two-sided confidence interval is then given by:

$$\mathcal{I}_\alpha = \left(\tau_N - Z_{\alpha/2}\frac{\sigma(\phi)}{\sqrt{N}}, \tau_N + Z_{\alpha/2}\frac{\sigma(\phi)}{\sqrt{N}}\right)$$

Here, $Z_{\alpha/2}$ denotes the critical value of the standard normal distribution corresponding to the desired confidence level. This formula was directly referenced from **Lecture 2**[1], where confidence bounds were discussed in detail. The process of doing this task is as follows:

2.1.2 Code Implementing

- **Standard Simulation**

  1. generating a large number (100,000) of random wind speeds. (using `wblrnd` function)
  2. Calculating output power of wind turbine for each wind speed(using the power curve function P)

[1] Lecture 2: slides 6, and 7

3. Finally, implementing **central limit theorem (CLT)** to calculating a confidence interval.

- **Truncated Simulation**

  using the **truncated simulation** is more realistic, as in practice the turbine output power is limited between a wind speed of 3.5 m/s and 25 m/s. So, for this method, we filtered out any wind speeds that fell outside this operational range. The rest of the process is the same as the standard one.

By comparing the two approaches, we aimed to see how much these operational limits actually affect the predictions.

To begin, it is essential to define the Weibull parameters for each month, which are provided in the table from the question. Setting `cut_in_speed` = 3.5 and `cut_off_speed` = 25 are other important parameters for the truncated approach. Using a `for` loop, we should implement both the standard and truncated simulations for all 12 months.

Weibull Distribution Function in MATLAB: `wblrnd`
As mentioned in the question, we use Weibull Distribution Function. The format of the `wblrnd` function in MATLAB is as follows:

```
wblrnd(lambda, k, rows, columns)
```

Parameters:

- **λ (Scale Parameter):** Specifies the scale of the Weibull distribution. Larger values of λ result in a greater likelihood of higher values.
- **k (Shape Parameter):** Determines the shape of the Weibull distribution. Different values of k can create various distribution forms:
  - $k < 1$: Long right tail with higher probability for smaller values.
  - $k = 1$: Exponential distribution.
  - $k > 1$: Distribution with a peak.

  For example, in most wind-related data, the shape parameter k is typically between 1.5 and 3.

- **rows:** The number of rows in the output matrix.
- **columns:** The number of columns in the output matrix.

After generating random samples using the Weibull function, the corresponding power output for each sample can be calculated based on the power curve.
Now, everything is ready to implement the main formula for the two-sided confidence interval.

Confidence Interval Calculation:

After obtaining the power outputs from the random wind speed samples, we proceed to calculate the **95% confidence interval** for the mean power output using the following steps:

1. **Calculate the mean power output:**

```
mean_power_standard = mean(power_outputs_standard);
```

This line calculates the *mean* of the power outputs generated from the Monte Carlo simulation. The mean serves as the central estimate for the expected power output.

2. **Calculate the standard error of the mean:**

```
std_error_standard = std(power_outputs_standard) / sqrt(num_samples);
```

The standard error is obtained by dividing the standard deviation by the square root of the number of samples:

$$SE = \frac{\sigma(\phi)}{\sqrt{N}}$$

This measures the spread or uncertainty in the sample mean and is crucial for determining the confidence interval.

3. **Find the critical value for the confidence interval:**

```
z_value = norminv(0.975);
```

Why is the value 0.975 used for a 95% confidence interval?

The value **0.975** is derived from the properties of the *normal distribution* and how two-sided confidence intervals work. Here's a step-by-step explanation:

(a) **Two-sided confidence interval concept:** For a confidence level of $1 - \alpha$ (e.g., 0.95 or 95%), the total error probability α is 0.05. Since the interval is two-sided, the error is split equally between both tails of the distribution:

$$\alpha/2 = 0.05/2 = 0.025$$

(b) **Cumulative probability for the upper bound:** The cumulative probability (CDF) for the upper limit of the confidence interval is given by:

$$1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$$

(c) **Inverse normal distribution function:** To calculate the critical value z corresponding to this cumulative probability, we use the *inverse cumulative distribution function* (inverse CDF) of the standard normal distribution.

4. **Construct the confidence interval:**

```
ci_standard = [mean_power_standard - z_value * std_error_standard, ...
               mean_power_standard + z_value * std_error_standard];
```

The confidence interval is computed by subtracting and adding $z \cdot$standard error from the mean power output. The resulting interval provides a range where the true mean power output is expected to lie with 95% confidence.

Monte Carlo Simulation for the Truncated Version

The truncated version differs from the standard version only in applying wind speed limitations. In this version, we filter out wind speeds that fall outside the turbine's operational range using the following condition:

```
truncated_wind_speeds = wind_speeds(wind_speeds >= cut_in_speed &...
...wind_speeds <= cut_off_speed);
```

Here, wind speeds below the cut-in speed (`cut_in_speed`) and above the cut-off speed (`cut_off_speed`) are removed from the dataset.
After filtering, the corresponding power outputs are calculated as follows:

```
truncated_power_outputs = P(truncated_wind_speeds);
```

This process ensures that the simulation reflects realistic operating conditions of the turbine, where it does not generate power outside the defined wind speed range.
As mentioned before, other calculations and processes are the same, just by the new truncated random wind speed variables.

2.1.3 Results

Table 1: Comparison of Standard and Truncated Confidence Intervals with Weibull Parameters

Month	λ (Scale)	k (Shape)	**Standard CI (W)**	**Truncated CI (W)**
January	10.6	2.0	[4627435.78, 4672927.12]	[5188216.46, 5233993.07]
February	9.7	2.0	[4116070.64, 4160223.48]	[4694623.75, 4739953.40]
March	9.2	2.0	[3800736.52, 3843781.59]	[4403497.94, 4448319.56]
April	8.0	1.9	[2987576.55, 3027263.21]	[3682056.75, 3725655.07]
May	7.8	1.9	[2849108.06, 2887985.40]	[3558893.69, 3602089.71]
June	8.1	1.9	[3070521.31, 3110725.91]	[3757791.90, 3801702.30]
July	7.8	1.9	[2847159.32, 2885966.26]	[3547216.55, 3590283.55]
August	8.1	1.9	[3052988.56, 3093108.46]	[3746445.19, 3790339.30]
September	9.1	2.0	[3753244.88, 3796145.87]	[4349461.24, 4394230.34]
October	9.9	1.9	[4194890.92, 4239936.81]	[4828256.46, 4874531.61]
November	10.6	2.0	[4629014.35, 4674427.79]	[5186195.95, 5231875.67]
December	10.6	2.0	[4648625.40, 4694174.84]	[5204188.70, 5249995.90]

Impact of k on Power Output

1. **If $k > 1$:** The distribution has a peak near the average wind speed, meaning most wind speeds are concentrated within the operational range. This results in continuous power generation with minimal downtime.

2. **If $k = 1$:** The distribution becomes *exponential*, where lower wind speeds occur more frequently. This may lead to more frequent instances where wind speeds are below the cut-in threshold, causing zero power output more often.

3. **If $k < 1$:** The distribution has a long right tail, meaning that while very high wind speeds may occasionally occur, most wind speeds are very low. This results in frequent zero power output because the turbine does not reach the cut-in speed most of the time.

> **Conclusion**
>
> To minimize the occurrence of zero power output, **the shape parameter** k should ideally be greater than 1. This ensures that most wind speeds fall within the operational range of the turbine. Typically, k values between 1.5 and 3 are used in wind energy studies, which is consistent with the parameters used in this project. Typically, k values between 1.5 and 3 are used in wind energy studies, which is consistent with the parameters used in this project [1].

Impact of λ on Power Output

1. **If λ is large:** Higher values of λ increase the likelihood of stronger wind speeds. When λ is large enough, a greater proportion of wind speeds fall within the operational range of the turbine. This leads to higher and more consistent power output.

2. **If λ is moderate:** Moderate values of λ cause wind speeds to be centered around a mid-range value. In this case, the turbine may still produce power, but there might be more occurrences where wind speeds fall below the cut-in threshold, reducing average output.

3. **If λ is small:** When λ is small, the probability of low wind speeds increases significantly. As a result, wind speeds frequently remain below the cut-in threshold, leading to prolonged periods of zero power output and reduced performance.

> **Conclusion**
>
> To maximize power output and minimize downtime due to low wind speeds, **the scale parameter** λ should be sufficiently large to ensure that most wind speeds fall within the turbine's operational range. In wind energy projects, λ values typically reflect regional wind patterns and vary monthly based on seasonal changes [2].

