Monte Carlo Hand-in Assignment 1

Eliot Montesino Petrén, el6183mo-s, 990618-9130, LU Faculty of Engineering F19 , eliot.mp99@gmail.com,
Kasper Nordenram, ka0884no-s, 990212-7910, LU Faculty of Engineering F19

I. Introduction

This report explores Monte Carlo and empirical methods in mathematical statistics, applied to the specific context of wind turbine power production. Monte-Carlo methods are famous for practical applicability and this assignment tries to demonstrate the ability of these methods to solve real-world problems. The report is a hand-in in the course of *"Monte Carlo and Empirical Methods for Stochastic Inference, FMSN50"* at Lund University Faculty of Engineering.

Specifically, the assignments analyzes the efficiency and output variability of wind turbines using various advanced stochastic methods. Wind speed is assumed Weibull distributed for modeling wind speeds, and this is of course a critical factor in assessing the performance of wind turbines. Furthermore, the investigation extends to optimization of power output, guided by the Betz limit and empirical power coefficients. The methods used are based on Monte Carlo simulations, truncated distributions, control variate techniques, and importance sampling, contributing to the precision of the estimations. This approach also evaluates the variability and reliability of wind turbine performance in different environmental conditions.

Let's go!

Q1 Random number generation

a: Using the definition of the conditional distribution,

$$f_{X|X\in I}(x) = \frac{f_X(x)}{\mathbb{P}(X \in I)}. \tag{1}$$

For the cumulative distribution function,

$$\begin{aligned} &F_{X|X\in I}(x) = \{x \text{ is in } I\} = \int_a^x f_{X|X\in I}(x)dx = \\ &\int_a^x \frac{f_X(x)}{\mathbb{P}(X\in I)}dx = \frac{1}{F_X(b)-F_X(a)}\int_a^x f_X(x)dx = \\ &\frac{F_X(x)-F_X(a)}{F_X(b)-F_X(a)} \end{aligned} \tag{2}$$

b: Attempting to invert the function, we assert

$$F_X(F_X^{-1}(u)) = u. \tag{3}$$

Using the expression from above,

$$\begin{aligned} &u = \frac{F_X(F^{-1}_{X|X\in I}(u)) - F_X(a)}{F_X(b)-F_X(a)} \iff \\ &F^{-1}_{X|X\in I}(u) = F_X^{-1}((1-u)F_X(a) + uF_X(b)) \end{aligned} \tag{4}$$

By generating random variables $u \in U(0,1)$, we get samples from $X|X \in I$ as $F_X^{-1}(u)$. Calling these samples $Z = F_X^{-1}(U)$, we see that they will achieve the correct distribution by

$$\begin{aligned} &F_Z(z) = \mathbb{P}(Z < z) = \mathbb{P}(F_X^{-1}(U) < z) = \\ &\mathbb{P}(U < F_X(z)) = F_X(z). \end{aligned} \tag{5}$$

This method is known as inverse transform sampling. [1]

Q2 Power production of a wind turbine

We model the windspeed as being Weibull distributed, so $V \in \text{Weibull}(\lambda, k)$, and the function $P(v)$, mapping the wind-speed to the power output. We define the interval $I = [a, b]$, where $a = 4$ and $b = 25$. We will use the approximation

$$\tau := \mathbb{E}X \simeq \frac{1}{N}\sum_{i=1}^{N} X_i := \tau_N, \tag{6}$$

where X and $X_i,\ i \in \{1, 2, ..., N\}$ are all i.i.d. random variables [2]. We also use the asymptotic normality of τ_N, with the variance estimation

$$\mathbb{V}\tau_N \simeq \frac{s_N^2}{N} := \frac{1}{N^2}\sum_{i=1}^{N}(X_i - \bar{X})^2, \tag{7}$$

to find the 99% approximate confidence interval. Since the number of i.i.d samples are large, a normal distribution can be assumed in determining the 99% confidence interval according to...

$$I_{99\%} := [\tau_N - \lambda_{0.005}\frac{s_N}{\sqrt{N}}, \tau_N + \lambda_{0.005}\frac{s_N}{\sqrt{N}}]. \tag{8}$$

a: We define and calculate

$$\begin{aligned} &\tau := \mathbb{E}_{f_V}P(V) = \int_{-\infty}^{\infty} P(v)f_V(v)dv = \\ &\int_I P(v)f_V(v)dv = \int_{-\infty}^{\infty} P(v)\mathbb{P}(V\in I)f_{V|V\in I}(v)dv = \\ &\mathbb{E}_{f_{V|V\in I}}P(V)\mathbb{P}(V\in I) = \mathbb{P}(V\in I)\mathbb{E}_{f_{V|V\in I}}P(V), \end{aligned} \tag{9}$$

where he last equality holds because P is 0 outside of I. We will sample first from the entire support of V, then only from I, which we can do using the expression from Subsection I. To do this, we have to use the inverted Weibull CDF, which is a provided function in Matlab "wblinv", but the inversion can also be done without the built in function and that derivation is also provided as follows...

