# MS6051 – Statistical Inference Assignment

Group: X Student IDs: XXX, YYY

**Abstract**

This report presents the results, analysis, and interpretations for the MS6051 Statistical Inference Assignment. All simulation work was performed in R, with code included in the accompanying .txt file.

## Contents

# 1 Question 1 – Exponential Data: Estimation Bias and Efficiency

## 1.1 Maximum Likelihood Estimator

This is the expression for the maximum likelihood estimator (MLE) of the rate parameter $\lambda$ for an exponential sample:

$$\hat{\lambda}_{\text{MLE}} = \frac{1}{\bar{X}}.$$

## 1.2 Generation of a Single MLE Estimate

To ensure reproducibility, we set the seed to the sum of our two student ID numbers:

$$\texttt{set.seed}(25203533 + 25189506).$$

Using $\lambda = 1$ and a sample size of $n = 10$, the resulting MLE estimate is

$$\hat{\lambda}_{n=10} = 0.7068795.$$

## 1.3 Simulation Study (1000 Replicates)

We then estimated the mean, variance, and mean squared error (MSE) of the MLE by repeating the estimation procedure 1000 times. The simulation produced the following results:

$$\widehat{\mathbb{E}}[\hat{\lambda}] = 1.108618, \qquad \widehat{\text{Var}}(\hat{\lambda}) = 0.1429993, \qquad \widehat{\text{MSE}}(\hat{\lambda}) = 0.1546541.$$

**Estimated Mean, Variance, and MSE** For reference, the analytical definitions of the bias and the mean squared error are:

$$\text{Bias}(\hat{\lambda}) = \mathbb{E}[\hat{\lambda}] - \lambda,$$

## 1.4 Modify the sample size effects on Bias, MSE and efficiency

For the sample sizes $n = 20, 50, 200, 400$, and 1000, we generated the corresponding estimates and plotted the bias, efficiency, and MSE.

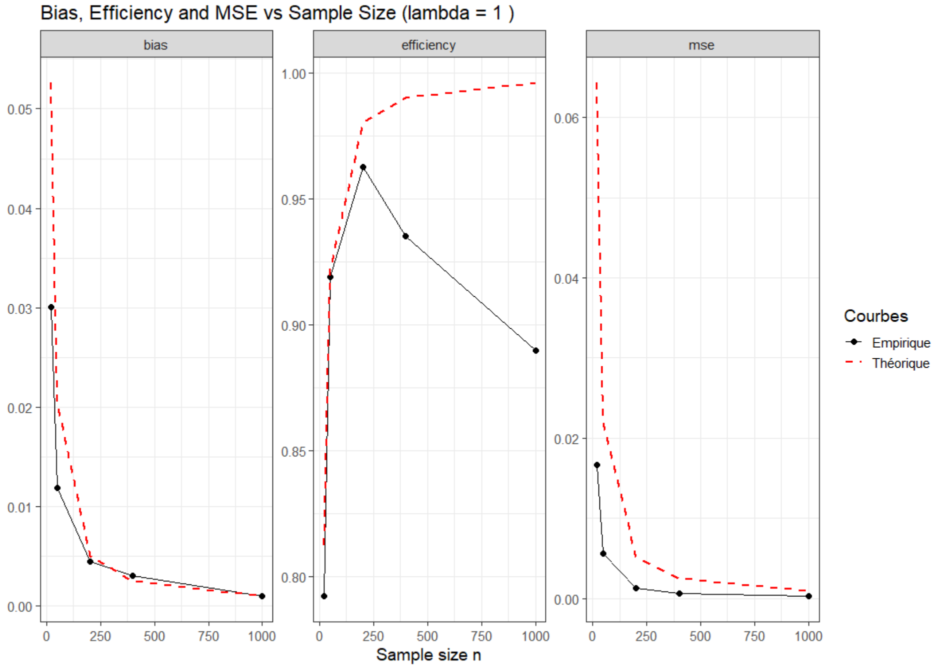


Figure 1: Bias, efficiency, and MSE for different sample sizes.

From these plots, we observe that the bias decreases as the sample size increases. This behaviour is consistent with the analytical expression for the bias of the MLE under an exponential model:

$$\mathrm{Bias}(\hat{\lambda}) = \frac{\lambda}{n-1},$$

which clearly approaches zero as $n \to \infty$.