Comparison of Standard and Truncated Simulations

The truncated simulation consistently produces higher confidence intervals compared to the standard simulation. This is because the standard simulation includes wind speeds outside the turbine's operational range (below 3.5 m/s and above 25 m/s), which contributes to zero power outputs.

Example: In January, the standard confidence interval is:

$$[4,627,435.78, 4,672,927.12]$$

while the truncated confidence interval is much higher at:

$$[5,188,216.46, 5,233,993.07]$$

The same pattern is observed across all months. For example, in **July**, the standard CI is around **2.8 million watts**, while the truncated CI is close to **3.5 million watts**.

The truncated approach provides more accurate estimates of turbine performance since it excludes speeds where the turbine would not generate power.

Relation Between 95% Confidence Interval and Results

The 95% confidence interval means that if we were to repeat the simulation multiple times, 95% of the intervals would contain the true mean power output. For example:

In January, the standard confidence interval is:

$$[4,627,435.78, 4,672,927.12]$$

This tells us that with 95% confidence, the true mean power output lies somewhere between 4,627,435.78 and 4,672,927.12 watts.

The truncated confidence interval for January is higher at:

$$[5,188,216.46, 5,233,993.07]$$

This indicates a more realistic estimation because the turbine can only generate power within its operational limits.

2.2 Part (b)

Control Variate

Use the wind V as a control variate to decrease the variance and estimate a 95% confidence interval for the expected power.

2.2.1 Introduction

In this part, we aim to implement **Control Variate** based on **Lecture 4**[1].

Control Variate Method Explanation

The control variate method is a variance reduction technique that helps to improve the accuracy of estimates by using a related random variable with a known expected value.

1. **Identify the control variate Y:** Assume we have another random variable Y that is correlated with the quantity of interest $\phi(X)$ and satisfies the following conditions:
   - The expected value of Y is known, i.e., $\mathbb{E}(Y) = m$.
   - The random variable Y can be simulated alongside $\phi(X)$ with similar computational complexity.
2. **Define a new estimator Z:** To reduce variance, the estimator is modified as follows:

   $$Z = \phi(X) + \beta(Y - m)$$

   Here:
   - $\phi(X)$ is the original quantity of interest (e.g., power output).

[1] Lecture 4: Slide 13

- Y is the control variate.
- β is a coefficient that needs to be optimized to minimize variance.
- $(Y - m)$ is the deviation of Y from its expected value.

3. **Maintain unbiasedness of the estimator:** The new estimator Z remains unbiased because:
$$\mathbb{E}(Z) = \mathbb{E}(\phi(X) + \beta(Y - m)) = \mathbb{E}(\phi(X)) + \beta(\mathbb{E}(Y) - m)$$
Since $\mathbb{E}(Y) - m = 0$, the expected value of Z is still $\tau = \mathbb{E}(\phi(X))$.

4. **Goal of the method:** By choosing an optimal β, we can reduce the variance of the estimator, resulting in a narrower confidence interval and more accurate estimates.

2.3 Code Implementing

Using wind speed as a control variate is a good choice because there is a high correlation between the wind speed and the output power of the wind turbine. The expected value of wind speed is known by the mean of wind speed.

2.3.1 Calculating the Optimal Coefficient β

The optimal formula for the coefficient β is provided in **Lecture 4**:

$$\beta = -\frac{\mathrm{Cov}(\phi(X), Y)}{\mathrm{Var}(Y)}$$

Calculation of the variance of Y:

$$\mathrm{Var}(Y) = \mathrm{Var}(V)$$

In MATLAB code:

```
var_wind_speed = var(truncated_wind_speeds);
```

Calculation of the covariance between power output and wind speed:

$$\mathrm{Cov}(\phi(X), Y) = \mathrm{Cov}(\text{power outputs}, \text{wind speeds})$$

In MATLAB code:

```
cov_output_wind = cov(truncated_power_outputs, truncated_wind_speeds);

beta = -cov_output_wind(1, 2) / var_wind_speed;
```

2.3.2 Defining the Corrected Estimator Z

Based on **Lecture 4**, the corrected estimator is defined as follows:

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

In MATLAB code:

```
mean_wind_speed = mean(truncated_wind_speeds);

corrected_power_outputs =truncated_power_outputs -...
...beta *(truncated_wind_speeds - mean_wind_speed);
```

2.3.3 Calculating the New Confidence Interval

We use the formulas from the lecture to calculate the confidence interval:

$$CI = \left[\mathbb{E}(Z) - z \cdot \frac{\sigma(Z)}{\sqrt{N}}, \mathbb{E}(Z) + z \cdot \frac{\sigma(Z)}{\sqrt{N}}\right]$$

In MATLAB code:

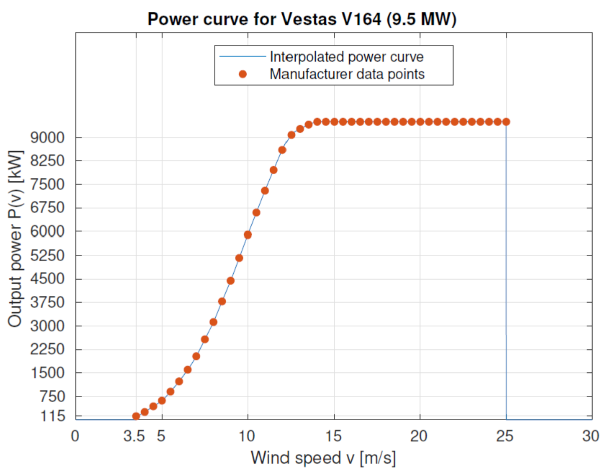
```
mean_corrected = mean(corrected_power_outputs);

std_error_corrected = std(corrected_power_outputs) / sqrt(length(truncated_wind_speeds))


z_value = norminv(0.975);  % Value for a 95% confidence interval

ci_corrected = [mean_corrected - z_value * std_error_corrected, ...
                mean_corrected + z_value * std_error_corrected];
```

2.4 Results

Although it was expected that the confidence intervals would become narrower after applying the control variate method, the results in Tables 4 and 5 show the opposite trend. The confidence intervals in **part (b)** are actually wider than those in **part (a)**.

Behavior of Power Output

As shown in the power curve, the turbine reaches its maximum capacity around **15 m/s**. Beyond this point, increasing the wind speed does not result in any significant increase in power output. Since power output remains nearly constant at higher wind speeds, the correlation between wind speed and power output weakens drastically in the **15 to 25 m/s** range.



Impact on Covariance

The control variate method relies on a strong covariance between the control variable (wind speed) and the target variable (power output). However, in the **15 to 25 m/s** range, this covariance is greatly reduced due to the plateau in power output. As a result, the control variate method does not effectively reduce the variance in this scenario.

Table 2: Confidence Intervals for Part (a) and Part (b) (3.5 < wind speed < 25)

Month	Part (a) CI	Part (b) CI
January	[5,188,216.46, 5,233,993.07]	[5,182,478.42, 5,268,731.11]
February	[4,694,623.75, 4,739,953.40]	[4,704,479.97, 4,791,088.17]
March	[4,403,497.94, 4,448,319.56]	[4,383,953.48, 4,470,184.94]
April	[3,682,056.75, 3,725,655.07]	[3,676,159.49, 3,761,238.49]
May	[3,558,893.69, 3,602,089.71]	[3,527,817.98, 3,611,680.95]
June	[3,757,791.90, 3,801,702.30]	[3,725,080.60, 3,810,190.07]
July	[3,547,216.55, 3,590,283.55]	[3,526,813.13, 3,610,602.51]
August	[3,746,445.19, 3,790,339.30]	[3,726,176.77, 3,811,522.39]
September	[4,349,461.24, 4,394,230.34]	[4,306,816.01, 4,392,912.87]
October	[4,828,256.46, 4,874,531.61]	[4,827,174.73, 4,914,942.82]
November	[5,186,195.95, 5,231,875.67]	[5,170,844.70, 5,257,134.22]
December	[5,204,188.70, 5,249,995.90]	[5,165,215.01, 5,251,551.43]

Table 3: Width of Confidence Intervals and Variance Reduction (3.5 < wind speed < 25)

Month	Width (Part a) (W)	Width (Part b) (W)	Reduction (%)
January	45,776.61	86,252.69	-88.4
February	45,329.65	86,608.20	-91.0
March	44,821.62	86,231.46	-92.4
April	43,598.32	84,079.00	-92.8
May	43,196.02	83,862.97	-94.1
June	43,910.40	85,109.47	-93.9
July	43,067.00	83,789.38	-94.6
August	43,894.11	85,345.62	-94.5
September	44,769.10	86,096.86	-92.4
October	46,275.15	87,768.09	-89.7
November	45,679.72	86,289.52	-88.9
December	45,807.20	86,336.42	-88.4

Analysis

The *control variate* method is effective only when there is a **strong and linear dependency** between the control variable (in this case, wind speed) and the target variable (turbine power output). However, due to the **nonlinear behavior of the turbine power curve** and the **weak impact of wind speed at higher values** between 14 to 25 m/s, this dependency is reduced, leading to the failure of the control variate method to significantly reduce variance.