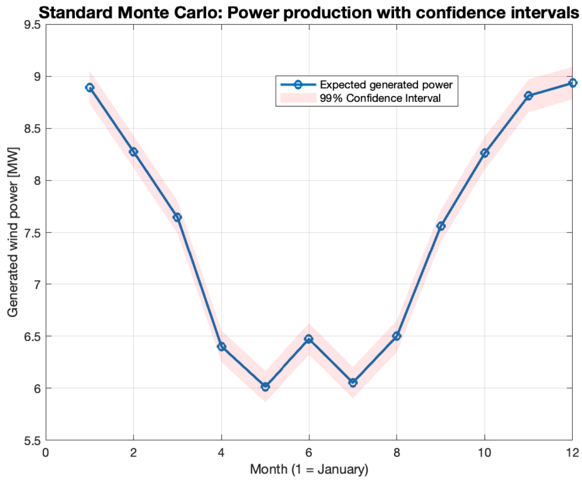

Fig. 1. Estimated power production of a wind turbine for each month with 99% confidence bands. No variance reduction methods have been applied, yet.

$$u = F(F^{-1}(u)) = \{F^{-1}(u) > 0\} = 1 - \exp\left(-\left(\frac{F^{-1}(u)}{\lambda}\right)^k\right) \iff F^{-1}(u) = \lambda(-\ln(1-u))^{1/k}. \quad (10)$$

The estimated time series of generated wind power output (not using truncated sampling yet) with confidence intervals was calculated and can be seen in Figure 1 and the estimated confidence intervals can be seen in Table I.

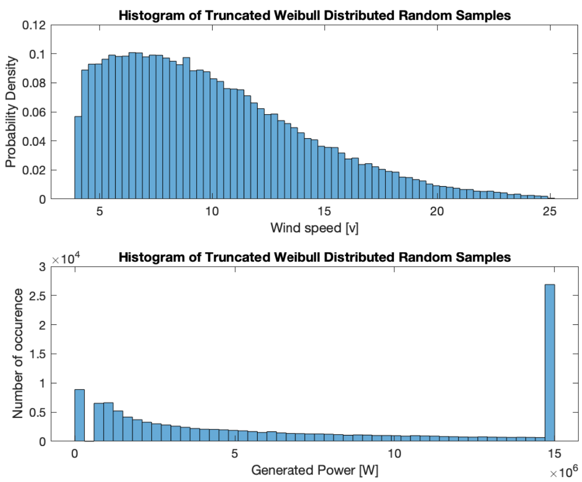

Fig. 2. Histograms detailing the samples from the truncated Weibull distribution, as well as the distribution of their power output. Note that samples from the truncated distribution are within the conditional range $[4, 25]$.

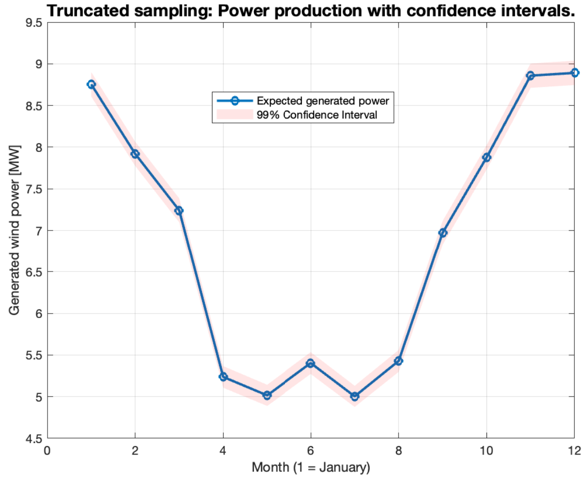

Fig. 3. Estimated power production of a wind turbine for each month with 99% confidence bands, produced from the truncated distribution with 10000 samples.

Note that from here on the number of drawn samples is $10,000$ and this will be the case until the last task. The number of samples drawn must be the same to be able to compare the different methods, since a larger sample will result in a lower variance and a more trustworthy expected value estimation. This specific number of samples were chosen partly because of computational time reasons and partly because to leave some room for noticeable improvement in this task.