Similarly, the MSE also decreases with $n$. Using the analytical formula

$$\mathrm{MSE}(\hat{\lambda}) = \frac{(n\lambda)^2}{(n-1)(n-2)} = \lambda^2 \left(1 - \frac{3}{n} + \frac{2}{n^2}\right)^{-1},$$

we see that the MSE converges to $\lambda^2$ as $n$ becomes large.

The behaviour of the efficiency is slightly more involved. Using the closed-form expression

$$\mathrm{Eff}(\hat{\lambda}) = \frac{(n-1)^2(n-2)}{n^3} = 1 - \frac{4}{n} + \frac{5}{n^2} - \frac{2}{n^3},$$

we observe that the efficiency increases monotonically with $n$, approaching 1 as the sample size grows. This confirms that the MLE becomes asymptotically efficient. To conclude on the effect of sample size, increasing $n$ consistently leads to a more accurate estimator for the exponential distribution. This behaviour is fully consistent with the theoretical results discussed in class and is clearly confirmed by our simulation study. The only exception is the efficiency, which does not follow the expected asymptotic pattern in our plots. This can be explained by the sensitivity of the estimator and the randomness of the simulated exponential samples: with a different sample, the empirical efficiency would likely align more closely with the theoretical expression. In other words, the deviation observed here is due to sampling variability rather than a contradiction of the asymptotic result.

### 1.5 Effect of Modifying the Value of $\lambda$

In this section, we modify the rate parameter to $\lambda = 0.5$ and $\lambda = 2$, and repeat the calculations from Section 1.4. We then comment on how changing $\lambda$ influences the bias, MSE, and efficiency of the MLE.

**Case** $\lambda = 0.5$

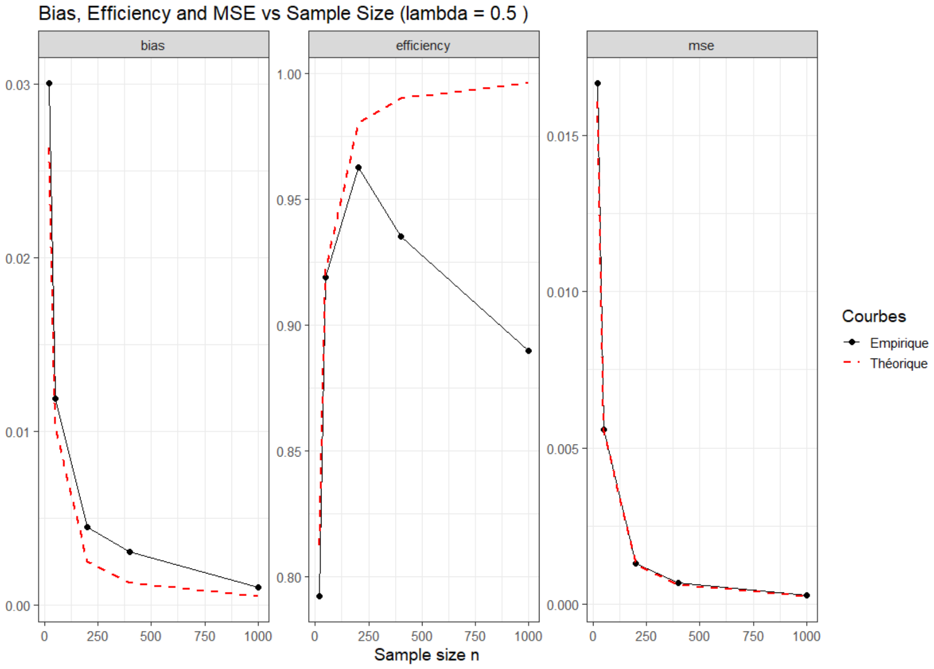


Figure 2: Bias, efficiency, and MSE for $\lambda = 0.5$.

We observe that changing $\lambda$ has no effect on the efficiency, which depends only on the sample size. However, both the bias and the MSE decrease when $\lambda$ is reduced. This is expected: the analytical formulas include $\lambda$ directly for the bias, and $\lambda^2$ for the MSE, so halving $\lambda$ leads to a four-fold reduction in MSE.