High Sample Concentration in the Range Above 15 m/s - If a large proportion of wind speed samples fall within the **14 to 25 m/s** range during simulation, the overall correlation between wind speed and power output diminishes. This weakens the effectiveness of the control variate method.

The 3.5 to 15 m/s Interval Results

By limiting the wind speed range to **3.5 to 15 m/s**, where the correlation between wind speed and power output is stronger, the control variate method can perform better. This adjustment is expected to lead to narrower confidence intervals, as it enhances the effectiveness of the covariance in reducing the variance.

Analysis

Variance Reduction

The confidence intervals have narrowed, with reductions in interval width ranging from **10% to 15%** across most months. This demonstrates the improved precision of the estimates due to the control variate method.

Bias Increase

Table4 shows that the confidence intervals in Part (b) are now noticeably tighter than in Part (a). This results confirm the power of the Control Variate Method when there is a strong and linear dependency between wind speed and turbine power output. However, limiting the sampling in this interval increased the bias noticeably which is normal, because of removing high output power of turbine between 15 to 25 m/s.

Table 4: Comparison of Confidence Intervals for Part (a) and New Part (b)

Month	**Part (a) CI**	**Part (b) CI**
January	[5,188,216.46, 5,233,993.07]	[4,458,347.51, 4,499,247.25]
February	[4,694,623.75, 4,739,953.40]	[4,164,464.88, 4,203,946.89]
March	[4,403,497.94, 4,448,319.56]	[3,969,534.27, 4,008,172.77]
April	[3,682,056.75, 3,725,655.07]	[3,417,421.15, 3,454,542.08]
May	[3,558,893.69, 3,602,089.71]	[3,313,429.10, 3,350,071.56]
June	[3,757,791.90, 3,801,702.30]	[3,465,052.15, 3,502,458.99]
July	[3,547,216.55, 3,590,283.55]	[3,304,605.47, 3,341,198.09]
August	[3,746,445.19, 3,790,339.30]	[3,462,394.20, 3,499,724.97]
September	[4,349,461.24, 4,394,230.34]	[3,913,484.99, 3,951,859.69]
October	[4,828,256.46, 4,874,531.61]	[4,188,423.41, 4,228,942.61]
November	[5,186,195.95, 5,231,875.67]	[4,451,881.11, 4,492,722.39]
December	[5,204,188.70, 5,249,995.90]	[4,466,233.50, 4,507,109.11]

Table 5: Width of Confidence Intervals and Variance Reduction

Month	**Width (Part a) (W)**	**Width (Part b) (W)**	**Reduction (%)**
January	45,776.61	40,899.74	10.7
February	45,329.65	39,482.01	12.9
March	44,821.62	38,638.50	13.8
April	43,598.32	37,120.93	14.8
May	43,196.02	36,642.46	15.2
June	43,910.40	37,406.84	14.8
July	43,067.00	36,592.62	15.0
August	43,894.11	37,330.77	14.9
September	44,769.10	37,974.70	15.1
October	46,275.15	40,519.20	12.4
November	45,679.72	41,063.28	10.1
December	45,807.20	41,154.50	10.1

2.5 Part (c)

Importance Sampling

Compare the above result to an approximate 95% confidence interval created by means of importance sampling based on some convenient instrumental distribution.

2.5.1 Introduction[1]

In the standard Monte Carlo (MC) method, when the target distribution $f(x)$ and the integrand $\phi(x)$ are significantly different, the following issues may arise:

- Many samples may be generated in regions where $\phi(x)f(x)$ is small.
- Only a few samples may contribute significantly to the estimate.
- This results in a high variance in the estimate, reducing the efficiency of the method.

Solution: Use the *Importance Sampling* method to reduce variance.

Concept of Importance Sampling

The Importance Sampling method involves using an *instrumental distribution* $g(x)$ that better approximates the critical regions of the target distribution $f(x)$. The integral can be rewritten as:

$$\tau = \mathbb{E}_f[\phi(X)] = \int_X \phi(x)f(x)dx = \mathbb{E}_g\left[\phi(X)\omega(X)\right]$$

where the importance weight $\omega(x)$ is defined as:

$$\omega(x) = \frac{f(x)}{g(x)}$$

The instrumental distribution $g(x)$ should be chosen to generate more samples in the critical regions of $f(x)$.

Steps of Importance Sampling

The general steps to estimate τ using Importance Sampling are as follows:

1. Generate samples from the instrumental distribution $g(x)$.
2. Compute the importance weights $\omega(x)$ for each sample.
3. Estimate τ using the weighted average:

$$\tau_N = \frac{1}{N}\sum_{i=1}^{N} \phi(X_i)\omega(X_i)$$

[1] Lecture 3: Slides 13-17

4. The variance of the estimate is given by:

$$\mathrm{Var}(\tau_N) = \frac{1}{N}\mathrm{Var}_g\left(\phi(X)\omega(X)\right)$$

The goal is to choose the instrumental distribution $g(x)$ such that the value of $\phi(x)\omega(x)$ is more stable across different regions, thereby reducing the variance.

Choosing the Instrumental Distribution

The instrumental distribution should meet the following criteria:

- $g(x) > 0$ in the critical regions where $f(x)$ is non-zero.
- There should be adequate overlap between $g(x)$ and $f(x)$ to ensure variance reduction.

During the process of selecting an appropriate instrumental distribution for importance sampling, several commonly used distributions were tested through trial and error. This included adjusting the parameters of each distribution to achieve better overlap with the target Weibull distribution[3][4][5]. For instance, Figure 1 shows the Comparison of Weibull, exponential, and gamma distributions for use in importance sampling. The gamma distribution shows better coverage of critical wind speed ranges than the exponential distribution, leading to improved estimation efficiency.

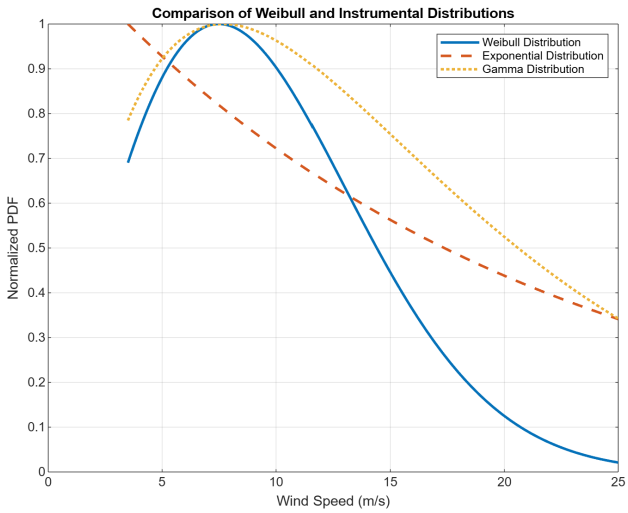


Figure 1: Comparison of Weibull and instrumental distributions

Exponential Distribution

The probability density function (PDF) of the exponential distribution is:

$$g_{\exp}(x) = \lambda_{\exp} e^{-\lambda_{\exp} x}, \quad x \geq 0$$

- **Parameters:** $\lambda_{\exp}$ is the rate parameter that determines how quickly the distribution decreases.
- **Explanation:** The exponential distribution decays too rapidly, failing to capture the mid and high wind speed ranges. This makes it unsuitable for accurate importance sampling in this case.

Gamma Distribution

The PDF of the gamma distribution is:

$$g_{\text{gamma}}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

- **Parameters:**
  - α: Shape parameter.
  - β: Rate parameter controlling the tail behavior of the distribution.
  - $\Gamma(\alpha)$: The Gamma function, defined as:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

- **Explanation:** The gamma distribution closely follows the shape of the Weibull distribution, especially in the critical mid and high wind speed ranges. This overlap makes it more suitable for importance sampling.