Now using truncated sampling, the results of the truncated samples were deemed reasonable according to Figure 2 where no samples can be seen outside of the interval. The mass is shifted to the right in the subplot below due to the nature of the power function. The estimated time series with confidence intervals using truncated sampling can be seen in fig 3 and the confidence intervals can be seen in Table II

b: We define the control variate

$$Z = P(V) + \beta(V - m). \quad (11)$$

The m that gives an unbiased estimate is

$$m = \mathbb{E}V = \Gamma(1 + \frac{1}{k})\lambda, \quad (12)$$

and the optimal β that gives minimum variance is

$$\beta = -\frac{\mathbb{C}(P(V), V)}{\mathbb{V}(V)}. \quad (13)$$

This has no closed-form solution, so we will use an online update scheme where

$$
\begin{aligned}
C_{l+1} &= \frac{l}{l+1} C_l + \frac{1}{l+1} \phi(X_{l+1})(Y_{l+1} - m) && \text{(14a)}\\
V_{l+1} &= \frac{l}{l+1} V_l + \frac{1}{l+1} (Y_{l+1} - m)^2 && \text{(14b)}\\
\beta_l &= -\frac{C_{l+1}}{V_{l+1}} && \text{(14c)}\\
Z_{l+1} &= \phi(X_{l+1}) + \beta_l (Y_{l+1} - m) && \text{(14d)}\\
\tau_{l+1} &= \frac{l}{l+1} \tau_l + \frac{1}{l+1} Z_{l+1} && \text{(14e)}
\end{aligned}
$$

where C_0 and V_0 are 0, and β_0 = 1. In this implementation, $Y_l = V_l$ and $\phi(X_l) = P(V_l)$. Neglecting the variance in this estimate (not exactly realistic, so keep that in mind), we get

$$
\begin{aligned}
\mathbb{E}Z &= \mathbb{E}\{P(V) + \beta(V - \mathbb{E}V)\} = \mathbb{E}P(V) && \text{(15a)}\\
\mathbb{V}Z &= \mathbb{V}\{P(V)\}\left(1 - \frac{\mathbb{C}(P(V),V)^2}{\mathbb{V}\{P(V)\}\mathbb{V}\{V\}}\right), && \text{(15b)}
\end{aligned}
$$

so the same expectation but a smaller or equal variance. Let us now proceed to estimation and see if the variance is reduced or not in this specific case.

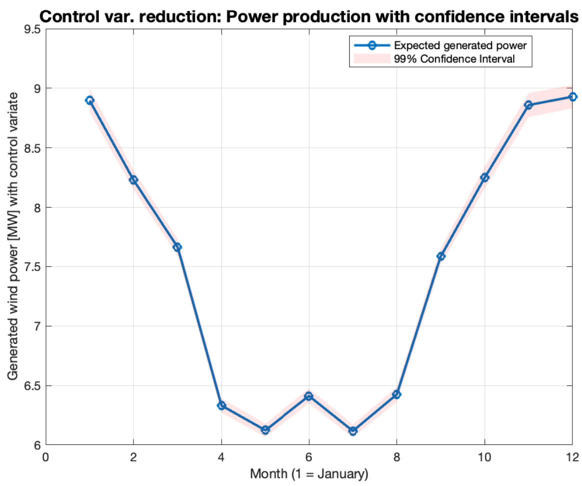

Fig. 4. Estimated power production for each month with 99% confidence bands, produced using a control variate with 10000 samples.

The estimated time series of generated wind power output with confidence intervals using control variables variance reduction was calculated and can be seen in Figure 4 and the estimated confidence intervals can be seen in Table IV.

c: As an instrumental distribution, we use a normal distribution $f_N(x)$ that gives a small variance [4]

$$\mathbb{V}P(V)\omega(V), \tag{16}$$

of the estimate. Here,

$$\omega(V) = \frac{f_V(V)}{f_N(V)}. \tag{17}$$

We thus want to minimize

$$\mathbb{V}\frac{P(V)f_V(V)}{f_N(V)} = \mathbb{V}W. \tag{18}$$

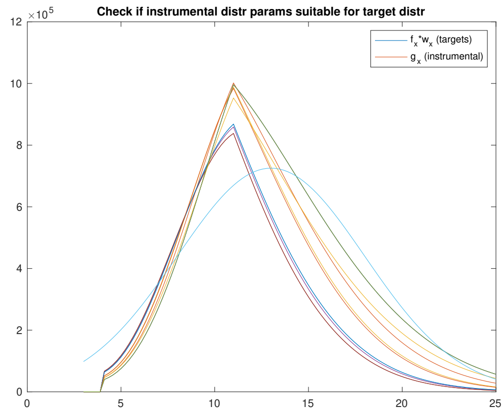

Fig. 5. Comparison instrumental distribution and the target distributions for each month.

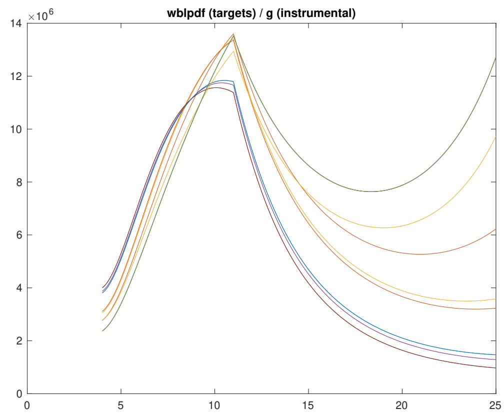

Fig. 6. $\frac{P(v)f(v)}{g(v)}$, this should be as constant as possible on the interval $[4, 25]$ for our estimate to have small variance.