**Case $\lambda = 2$**

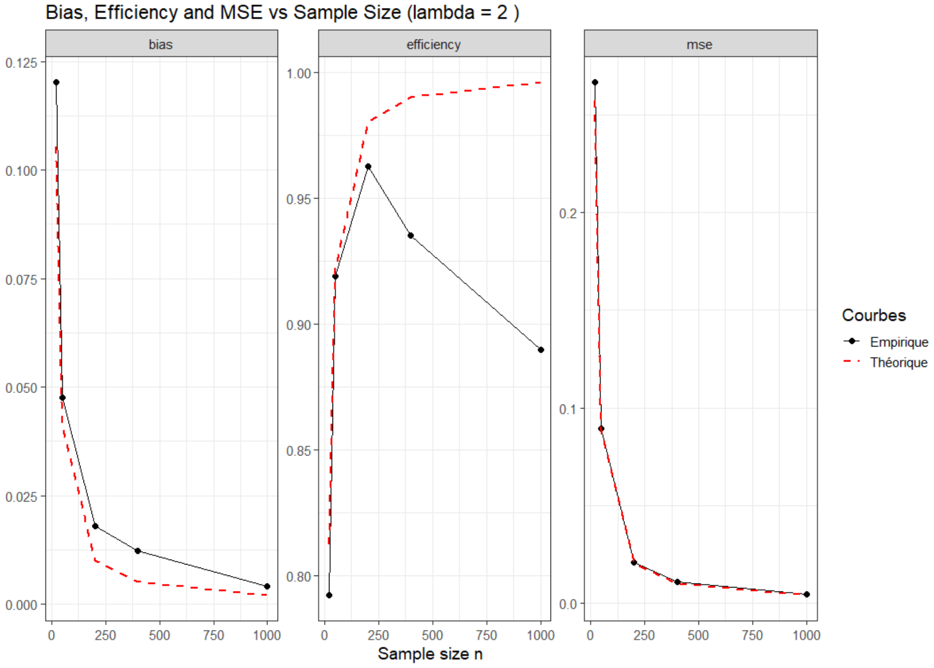


Figure 3: Bias, efficiency, and MSE for $\lambda = 2$.

As expected, increasing $\lambda$ leads to larger bias and larger MSE values compared to the case $\lambda = 1$. Since the MSE grows proportionally to $\lambda^2$, doubling $\lambda$ results in approximately four times larger MSE, which is exactly what we observe here.

Overall, these results illustrate that the rate parameter $\lambda$ has a strong influence on the magnitude of both the bias and the MSE. In practical terms, increasing $\lambda$ (i.e., increasing the expected number of events per unit time) leads to less favourable estimation performance, whereas smaller values of $\lambda$ improve it.

# 2 Question 2 – Comparing MLE and MoM

## 2.1 Density and Random Selection of $x_{\max}$

The PDF is given by

$$f(x) = cx^{\theta-1}, \quad x \in [0, x_{\max}],\, \theta > 0$$

Here $x_{\max}$ is chosen as 8 by running the R code for getting a random value between 2 and 15 (With appropriate seed value being set).
For finding c, we set the integral oof the PDF to 1

$$\int_0^8 cx^{\theta-1}\, dx = 1$$

Integrating with respect to x:

$$c\left[\frac{x^\theta}{\theta}\right]_0^8 = 1$$

Solving for c:

$$c = \frac{\theta}{8^\theta}$$

Now the PDF is given by

$$f(x) = \frac{\theta}{8^\theta} x^{\theta-1}, \quad x \in [0, 8],\, \theta > 0$$

## 2.2 Derivation of MLE and MoM

**Method of Moments Estimator:**
The first population moment is $E[X]$:

$$E[X] = \int_0^8 x \cdot \frac{\theta}{8^\theta} x^{\theta-1}\, dx = \frac{\theta}{8^\theta}\left[\frac{x^{\theta+1}}{\theta+1}\right]_0^8$$

$$E[X] = \frac{\theta}{8^\theta}\frac{8^{\theta+1}}{\theta+1} = \frac{8\theta}{\theta+1}$$

Equating to the first sample moment $\bar{X}$:

$$\frac{8\hat{\theta}_{\text{MoM}}}{\hat{\theta}_{\text{MoM}}+1} = \bar{X}$$

Solving for $\hat{\theta}_{\text{MoM}}$:

$$\hat{\theta}_{\text{MoM}}(8-\bar{X}) = \bar{X}$$

$$\boxed{\hat{\theta}_{\text{MoM}} = \frac{\bar{X}}{8-\bar{X}}}$$

**Maximum Likelihood Estimator Estimator:**
The log-likelihood function $\ell(\theta)$ is:

$$\ell(\theta) = n\ln\theta - n\theta\ln 8 + (\theta-1)\sum_{i=1}^{n}\ln x_i$$

Taking the first derivative with respect to $\theta$:

$$\frac{\partial\ell}{\partial\theta} = \frac{n}{\theta} - n\ln 8 + \sum_{i=1}^{n}\ln x_i$$