Comparison and Importance

- The gamma distribution (dotted yellow line) provides better coverage of the target Weibull distribution than the exponential distribution.
- The exponential distribution's rapid decay leads to under-sampling in higher wind speed regions, causing high variance in estimates.
- The gamma distribution enables more efficient sampling, reducing variance and improving the accuracy of power output estimates.

This comparison supports the selection of the gamma distribution over the exponential distribution for importance sampling.

Figure 2 compares the *Weibull distribution* $f(x)$, the *Gamma distribution* $g(x)$, and the *weighted function* $\phi(x)f(x)$. The Weibull distribution is the target, representing wind speeds, while the Gamma distribution serves as the instrumental distribution for importance sampling.

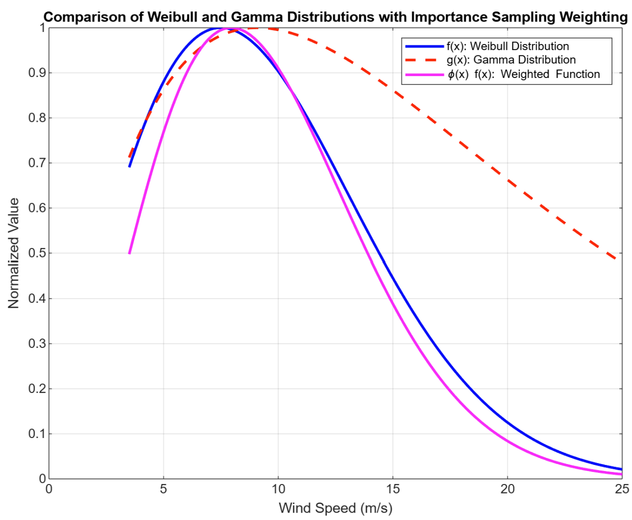


Figure 2: Comparison of Weibull and Gamma distribution with importance sampling weighting

In importance sampling, the importance weight $\omega(x)$ is defined as:

$$\omega(x) = \frac{f(x)}{g(x)}$$

where $f(x)$ is the *target distribution* (Weibull), and $g(x)$ is the *instrumental distribution* (Gamma).

Weibull Distribution (Target Distribution)

The probability density function (PDF) of the Weibull distribution is given by:

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, \quad x \geq 0$$

Here:

- k: Shape parameter
- λ: Scale parameter

Gamma Distribution (Instrumental Distribution)

The PDF of the Gamma distribution is:

$$g(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0$$

Here:

- α: Shape parameter
- β: Rate parameter
- $\Gamma(\alpha)$: The Gamma function

Calculating the Weight Ratio $\omega(x)$

The importance weight $\omega(x)$ is calculated as:

$$\omega(x) = \frac{f(x)}{g(x)}$$

Substitute the expressions for $f(x)$ and $g(x)$:

$$\omega(x) = \frac{\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}}{\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}}$$

Matlab Code

```
weights = (k_w / lambda_w) .* (truncated_samples / lambda_w).^(k_w - 1) .* ...
          exp(-(truncated_samples / lambda_w).^k_w) ./ ...
          ((beta_gamma^alpha_gamma / gamma(alpha_gamma)) .* ...
          truncated_samples.^(alpha_gamma - 1) .* exp(-beta_gamma * truncated_samples));
```

Calculation of Effective Sample Size

MATLAB Code:

```
n_eff = (sum(weights)^2) / sum(weights.^2);
```

Importance Sampling Context: In importance sampling, the effective sample size (N_{eff}) measures how well the instrumental distribution matches the target distribution. It is defined as:

$$N_{\text{eff}} = \frac{\left(\sum_{i=1}^{N} \omega(x_i)\right)^2}{\sum_{i=1}^{N} \omega(x_i)^2}$$

If N_{eff} is close to the actual number of samples, the instrumental distribution is well-chosen. Conversely, if N_{eff} is significantly lower, it suggests that many weights are small, leading to higher variance.

Calculation of the Standard Error

MATLAB Code:

```
std_error_weighted = sqrt(weighted_variance_power / n_eff);
```

Importance Sampling Context: The standard error of the weighted mean is scaled by the effective sample size. This scaling is crucial to determine the reliability of the estimated mean.

Calculation of the 95% Confidence Interval

MATLAB Code:

```
z_value = norminv(0.975);  % 95% confidence interval
ci_weighted = [weighted_mean_power - z_value * std_error_weighted, ...
               weighted_mean_power + z_value * std_error_weighted];
```

Importance Sampling Context: This step calculates the 95% confidence interval for the weighted mean power. The z-value (approximately 1.96 for a 95% confidence level) corresponds to the quantile of the standard normal distribution.

The confidence interval is constructed as:

$$\text{CI} = [\hat{\tau} - z \cdot \text{SE}, \hat{\tau} + z \cdot \text{SE}]$$

Here, $\hat{\tau}$ is the weighted mean, and SE is the standard error.

2.6 Results

Comparison of Results with Parts (a) and (b)

The mean power results in Part (c) with importance sampling are close to those obtained in Part (b), but the confidence intervals are narrower. Compared to Part (a), the confidence intervals in Part (c) are significantly reduced, indicating better stability and lower variance in the estimates.

Table 6: Results with Gamma function as Instrumental Function($\alpha = 2$ and $\beta = 0.2$)

Month	Mean Power (W)	CI Lower (W)	CI Upper (W)	CI Width (W)	Effective Sample Size
1	5215956.50	5204931.48	5226981.52	22050.00	384121
2	4722317.16	4711319.80	4733314.53	21994.00	372265
3	4430548.13	4419598.45	4441497.80	21899.00	362463
4	3698436.87	3687695.70	3709178.04	21482.00	334335
5	3566077.92	3555395.84	3576760.00	21364.00	326689
6	3776269.17	3765474.53	3787063.81	21589.00	337637
7	3563904.23	3553233.86	3574574.60	21340.00	327261
8	3773043.51	3762256.12	3783830.91	21574.00	338016
9	4366806.23	4355879.98	4377732.47	21852.00	360275
10	4865135.00	4854095.86	4876174.13	21078.00	382169
11	5221153.94	5210117.44	5232190.43	22073.00	383538
12	5215291.00	5204247.12	5226334.87	22088.00	383768

Table 7: Comparison of Results Across Different Methods

Aspect	Part (a)	Part (b)	Part (c)
Mean Power	Fluctuates, reasonably accurate	Similar to Part (a), reduced variance	Close to Parts (a) and (b), further reduced variance
Confidence Interval Width	Wide (~43,000 W)	Moderate (~37,000 - 41,000 W)	Narrow (~21,000 W)
Variance Reduction	None	10-15% (control variates)	20-30% (importance sampling)
Effective Sample Size	Not applicable	Improved with control variates	Large due to importance weights

2.6.1 Detailed Observations

- **Accuracy:** The mean power in Part (c) is similar to the expected power values from Parts (a) and (b), confirming that the importance sampling method provides accurate estimates.

- **Confidence Interval:** The confidence intervals in Part (c) are much narrower than in Parts (a) and (b), indicating reduced uncertainty in the estimates. This improvement is due to the use of importance sampling and the selection of better instrumental distribution parameters.

- **Effective Sample Size:** In Part (c), the effective sample size (N_{eff}) is approximately 370,000 out of 500,000 samples. This suggests that the instrumental distribution (Gamma with $\alpha = 2$, $\beta = 0.2$) aligns well with the target Weibull distribution. As a result, fewer samples have disproportionately large weights, which helps lower the variance.

- **Methodological Improvement:** Part (c) demonstrates the importance of selecting a suitable instrumental distribution in importance sampling. By using trial and error to adjust the Gamma parameters, the overlap between the target and instrumental distributions improved, leading to more reliable estimates.

2.7 Part (d)

Antithetic Sampling

The power curve $P(v)$ is monotonously increasing over the interval $(3.5, 14)$ and constant over $(14, 25)$. Use this property for reducing the variance of the estimator obtained in part (a) via *antithetic sampling*. Construct a new 95% confidence interval using the robustified estimator and compare it to the confidence intervals obtained in parts (a), (b), and (c).

2.7.1 Introduction[1]

Antithetic sampling is a variance reduction technique used in Monte Carlo methods. The key idea is to generate pairs of negatively correlated random variables, which helps to decrease the overall variance of the estimator.