Clearly, if W is relatively constant for all V, the variance will be small. Thus, the proposal distribution (in this case a normal distribution) that minimizes the variance fits the target distribution (the distribution $P(V)f_V(V)$) well. The parameters of the normal distribution were set to $\mu = 13$, $\sigma = 5$, by visual inspection. The fit can be seen in Figure 5 and in Figure 6

TABLE I
STANDARD MONTE CARLO SAMPLING: 99% CONFIDENCE INTERVAL FOR GENERATED POWER (MW)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Lower Bound	8.7373	8.1196	7.4889	6.2498	5.8665	6.3217	5.9044	6.3482	7.4048	8.1034	8.6561	8.7792
Upper Bound	9.0495	8.4304	7.7997	6.5534	6.1639	6.6246	6.2032	6.6512	7.7146	8.4191	8.9704	9.0904
Difference	0.31222	0.31087	0.31080	0.30359	0.29739	0.30291	0.29882	0.30299	0.30978	0.31566	0.31434	0.31117

TABLE II
TRUNCATED SAMPLING: 99% CONFIDENCE INTERVAL FOR GENERATED POWER (MW)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Lower Bound	8.6076	7.7745	7.0986	5.1081	4.8891	5.2735	4.8776	5.2985	6.8309	7.7311	8.7117	8.7451
Upper Bound	8.9000	8.0641	7.3851	5.3668	5.1428	5.5346	5.1305	5.5604	7.1153	8.0248	9.0039	9.0364
Difference	0.29233	0.28965	0.28653	0.25869	0.25363	0.26111	0.25292	0.26189	0.28446	0.29372	0.29216	0.29127

TABLE III
CONTROL VARIATE SAMPLING: 99% CONFIDENCE INTERVAL FOR GENERATED POWER (MW)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Lower Bound	8.7982	8.1472	7.5972	6.2717	6.0665	6.3503	6.0628	6.3627	7.5192	8.1590	8.7595	8.8308
Upper Bound	8.9945	8.3107	7.7379	6.3913	6.1795	6.4744	6.1720	6.4881	7.6582	8.3418	8.9588	9.0279
Difference	0.19629	0.16358	0.14069	0.11953	0.11301	0.12408	0.10928	0.12544	0.13903	0.18283	0.19931	0.19710

TABLE IV
IMPORTANCE SAMPLING: 99% CONFIDENCE INTERVAL FOR GENERATED POWER (MW)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Lower Bound	8.7995	8.0613	7.5315	6.1673	5.9497	6.2742	5.9497	6.2742	7.4373	8.0849	8.7995	8.7995
Upper Bound	8.9488	8.2265	7.7116	6.3561	6.1413	6.4613	6.1413	6.4613	7.6199	8.2313	8.9488	8.9488
Difference	0.14938	0.16523	0.18009	0.18877	0.19158	0.18718	0.19158	0.18718	0.18258	0.14633	0.14938	0.14938

d: As $P(V)$ is monotone from 4 to 25 , we can use the antithetic variables

$$\hat{V} = \quad P(F^{-1}_{X|X\in I}(U)) \tag{19a}$$

$$\tilde{V} = \quad P(F^{-1}_{X|X\in I}(1-U)), \tag{19b}$$

to form

$$W = \frac{1}{2}(\hat{V} + \tilde{V}), \tag{20}$$

where $U \in U(0,1)$. The variance of W is [3]

$$\mathbb{V}W = \frac{1}{4}(\mathbb{V}\hat{V} + 2\mathbb{C}(\hat{V} + \tilde{V}) + \mathbb{V}\tilde{V}) \tag{21a}$$

The chosen antithetic variables will (lecture 4 slide 21) have a nonpositive covariance, so this achieves a strict variance reduction.

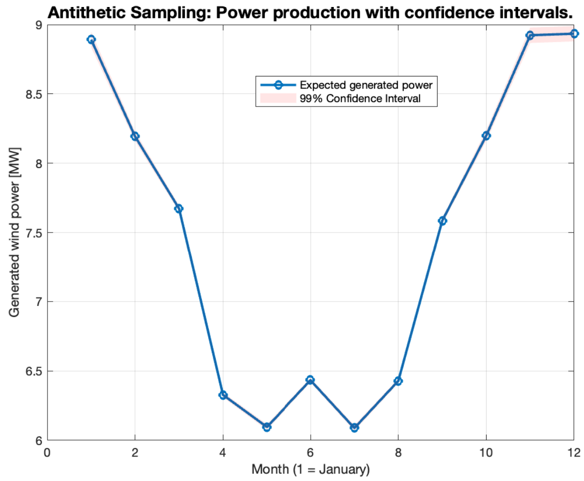

Fig. 7. Estimated power production for each month with 99% confidence bands, produced using antithetic sampling with 10000 samples.

The estimated time series of generated wind power output with confidence intervals using antithetic sampling for vari-

TABLE V
ANTITHETIC SAMPLING: 99% CONFIDENCE INTERVAL FOR GENERATED POWER (MW)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Lower Bound	8.8359	8.1579	7.6504	6.3100	6.0759	6.4174	6.0683	6.4125	7.5652	8.1595	8.8675	8.8788
Upper Bound	8.9471	8.2278	7.6949	6.3449	6.1165	6.4494	6.1093	6.4446	7.6061	8.2379	8.9788	8.9898
Difference	0.11122	0.06993	0.04445	0.03485	0.04059	0.03196	0.04102	0.03208	0.04092	0.07834	0.11132	0.11107

TABLE VI
PROBABILITY OF TURBINE POWER DELIVERY PER MONTH

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0.87929	0.86532	0.85265	0.79897	0.79122	0.80266	0.79122	0.80266	0.85021	0.85379	0.87929	0.87929

ance reduction was calculated and can be seen in Figure 7 and the estimated confidence intervals can be seen in Table V. It is apparent both from the figure and from the table that the antithetic sampling variance reduction method is the most effective in reducing variance compared to the other methods, at least specifically for this report and task with wind turbines.

It should also be noted that, at least with our implementation, that this type of sampling was the fastest to compute. This is likely due to the number of random samples that has to be drawn is half of the samples size due to the nature of antithetic sampling. Would the inverse CDF function not be easy to compute, then the computational time would increase dramatically. This will become relevant later in this report because antithetic sampling was considered for the last task as well but that would require optimization to invert the CDF function in that task, thus rendering the purpose of antithetic sampling useless.

e: The probability of producing wind power is found by integrating the density of V over I, as $P(v)$ is positive here. The estimated probabilities after calculation can be seen in Table VI.

$$\begin{aligned}&\mathbb{P}(P(V)>0)=\mathbb{P}(4<V<25)=\int_4^{25} f_V(v)dv=\\&\int_4^{25}\frac{k}{\lambda}\left(\frac{v}{\lambda}\right)^{k-1}\exp\left(-\left(\frac{v}{\lambda}\right)^k\right)=\left\{u=\left(\frac{v}{\lambda}\right)^k\right\}=\\&\int_{\left(\frac{4}{\lambda}\right)^k}^{\left(\frac{25}{\lambda}\right)^k}\exp(u)du=\exp\left(\frac{25}{\lambda}\right)^k-\exp\left(\frac{4}{\lambda}\right)^k. \qquad (22)\end{aligned}$$

f: The expectation of power in the wind is, using the given expression for power in the wind and expectation of a Weibull distribution,

$$\begin{aligned}&\mathbb{E}P_{tot}(V)=\mathbb{E}\frac{1}{2}\rho\pi\frac{d^2}{4}V^3=\\&\frac{1}{2}\rho\pi\frac{d^2}{4}\mathbb{E}V^3=\frac{1}{2}\rho\pi\frac{d^2}{4}\Gamma(1+\frac{3}{k})\lambda^3, \qquad (23)\end{aligned}$$

where ρ is the air density. Using this in combination with the method that was the most successful in reducing variance, **antithetic sampling**, the following average power coefficients were calculated with a 99% confidence interval in Table VII. Perhaps an even better comparison than this is to compare it to the maximum capacity that is possible in practice, the Betz limit. Therefore a new table was also produced that compares the power coefficient against the Betz limit instead, see Table VIII. The total average when compared to the Betz limit (that is the expected value of the whole year) was equal to $0.4865 = 49\%$, and not comparing against the Betz limit this corresponds to a yearly average power coefficient of 29% and the report states that "real wind turbines can usually utilise at most about 40 to 45% of the total power". We are below what one would desire a wind turbine to produce.

g: We estimate the capacity factor (c.f.) and the availability factor (a.f.),

$$\text{c.f.}\simeq\frac{1}{12}\sum_{i=1}^{12}\frac{\tau_{N,i}}{15MV\cdot 1\text{ month}} \qquad (24a)$$

$$\text{a.f.}\simeq\frac{1}{12}\sum_{i=1}^{12}\left[\exp\left(\frac{25}{\lambda_i}\right)^{k_i}-\exp\left(\frac{4}{\lambda_i}\right)^{k_i}\right], \qquad (24b)$$

so we average over the months. Note that the second expression is an approximation only because the Weibull model is an approximation. We have used antithetic sampling to find the τ:s. The variances of these estimates are