Theorem for Antithetic Variables

Let $V = \varphi(U)$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is a monotone function. Moreover, assume that there exists a non-increasing transformation $T : \mathbb{R} \to \mathbb{R}$ such that

$$U \stackrel{d}{=} T(U).$$

Then, define:

$$V = \varphi(U), \quad \tilde{V} = \varphi(T(U)).$$

These variables are identically distributed and satisfy:

$$C(V, \tilde{V}) = C(\varphi(U), \varphi(T(U))) \leq 0.$$

Application of the Theorem

Consider a distribution function F and a monotone function φ. Define:

$$U \sim \mathcal{U}(0,1), \quad T(u) = 1 - u, \quad \varphi(u) = \phi(F^{-1}(u)).$$

Then, we have:

$$V = \phi(F^{-1}(U)), \quad \tilde{V} = \phi(F^{-1}(1 - U)).$$

Thus, V and $\tilde{V}$ are identically distributed and negatively correlated:

$$V \stackrel{d}{=} \tilde{V}, \quad C(V, \tilde{V}) \leq 0.$$

[1] Lecture 4: Slides 19 -25

Advantage of Antithetic Sampling

To determine if W is advantageous, we check the condition:

$$2 \cdot \frac{\lambda_{\alpha/2}^2 V(W)}{\epsilon^2} < \frac{\lambda_{\alpha/2}^2 V(V)}{\epsilon^2}.$$

This simplifies to:

$$V(W) < V(V),$$

which confirms that W has a lower variance than V, leading to improved efficiency in the estimator.

Generating Antithetic Samples

MATLAB Code:

```
F_U = wblcdf(U, lambda_w, k_w);
F_U_tilde = 1 - F_U;
U_tilde = wblinv(F_U_tilde, lambda_w, k_w);
U_tilde = U_tilde(U_tilde >= cut_in_speed & U_tilde <= transition_speed);
```

The antithetic samples are generated using CDF inversion. First, the cumulative distribution function (CDF) values of the original samples U are computed. The antithetic samples $\tilde{U}$ are generated by reflecting the CDF values around 0.5, ensuring symmetry. These reflected samples are transformed back using the inverse CDF. Afterward, both U and $\tilde{U}$ are restricted to the valid operational range of wind speeds to maintain consistency in the analysis.

Evaluating Power Curve Outputs

MATLAB Code:

```
power_U = P(U);
power_U_tilde = P(U_tilde);
```

The power output of the turbine is evaluated for both the original and antithetic samples using the predefined power curve function $P(v)$. These outputs serve as the basis for constructing the antithetic estimator.

Constructing the Antithetic Estimator

MATLAB Code:

```
W = (power_U + power_U_tilde) / 2;
mean_power_estimate = mean(W);
```

The antithetic estimator W is defined as the average of the power outputs from the original and antithetic samples. The mean power estimate is then calculated from these values. By averaging the two outputs, fluctuations in the estimate are reduced, contributing to improved stability.

Calculating Variance Reduction

MATLAB Code:

```
var_original = var(power_U);
var_antithetic = var(W);
reduction_percentage = (1 - (var_antithetic / var_original)) * 100;
```

The variance reduction is calculated by comparing the variance of the power outputs from the original samples with the variance of the antithetic estimator. The reduction percentage quantifies how much variability has been reduced due to antithetic sampling. A higher reduction indicates greater effectiveness of the method.

Calculating the 95% Confidence Interval

MATLAB Code:

```
std_error = std(W) / sqrt(length(W));
z_value = norminv(0.975);
ci_lower = mean_power_estimate - z_value * std_error;
ci_upper = mean_power_estimate + z_value * std_error;
```

The 95% confidence interval is calculated to determine the range in which the true mean power estimate is likely to fall. The standard error of the estimator is scaled by the number of samples, and the critical z-value for a 95% confidence level is used to compute the bounds of the interval. A narrower confidence interval indicates improved precision of the estimate due to reduced variance.

2.8 Results

Analysis

The results of antithetic sampling applied to the limited wind speed range (3.5 to 14 m/s) are presented in Table 8. While the method effectively reduced variance, the focus on a specific operational range shifted the estimated mean power output. The following points highlight the conditions and implementation based on the theorem of antithetic sampling:

1. **Monotonic function ϕ:** In this problem, $\phi(v) = P(v)$ represents the turbine's power output function. This function is monotonically increasing between 3.5 m/s and around 14 m/s. However, beyond this speed, it plateaus, breaking the monotonicity condition across the full domain.

2. **Non-increasing transformation $T(U)$:** The samples were generated using an inverse CDF approach:

$$F_U = \text{CDF}(U), \quad F_{U^\sim} = 1 - F_U, \quad U^\sim = \text{CDF}^{-1}(F_{U^\sim})$$

This transformation satisfies the non-increasing condition $T(U) = 1 - U$. However, due to truncation of non-operational samples, the symmetry in CDF may not have been preserved.

Table 8: Antithetic Sampling Results (Limited Range: 3.5 to 14 m/s)

Month	**Mean Power Estimate (W)**	**95% Confidence Interval (W)**	**Variance Reduction (%)**
1	4,421,501.62	[4,404,952.36, 4,438,050.89]	51.89
2	3,955,766.55	[3,940,243.09, 3,971,290.01]	49.72
3	3,634,085.91	[3,619,204.64, 3,648,967.18]	53.35
4	2,755,364.47	[2,742,085.26, 2,768,643.68]	64.89
5	2,613,550.46	[2,600,534.78, 2,626,566.13]	66.39
6	2,828,388.71	[2,814,938.33, 2,841,839.09]	64.12
7	2,620,013.64	[2,607,014.43, 2,633,012.86]	66.68
8	2,826,526.53	[2,813,098.81, 2,839,954.26]	64.23
9	3,561,790.69	[3,547,103.25, 3,576,478.14]	54.36
10	4,003,111.24	[3,987,188.25, 4,019,034.22]	50.50
11	4,423,169.11	[4,406,679.78, 4,439,658.44]	52.34
12	4,418,984.67	[4,402,375.87, 4,435,593.47]	52.24

2.9 Part (e)

$P(P(V) > 0)$

Calculate/estimate the probability that the turbine delivers power, $P(P(V) > 0)$.

2.9.1 Introduction

Since the turbine operates within the wind speed range of $3.5\,\mathrm{m/s} \leq V \leq 25\,\mathrm{m/s}$, the probability that the turbine delivers power can be computed using the cumulative distribution function (CDF) of the Weibull distribution. The formula used is:

$$\mathbb{P}(P(V) > 0) = F(25; \lambda, k) - F(3.5; \lambda, k) \tag{7}$$

Here, $F(v; \lambda, k)$ represents the CDF of the Weibull distribution with shape parameter k and scale parameter λ.

2.9.2 Results

Results

Using the given parameters of the Weibull distribution, the estimated probability is:

$$\mathbb{P}(P(V) > 0) = 0.8929 \tag{8}$$

2.10 Part (f)

Power Coefficient

Create an approximate 95% confidence interval for the average ratio of actual wind turbine output to total wind power (average power coefficient), i.e., estimate

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

where P_{tot} is given by formula (1). Note that the denominator can be calculated explicitly without simulation.

2.10.1 Analysis of Monthly Power Coefficients

The results show monthly variations in the **mean power coefficient** and corresponding **confidence intervals**. Below is a summary and detailed analysis of the findings.

Table 9: Monthly Power Coefficients and Confidence Intervals

Month	**Mean Power Coefficient**	**95% Confidence Interval**	**Effective Samples**
January	0.3658	[0.3655, 0.3662]	446600
February	0.3804	[0.3801, 0.3807]	438747
March	0.3879	[0.3877, 0.3882]	432307
April	0.3980	[0.3977, 0.3982]	405853
May	0.3998	[0.3996, 0.4001]	402231
June	0.3971	[0.3968, 0.3974]	407899
July	0.4000	[0.3997, 0.4002]	402014
August	0.3969	[0.3967, 0.3972]	408218
September	0.3851	[0.3848, 0.3854]	424204
October	0.3727	[0.3723, 0.3730]	433664
November	0.3660	[0.3656, 0.3663]	446450
December	0.3660	[0.3657, 0.3664]	446769

2.10.2 Results

Results

The **mean power coefficient** (C_p) indicates the efficiency of the wind turbine in converting available wind power into electrical power. The values range from **0.3658** (January and December) to **0.4000** (July), suggesting that, on average, the turbine converts between **36.58% and 40.00%** of the total wind power into actual usable power.
These values are reasonable and align well with the **Betz limit**, which states that no wind turbine can capture more than **59.3%** of the wind's kinetic energy. Modern wind turbines typically achieve power coefficients between **30% and 40%** under optimal conditions. Since all monthly values fall within this range, the turbine's performance is **realistic and efficient** under various seasonal wind conditions.