TABLE VII
99% CONFIDENCE INTERVAL FOR AVERAGE POWER COEFFICIENT PER MONTH.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Lower Bound	0.2157	0.2603	0.2903	0.3419	0.3527	0.3362	0.3523	0.3359	0.2958	0.2327	0.2164	0.2167
Upper Bound	0.2184	0.2626	0.2920	0.3438	0.3551	0.3378	0.3547	0.3376	0.2974	0.2349	0.2192	0.2194

TABLE VIII
COMPARISON AGAINST BETZ LIMIT: AVG POWER COEFFICIENT/BETZ FOR EACH MONTH IN A 99& CONFIDENCE INTERVAL.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Lower Bound	0.3640	0.4393	0.4899	0.5770	0.5953	0.5673	0.5945	0.5668	0.4991	0.3926	0.3653	0.3657
Upper Bound	0.3685	0.4431	0.4927	0.5802	0.5992	0.5701	0.5985	0.5697	0.5018	0.3964	0.3698	0.3703

TABLE IX
99% CONFIDENCE INTERVAL OF CAPACITY FACTOR FOR EACH MONTH.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Lower Bound	0.5891	0.5439	0.5100	0.4207	0.4051	0.4278	0.4046	0.4275	0.5043	0.5440	0.5912	0.5919
Upper Bound	0.5965	0.5485	0.5130	0.4230	0.4078	0.4300	0.4073	0.4296	0.5071	0.5492	0.5986	0.5993

TABLE X
POINT ESTIMATES OF CAPACITY FACTOR AND AVAILABILITY FACTOR.

c.f.	0.50
a.f.	0.84

$$\mathbb{D}\frac{1}{12}\sum_{i=1}^{12}\frac{\tau_{N,i}}{15MV\cdot 1\text{ month}} = \frac{1}{12\cdot 15MV\cdot 1\text{ month}}\sqrt{\mathbb{V}\sum_{i=1}^{12}\tau_{N,i}} = \{\tau_{N,i}\text{ independent}\} = \frac{1}{12\cdot 15MV\cdot 1\text{ month}}\sqrt{\sum_{i=1}^{12}\frac{s_{N,i}^2}{N}} = \frac{1}{12\cdot 15MV\cdot 1\text{ month}\cdot\sqrt{N}}\sqrt{\sum_{i=1}^{12}s_{N,i}^2} \tag{25a}$$

$$\mathbb{D}\frac{1}{12}\sum_{i=1}^{12}\left[\exp\left(\frac{25}{\lambda_i}\right)^{k_i} - \exp\left(\frac{4}{\lambda_i}\right)^{k_i}\right] = 0, \tag{25b}$$

where the second standard deviation is of course not zero in reality, but the parameters are assumed to be known so this is the result. If the original modeling data was available, an interval could be produced. The point estimates are given in Table X and a 99% confidence interval of the monthly capacity factor is seen in IX.

The text from the task states "Wind turbines typically have a capacity factor of 20 to 40% and an availability greater than 90%." Clearly, the capacity factor is consistently sufficiently or even very high, while the availability factor is somewhat less convincing. It seems like a fairly reasonable place to build the wind power plant, assuming the slightly high intermittency of the power production is not a problem.

Q3 COMBINED POWER PRODUCTION OF TWO WIND TURBINES

In this section, we have used $100,000$ samples instead of $10,000$ to reduce variance even further, a higher number of samples took too long time to reasonably compute.

a: We can reduce the problem to be 1-dimensional by noting that expectation is linear, so

$$\mathbb{E}(P(V_1)+P(V_2)) = \mathbb{E}P(V_1)+\mathbb{E}P(V_2) = 2\mathbb{E}P(V_1). \tag{26}$$

Since the marginal densities of the winds are the same as the density used in Q2, we can use antithetic sampling for one-dimensional V and multiply by 2.

The point estimate of the expectation is 7.567 MW.

b: The covariance can be rewritten

$$\mathbb{C}(P(V_1),P(V_2)) = \mathbb{E}\{P(V_1)P(V_2)\} - \mathbb{E}P(V_1)\mathbb{E}P(V_2). \tag{27}$$

We now want samples from both V_1 and V_2. This can be done one variable at a time. The following argument shows this,

$$f_{V_2|V_1}(v_2|v_1) = \frac{f_{V_1,V_2}(v_1,v_2)}{f_{V_1}(v_1)} \iff f_{V_1,V_2}(v_1,v_2) = f_{V_1}(v_1)f_{V_2|V_1}(v_2|v_1). \tag{28}$$

Thus, we can reuse the samples from exercise Q2a (without truncation), assigning them to be the V_1-samples, and then pull V_2 samples from the conditional distribution given each of the V_1-samples. After discussion with Magnus, we decided to go forward with this, as this will be an interesting method, even though it might risk to be a worse method than sampling from, for example, a two-dimensional normal distribution with suitably chosen means and covariance matrix. We have used rejection sampling to get the V_2-samples. To do this, we needed to choose a $g(\cdot)$ and K such that

$$Kg(v_2) > f_{V_2|V_1}(v_2|v_1),\ \forall v_1, v_2. \tag{29}$$

This has been found by plotting many different $f_{V_2|V_1}(v_2|v_1)$, and finding a $Kg(\cdot)$ that dominates them, according to rejection sampling criterions. We see in Figure 8 that we can choose $g(\cdot)$ as a Weibull distribution with $\lambda = 12$ and $k = 1.7$, and $K = 2.4$. This set of choices also works asymptotically for large v. Choosing a K that is not too high will take less up compute time, and compute time is of course of the essence.

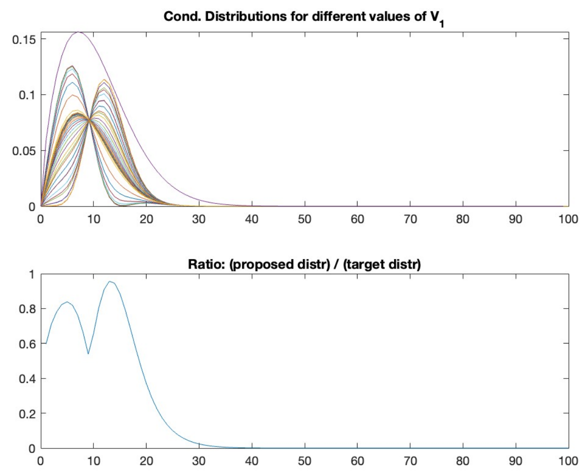

Fig. 8. Proposal distribution for the conditional distribution.

In Figure 9, we see that the samples indeed follow the target distribution.

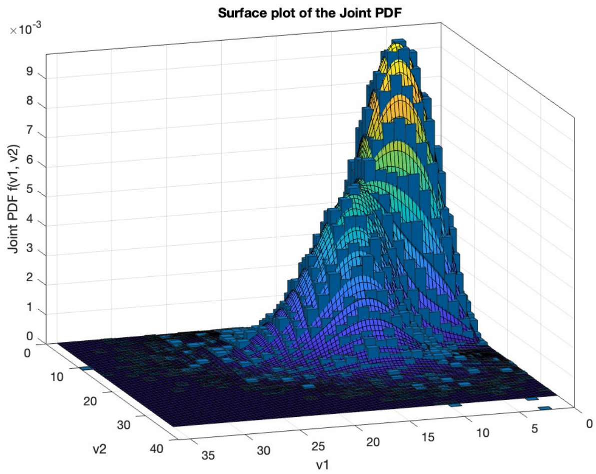

Fig. 9. A combined 2D histogram and surface plot of samples and target distribution.

The point estimate of the covariance is 19.11 (MW)^2.

c: We can rewrite the variance

$$\begin{aligned}\mathbb{V}\{P(V_1)+P(V_2)\} = \\ \mathbb{E}\{(P(V_1)+P(V_2))^2\} - \mathbb{E}\{P(V_1)+P(V_2)\}^2 = \\ \{\text{expanding the brackets and rearranging terms}\} = \\ \mathbb{V}P(V_1) + 2\mathbb{C}(P(V_1),P(V_2)) + \mathbb{V}P(V_2) = \\ 2[\mathbb{V}P(V_1) + \mathbb{C}(P(V_1),P(V_2))] = \\ 2[\mathbb{E}P(V_1)^2 - (\mathbb{E}P(V_1))^2 + \mathbb{C}(P(V_1),P(V_2))]. \quad (30)\end{aligned}$$

We can reuse the covariance from 3(b), as well as the expectation of $P(V_1)$ from 2(d).

The point estimate of the variance is 111.602 (MW)^2.

d: We can estimate the probability $\mathbb{P}(P(V_1)+P(V_2) > 15)$ using the points we have from b). We can also estimate $\mathbb{P}(P(V_1)+P(V_2) < 15)$. Since there is a non-zero probability mass for $P(V_1)+P(V_2) = 15$ (we can find ϵ-balls in which the joint density is constant at 15, for example where $V_1 < 4$ and $11 < V_2 < 25$), the other two events are not complementary and will not sum to 1. We define

$$\begin{aligned}\tau = \mathbb{P}(P(V_1)+P(V_2) > 15) = \\ \mathbb{E}I_{P(V_1)+P(V_2)>15} = \mathbb{E}\phi(V_1,V_2) \quad (31)\end{aligned}$$

where I_A is the indicator function for the event A. We can estimate this expectation as

$$\tau_N = \frac{1}{N}\sum_{i=1}^{N} I_{P(V_{1,i})+P(V_{2,i})>15} = \frac{1}{N}\sum_{i=1}^{N}\phi(V_{1,i},V_{2,i}), \quad (32)$$

and we get the sample variance as

$$s_N^2 = \frac{1}{N}\sum_{i=1}^{N}(I_{P(V_{1,i})+P(V_{2,i})>15} - \bar{I}_{P(V_{1,i})+P(V_{2,i})>15})^2, \quad (33)$$