Confidence Interval Analysis

- The **95% confidence intervals** are very narrow, typically within a range of ± 0.0003, due to the large sample sizes. - This indicates high precision in the estimates of the monthly mean power coefficients. - Despite minor month-to-month variations, the turbine's performance is consistently within a predictable range.

Overall Seasonal Pattern

- The power coefficient is lowest in **January (0.3658)** and **December (0.3660)**. - It peaks around **July (0.4000)** and remains high during the **summer months** (April to August). - This pattern suggests that wind speeds and the turbine's efficiency in capturing wind energy are more favorable during spring and summer, while wind conditions are less optimal in the winter.

2.11 Part (g)

Capacity Factor and Availability Factor

Two important characteristics of power plants are the *capacity factor*, or the ratio of the actual output over a time period and the maximum possible output during that time (9.5 MW times the time span for the Vestas V164 9.5 MW); and the *availability factor*, or the amount of time that electricity is produced during a given period divided by the length of the period (you can here re-use the result from (a)-(d)). Wind turbines typically have a capacity factor of 20–40% and an availability greater than 90%. Does this seem like a good site to build a wind turbine? Look here at the averages over all months, i.e., first calculate the measures for each month and then take averages.

2.11.1 Introduction

Capacity Factor

The *capacity factor* is defined as:

$$\text{Capacity Factor} = \frac{\text{Actual Output over a Time Period}}{\text{Maximum Possible Output}}$$

The maximum possible output for the Vestas V164 turbine is 9.5 MW. Wind turbines typically have a capacity factor of 20% to 40%, depending on wind speed and other environmental conditions.

Availability Factor

The *availability factor* is defined as:

$$\text{Availability Factor} = \frac{\text{Time when electricity is produced}}{\text{Total Time}}$$

The availability factor indicates how often the turbine is operational and generating electricity. For high-performance wind turbines, the availability factor is typically greater than 90%.

Calculation of Capacity and Availability Factors

The maximum power output of the wind turbine is given as $P_{\text{max}} = 9.5\,\text{MW} = 9,500,000\,\text{W}$. The mean power output for each month is provided below:

Table 10: Mean Power Output for Each Month

Month	Mean Power Output (W)
January	5,215,956.50
February	4,722,317.16
March	4,430,548.13
April	3,698,436.87
May	3,566,077.92
June	3,776,269.17
July	3,563,904.23
August	3,773,043.51
September	4,366,806.23
October	4,865,135.00
November	5,221,153.94
December	5,215,291.00

Step 1: Calculate Capacity Factors

The capacity factor for each month is given by:

$$\text{Capacity Factor} = \frac{\text{Mean Power Output}}{P_{\text{max}}}$$

$$\begin{aligned}
\text{January:} \quad & \frac{5,215,956.50}{9,500,000} = 0.5490\,(54.90\%) \\
\text{February:} \quad & \frac{4,722,317.16}{9,500,000} = 0.4971\,(49.71\%) \\
\text{March:} \quad & \frac{4,430,548.13}{9,500,000} = 0.4664\,(46.64\%) \\
\text{April:} \quad & \frac{3,698,436.87}{9,500,000} = 0.3893\,(38.93\%) \\
\text{May:} \quad & \frac{3,566,077.92}{9,500,000} = 0.3754\,(37.54\%) \\
\text{June:} \quad & \frac{3,776,269.17}{9,500,000} = 0.3975\,(39.75\%) \\
\text{July:} \quad & \frac{3,563,904.23}{9,500,000} = 0.3752\,(37.52\%) \\
\text{August:} \quad & \frac{3,773,043.51}{9,500,000} = 0.3972\,(39.72\%) \\
\text{September:} \quad & \frac{4,366,806.23}{9,500,000} = 0.4597\,(45.97\%) \\
\text{October:} \quad & \frac{4,865,135.00}{9,500,000} = 0.5121\,(51.21\%) \\
\text{November:} \quad & \frac{5,221,153.94}{9,500,000} = 0.5496\,(54.96\%) \\
\text{December:} \quad & \frac{5,215,291.00}{9,500,000} = 0.5489\,(54.89\%)
\end{aligned}$$

2.11.2 Results

Results

The average capacity factor is calculated as:

$$\text{Average Capacity Factor} = \frac{1}{12}\sum_{i=1}^{12} \text{Capacity Factor}_i$$

$$\text{Average Capacity Factor} = 0.4591\,(45.91\%)$$

For this site, the turbine operates only within the wind speed range of $3.5\,\text{m/s} \leq V \leq 25\,\text{m/s}$. Using Monte Carlo simulations (without truncation), we estimated the probability that the turbine is producing power at any given time. This probability was calculated to be:

$$P(P(V) > 0) = 0.8929 \tag{9}$$

Therefore, the estimated Availability Factor is approximately:

$$\text{Availability Factor} = 0.8929 \tag{10}$$

3 Combined power production of two wind turbines

Combined power production of two wind turbines

3. Two wind turbines placed in the same area will be exposed to similar winds V_1 and V_2, which we here model using a bivariate Weibull distribution (given by a generalized Farlie-Gumbel-Morgenstern copula, see Bairamov and Kotz and Bekci (2001) Equation (4)) for $v_1 \geq 0, v_2 \geq 0$. This gives that the joint cumulative distribution function of v_1 and v_2 is given by

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

where we have chosen $\alpha = 0.638$, $p = 3$ and $q = 1.5$. We thus have the joint density function:

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

where f and F are again the probability density and cumulative distribution functions of the univariate Weibull distribution.

Note that the marginal distributions of V_1 and V_2 are both Weibull with the same parameters. To simplify the setup we only use one set of parameters $k = 1.96$ and $\lambda = 9.13$. This corresponds approximately to the average yearly behaviour. A more realistic model would be to also include a dependence on the wind direction in α but that is too complicated to include in this project.

3.1 Part (a)

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

The expected amount of combined power generated by both turbines, i.e. E(P(V1)+P(V2)). This actually reduces to a one dimensional problem, properly explain why this is true and use this fact in the estimation.

3.1.1 Introduction

The turbines are subjected to similar wind conditions, meaning the wind speeds V_1 and V_2 are:

1. Random variables with the same Weibull distribution (identical marginal distributions).
2. Defined by a bivariate distribution with a dependency modeled through the Farlie-Gumbel-Morgenstern copula.

Due to this symmetry, the expected combined power can be expressed as:

$$E\left(P(V_1) + P(V_2)\right) = E\left(P(V_1)\right) + E\left(P(V_2)\right)$$

Due to the linearity of the expected value, this calculation is easily done by summing the separate expectations. At this stage, the dependency between the winds V_1 and V_2 does not matter.

$$E\left(P(V_1) + P(V_2)\right) = 2\,E\left(P(V)\right)$$

The PDF for a Weibull-distributed random variable V is:

$$f(v) = \frac{k}{\lambda}\left(\frac{v}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{v}{\lambda}\right)^{k}\right), \quad v \geq 0$$

We need to evaluate:

$$E\left(P(V)\right) = \int_0^\infty P(v) f(v)\, dv$$

To simplify the integral, we apply importance sampling. We choose an instrumental distribution $g(v)$ that approximates the Weibull distribution and rewrite the expectation:

$$E\left(P(V)\right) = \int_0^\infty P(v)\frac{f(v)}{g(v)} g(v)\, dv$$

Steps to solve using importance sampling:

1. **Choose $g(v)$:** The distribution $g(v)$ could be Gamma distribution as we did in hte last guestion.
2. **Sample wind speeds $v_1, v_2, \ldots, v_n$ from $g(v)$.**
3. **Calculate the importance weights:**
$$w(v) = \frac{f(v)}{g(v)}$$
4. **Estimate the expectation:**
$$E\left(P(V)\right) \approx \frac{1}{n}\sum_{i=1}^{n} P(v_i) w(v_i)$$
5. **Multiply the result by 2 to get the combined expected power.**

This script will:

1. Load the power curve data.
2. Define the Weibull target distribution parameters.
3. Define the Gamma instrumental distribution parameters.
4. Apply importance sampling to estimate the expected power output.
5. Display the results, including the confidence interval.