We can then stratify our estimate. We see that on the set $R_1 := ([4,25]\times[4,25]) \setminus ([4,11]\times[4,11])$, the indicator is always 1, and on the set $R_0 := ([0,\infty]\times[0,\infty]) \setminus ([4,25]\times[4,25])$, the indicator is always 0. That leaves only $[4,11]\times[4,11]$ with nonzero variance. We can find the probability of the zero-variance regions as

$$\begin{aligned}\mathbb{P}(R_1) = F_{V_1,V_2}(25,25) - F_{V_1,V_2}(11,11) - \\ [F_{V_1,V_2}(25,4) - F_{V_1,V_2}(11,4)] - \\ [F_{V_1,V_2}(4,25) - F_{V_1,V_2}(4,11)] \quad (34a)\end{aligned}$$

$$\begin{aligned}\mathbb{P}(R_0) = 1 - [F_{V_1,V_2}(25,25) - F_{V_1,V_2}(25,4) - \\ F_{V_1,V_2}(4,25) + F_{V_1,V_2}(4,4)]. \quad (34b)\end{aligned}$$

Thus, using the law of total probability and using Neyman allocation we find

$$\begin{aligned}\tau_N^{\text{strat}} = 1\cdot\mathbb{P}(R_1) + 0\cdot\mathbb{P}(R_0) + \\ [1-\mathbb{P}(R_1)-\mathbb{P}(R_0)]\frac{1}{N}\sum_{i=1}^{N} I_{P(V_{1,i})+P(V_{2,i})>15}, \quad (35)\end{aligned}$$

with the variance

$$\mathbb{V}\tau_N^{\text{strat}} = \frac{1-\mathbb{P}(R_1)-\mathbb{P}(R_0)}{N}\mathbb{V}_{[4,11]\times[4,11]} I_{P(V_1)+P(V_2)>15} \quad (36)$$

Now, we have time to find out how to sample from this region, which we did not have time to do, but it should be possible with rejection or importance sampling. Instead, we have used $V_1 + V_2$ as a control variate and defined

$$Z = I_{P(V_1)+P(V_2)>15} + \beta(V_1 + V_2 - m), \tag{37}$$

where m is set to the expectation of the control variate, so $2\Gamma(1+\frac{1}{k})\lambda$. We will not bother with tracking the convergence, so we will do one block-estimate of all the parameters in the end. Thus,

$$C = \frac{1}{N}\sum_i I_{P(V_{1,i})+P(V_{2,i})>15}(V_{1,i} + V_{2,i} - m) \tag{38a}$$

$$V = \frac{1}{N}\sum_i (V_1 + V_2 - m)^2 \tag{38b}$$

$$\beta = -\frac{C}{V} \tag{38c}$$

$$Z_i = I_{P(V_1)+P(V_2)>15} + \beta(V_{1,i} + V_{2,i} - m) \tag{38d}$$

$$\tau_N = \frac{1}{N}\sum_i Z_i. \tag{38e}$$

We will then use the CLT and estimate the interval using the variance of the control variate. We saw that the correlation was quite weak, so the variance reduction was very small (probably because the quantified correlation is linear, and there are regions where the indicator is zero for both small and large wind speeds). Perhaps $P(V_{1,i})+P(V_{2,i})$ with the biased mean estimated earlier would have been a better control variate, but we have not had time to look into this.

Finally, the best interval estimate we could produce of the probability of the combined power generation being over 15 MW, and the probability of it being below 15 MW, are found in Table XI.

TABLE XI

99% CONFIDENCE INTERVAL OF THE PROBABILITY OF THE COMBINED PRODUCTION OF WIND TURBINES PRODUCING OVER $15MW$, AND BELOW $15MW$, RESPECTIVELY.

Over 15 MW: Lower bound	0.9592
Over 15 MW: Upper bound	0.9622
Below 15 MW: Lower bound	0.03762
Below 15 MW: Upper bound	0.04078

The probabilities actually sum to 1, as there are no occurrences of 15 MW, but in reality there is a nonzero probability for this.

II. REFERENCES

[1] Nov. 2023. URL: https://en.wikipedia.org/wiki/Inverse_transform_sampling#cite_note-aalto-1.

[2] Magnus Wiktorsson. *Lecture 1, slide 8*. 2024.

[3] Magnus Wiktorsson. *Lecture 4, slide 19*. 2024.

[4] Magnus Wiktorsson. *Lecture 4, slide 6*. 2024.