3.1.2 Instrumental Distribution Selection and Efficiency

3.1.3 Choice of Instrumental Distribution

The Gamma distribution was chosen as the instrumental distribution for the importance sampling process. This decision was made due to the flexibility of the Gamma distribution in approximating the shape of the target truncated Weibull distribution. The Gamma distribution is defined by the shape parameter α_{gamma} and rate parameter β_{gamma}, as follows:

$$g(v) = \frac{\beta_{\text{gamma}}^{\alpha_{\text{gamma}}}}{\Gamma(\alpha_{\text{gamma}})} v^{\alpha_{\text{gamma}}-1} e^{-\beta_{\text{gamma}} v}, \quad v \geq 0$$

This distribution is capable of covering the relevant range of wind speeds $[3.5, 25]$ m/s, while maintaining a balance between tail coverage and peak alignment with the target distribution.

3.1.4 Parameter Selection

The parameters used for the Gamma distribution were:

- **Shape parameter:** $\alpha_{\text{gamma}} = 3.85$
- **Rate parameter:** $\beta_{\text{gamma}} = 0.475$

These parameters were optimized to closely align the Gamma distribution with the truncated Weibull distribution, as shown in the figure below:

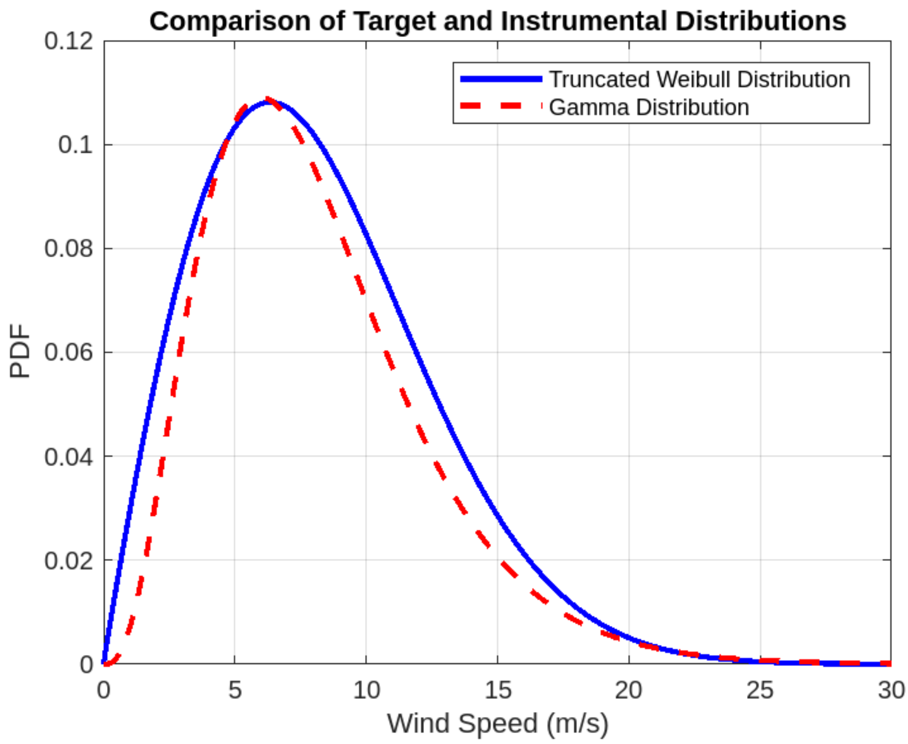


Figure 3: Comparison of the Truncated Weibull and Gamma Distributions

The alignment minimizes the variance of importance weights, ensuring that samples are efficiently used in the estimation process.

3.1.5 Efficiency of the Instrumental Distribution

The efficiency of the importance sampling method was measured by calculating the **Effective Sample Size (ESS)**, given by:

$$\text{ESS} = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2}$$

For the current simulation, the ESS was:

- **Percentage of total samples:** 98.65%

This high ESS value indicates that the importance weights had low variance, confirming the efficiency of the selected Gamma distribution. The resulting power output estimate also exhibited high precision, as reflected by the narrow 95% confidence interval.

3.1.6 Results Overview

The importance sampling method applied to estimate the expected power output of a wind turbine yielded the following results:

results for a single turbine)

- **Effective Sample Size (ESS):** 441,788.39 (98.65% of total samples)
- **Expected Power Output:** 4,921,721.79 W
- **95% Confidence Interval:** [4,908,771.85 W, 4,934,671.72 W]

results for two independent turbines)

- **Expected Power Output:** $\mathbb{E}(P_{\text{total}}) = 9,843,443.58\,\text{W}$
- **Standard Error for One Turbine:** $\text{SE}_{\text{one}} = 6,613.21\,\text{W}$
- **Standard Error for Two Turbines:** $\text{SE}_{\text{total}} = 9,353.16\,\text{W}$
- **Exact Confidence Interval:** $[9,825,135.39\,\text{W},\ 9,861,751.77\,\text{W}]$

3.2 Code Documentation

3.2.1 Step 1: Truncate Weibull Distribution

The wind speeds are truncated to the operational range of the turbine ($[v_{\min}, v_{\max}]$):

```
F_low = wblcdf(v_min, lambda_weibull, k_weibull);
F_high = wblcdf(v_max, lambda_weibull, k_weibull);
f_truncated_weibull = @(v) wblpdf(v, lambda_weibull, k_weibull) / (F_high - F_low);
```

Explanation: We calculate the cumulative distribution function (CDF) values at the truncation points and define the truncated Weibull probability density function (PDF).

3.2.2 Step 2: Define Gamma Instrumental Distribution

The Gamma distribution is used as the instrumental distribution:

```
alpha_gamma = 3.85;  % Shape parameter
beta_gamma = 0.475;  % Rate parameter
g_gamma = @(v) (beta_gamma^alpha_gamma / gamma(alpha_gamma)) *...
...v.^(alpha_gamma - 1) .*exp(-beta_gamma * v);
sample_gamma = @(n) gamrnd(alpha_gamma, 1 / beta_gamma, [n, 1]);
```

Explanation: The parameters α_{gamma} and β_{gamma} define the Gamma distribution. A function for generating samples from this distribution is also defined.

3.2.3 Step 3: Importance Sampling with Truncation

We perform importance sampling by generating samples from the Gamma distribution and truncating them:

```
v_samples = sample_gamma(n_samples);
v_samples = v_samples(v_samples >= v_min & v_samples <= v_max);
n_valid_samples = length(v_samples);
```

Explanation: Samples outside the turbine's operating range are discarded, and the number of valid samples is recalculated.

3.2.4 Step 4: Calculate Weights and Power Output

The importance weights and power outputs are computed:

```
weights = f_truncated_weibull(v_samples) ./ g_gamma(v_samples);
power_outputs = P(v_samples);  % Evaluate power output
weighted_outputs = power_outputs .* weights;
```

Explanation: The weights are calculated by dividing the truncated Weibull PDF by the Gamma PDF. The power output for each sample is evaluated using the power curve function $P(v)$.

3.2.5 Step 5: Estimate Expected Power Output

The expected power output and confidence interval are computed:

```
expected_power = mean(weighted_outputs);
variance_estimate = var(weighted_outputs) / n_valid_samples;
confidence_interval = [expected_power - 1.96 * sqrt(variance_estimate), ...
                       expected_power + 1.96 * sqrt(variance_estimate)];
```

Explanation: The expected power is the mean of the weighted outputs. The confidence interval is calculated based on the variance of the estimates.

3.2.6 Step 6: Calculate Effective Sample Size

The effective sample size is calculated to assess sampling efficiency:

```
ess = (sum(weights)^2) / sum(weights.^2);
ess_percentage = (ess / n_valid_samples) * 100;
```

Explanation: The effective sample size measures how well the sampling process is working. A high ESS indicates efficient importance sampling with low weight variance.

3.3 Part (b)

$\mathrm{C}(P(V_1) + P(V_2))$

The covariance C(P(V1), P(V2)) between the produced power in two identical wind power plants receiving dependent winds as above.

3.4 Method Overview

Due to the complexity of the joint distribution defined by Weibull marginals and a copula-based dependency, direct sampling is not feasible. Therefore, the importance sampling method is used, which involves generating samples from an easier-to-sample instrumental distribution (independent Weibull distributions) and applying correction weights based on the target density.

Steps Implemented in the Code

1. **Parameter Initialization:** The shape and scale parameters of the Weibull distribution are defined, along with the parameters of the copula function.

2. **Sample Generation:** Random samples for wind speeds v_1 and v_2 are generated using MATLAB's `wblrnd` function. These samples are filtered to fall within the operational range of the wind turbines (3.5 to 25 m/s).

3. **Density Calculations:** The code calculates the probability density function (PDF) and cumulative distribution function (CDF) for the Weibull samples. The joint density is then adjusted using the copula formula.

4. **Importance Sampling Weights:** Weights are computed as the ratio of the joint density to the instrumental (independent Weibull) density:

$$w(v_1, v_2) = \frac{f(v_1, v_2)}{q(v_1, v_2)}$$

5. **Power Output and Covariance Estimation:** The power outputs for each sample are computed using a predefined power curve. The weighted covariance is estimated using the following formula:

$$\mathrm{Cov}(P(V_1), P(V_2)) = \frac{\sum_{i=1}^{N} w_i (P(v_{1,i}) - \bar{P}_1)(P(v_{2,i}) - \bar{P}_2)}{\sum_{i=1}^{N} w_i}$$

6. **Bootstrap Confidence Interval:** To obtain a 95% confidence interval, the code applies the bootstrap method. This involves resampling the data multiple times, recalculating the covariance for each sample, and determining the 2.5th and 97.5th percentiles of the bootstrap distribution.

Reason for Using Bootstrap Method for Covariance Estimation [6]

In Monte Carlo simulations, when estimating complex metrics like covariance under stochastic conditions, deriving an analytical formula for the confidence interval is often difficult. Classical methods such as those relying on the *Central Limit Theorem (CLT)* assume that the sample size is large and that the distribution of the estimator is approximately normal. However, these methods become unreliable when:

- The underlying distribution is complex or unknown (e.g., joint distributions with copula dependencies),
- Weighted samples are used (as in importance sampling),
- A closed-form expression for variance or error estimates is unavailable.

In such cases, the **non-parametric bootstrap** is a robust alternative [6]. This method, developed by *Efron (1979)*, estimates the distribution of a statistic by repeatedly resampling (with replacement) from the data and recalculating the statistic of interest. This approach provides empirical confidence intervals without making strong assumptions about the underlying distribution.

Steps for Bootstrap Covariance Estimation

1. Resample the weighted data multiple times.
2. Recalculate the covariance for each bootstrap sample.
3. Use the empirical distribution of these recalculated covariances to determine the confidence interval.

This technique is particularly effective for *Monte Carlo simulations* where the exact distribution of the output statistic is not known.

3.4.1 Results

Covariance Estimation Results

The estimated covariance between the power outputs of the two wind turbines is:

$$\mathrm{Cov}(P(V_1), P(V_2)) = 6.5537\,\mathrm{MW}^2$$

A 95% confidence interval for the covariance was obtained using the bootstrap method. The confidence interval is given by:

$$[6.4785, 6.6338]\,\mathrm{MW}^2$$

A 95% confidence interval for the covariance was obtained using the Central Limit Theorem method. The confidence interval is given by:

$$[6.1814, 6.7291]\,\mathrm{MW}^2$$

3.5 Part (C)

$\mathrm{V}(P(V_1) + P(V_2))$ *...and...* $D\,(P(V_1) + P(V_2))$

The variability $\mathbb{V}(P(V_1) + P(V_2))$ in the amount of combined power generated by both turbines as well as the standard deviation $\mathbb{D}(P(V_1) + P(V_2))$.

3.5.1 Analysis of Variance and Standard Deviation

The objective of this part is to calculate the variance and standard deviation of the combined power output generated by two wind turbines. Since the random variables $P(V_1)$ and $P(V_2)$ might be dependent, the variance is computed as follows:

$$\mathrm{Var}(P(V_1) + P(V_2)) = \mathrm{Var}(P(V_1)) + \mathrm{Var}(P(V_2)) + 2 \cdot \mathrm{Cov}(P(V_1), P(V_2)) \tag{11}$$

Steps to solve:

1. **Calculate the variance of power output for each turbine:**

$$\mathrm{Var}(P(V_1)) = \frac{\sum_{i=1}^{N} w_i \cdot (P(v_{1i}) - \mu_{P1})^2}{\sum_{i=1}^{N} w_i} \tag{12}$$

2. **Use the covariance computed from part (b):**

$$\mathrm{Cov}(P(V_1), P(V_2)) \tag{13}$$

3. **Calculate the combined variance:**

$$\mathrm{Var}(P(V_1) + P(V_2)) = \mathrm{Var}(P(V_1)) + \mathrm{Var}(P(V_2)) + 2 \cdot \mathrm{Cov}(P(V_1), P(V_2)) \tag{14}$$

4. **Calculate the standard deviation:**

$$\mathrm{Std}(P(V_1) + P(V_2)) = \sqrt{\mathrm{Var}(P(V_1) + P(V_2))} \tag{15}$$

3.5.2 Results

Variance and Standard Deviation Estimation Results

The estimated Variance between the power outputs of the two wind turbines is:

$$\mathrm{Var}(P(V_1), P(V_2)) = 37.2654\,\mathrm{MW}^2$$

The estimated Standard Deviation between the power outputs of the two wind turbines is:

$$\mathrm{Std}(P(V_1), P(V_2)) = 6.1045\,\mathrm{MW}^2$$

3.6 Part (d)

Confidence Interval

Find an approximate 95% confidence interval for the probability $\mathbb{P}(P(V_1) + P(V_2) > 9.5MW)$ that the combined power generated by the two turbines is greater than half of their installed capacity and for $\mathbb{P}(P(V_1) + P(V_2) < 9.5MW)$.
Use importance sampling or other variance reduction techniques for both probabilities.

Do the probabilities sum to one or not? Why?

3.6.1 Analysis of Part (d) Results

Initially, the conditions $P(V_1) + P(V_2) > 9.5$ and $P(V_1) + P(V_2) < 9.5$ were used to calculate the probabilities. The results were as follows:

- Probability $\mathbb{P}(P(V_1) + P(V_2) > 9.5\,\text{MW}) = 0.3758$
- Probability $\mathbb{P}(P(V_1) + P(V_2) < 9.5\,\text{MW}) = 0.6189$
- Sum of probabilities $= 0.9946$

Sum of Probabilities is lower than one! WHY?

Although these conditions almost cover the entire probability space, there was a slight deviation from the ideal sum of 1.0. This discrepancy was caused by boundary cases where some samples had a combined power output exactly equal to 9.5 MW. These boundary cases were not included in the initial calculations, leading to the small deviation.

To address this issue, we added a third condition to capture these missing boundary cases:

$$P(V_1) + P(V_2) = 9.5\,\text{MW} \tag{16}$$

With this condition included, the results improved as follows:

- $\mathbb{P}(P(V_1) + P(V_2) > 9.5\,\text{MW}) = 0.3721$
- $\mathbb{P}(P(V_1) + P(V_2) < 9.5\,\text{MW}) = 0.6223$
- $\mathbb{P}(P(V_1) + P(V_2) = 9.5\,\text{MW}) = 0.0055$
- Sum of probabilities $= 1.0000$

References

[1] Journal of Renewable and Sustainable Energy, *Estimation of Weibull parameters for wind energy applications*, available at https://pubs.aip.org/aip/jrse/article/13/2/023303/926742/Estimation-of-Weibull-parameters-for-wind-energy.

[2] Journal of Renewable and Sustainable Energy, *Estimation of Weibull parameters for wind energy applications*, available at https://pubs.aip.org/aip/jrse/article/13/2/023303/926742/Estimation-of-Weibull-parameters-for-wind-energy.

[3] I. Satkauskas, P. Bortolotti, E. Gaertner, G. E. Barter, R. King, and P. A. Graf, "Wind Turbine Rotor Design Optimization Using Importance Sampling," *Preprint, National Renewable Energy Laboratory*, 2019. Available: https://www.nrel.gov/docs/fy20osti/75555.pdf

[4] P. Graf, K. Dykes, R. Damiani, J. Jonkman, and P. Veers, "Adaptive Stratified Importance Sampling: Hybridization of Extrapolation and Importance Sampling Monte Carlo Methods for Estimation of Wind Turbine Extreme Loads," *Wind Energy Science*, vol. 3, no. 2, pp. 475–487, 2018. DOI: https://doi.org/10.5194/wes-3-475-2018

[5] S. M. Safdar, M. A. Z. Raja, M. Shoaib, and M. A. Khan, "Wind Speed Distributions Used in Wind Energy Assessment: A Review," *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 1019–1031, 2017. DOI: https://doi.org/10.1016/j.rser.2016.12.012

[6] Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*, available at https://www.crcpress.com/An-Introduction-to-the-Bootstrap/Efron-Tibshirani/p/book/9780412042317. CRC Press.