Setting the derivative to zero and solving for $\hat{\theta}_{\text{MLE}}$:

$$\frac{n}{\hat{\theta}_{\text{MLE}}} = n\ln 8 - \sum_{i=1}^{n}\ln x_i$$

$$\boxed{\hat{\theta}_{\text{MLE}} = \frac{n}{n\ln 8 - \sum_{i=1}^{n}\ln x_i}}$$

## 2.3 Inverse CDF and Variable Generation

**Cumulative Distribution Function (CDF):**
The CDF $F(x)$ is found by integrating the PDF

$$f(t;\theta) = \frac{\theta}{8^\theta}t^{\theta-1}$$

$$F(x) = \frac{\theta}{8^\theta}\int_0^x t^{\theta-1}\, dt$$

$$F(x) = \frac{\theta}{8^\theta}\left[\frac{t^\theta}{\theta}\right]_0^x$$

The CDF is given by

$$F(x) = \left(\frac{x}{8}\right)^{\theta}, \quad \text{for } x \in [0, 8]$$

**Variable generation using Inverse CDF:**
To find the inverse CDF $F^{-1}(u)$, we set $u = F(x)$, where u $\sim$ Uniform(0,1) and solve for $x$:

$$u = \left(\frac{x}{8}\right)^{\theta}$$

Solving for $x$,

$$x = F^{-1}(u) = 8u^{1/\theta}$$

**Sample Generation and Distribution:**
Samples of 5000 were generated for different values of $\theta$ = (0.5, 1, 2, 4). The distribution of these samples are as follows

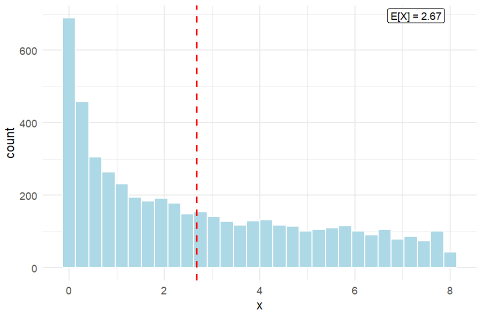


Figure 4: Distribution for $\theta = 0.5$

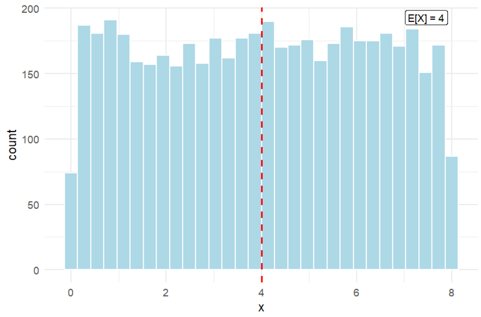


Figure 5: Distribution for $\theta = 1$

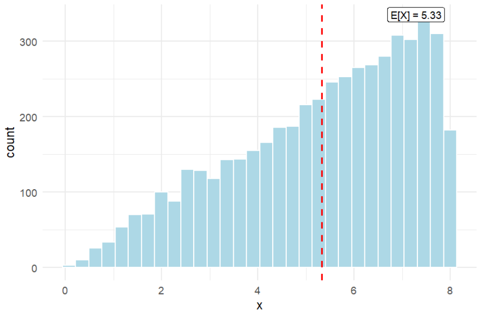


Figure 6: Distribution for $\theta = 2$

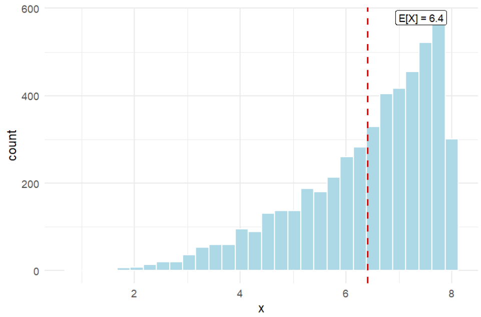


Figure 7: Distribution for $\theta = 4$

The plots clearly show that for $\theta < 1$, the distribution is right skewed. For $\theta > 1$, the distributions are left skewed. For $\theta = 1$, the distribution is uniform.

## 2.4 Estimators for a random sample

A random sample of size 100 was generated with $\theta = 1$ and the Method of Moments and Maximum Likelihood estimators are calculated.

$$\boxed{\hat{\theta}_{\text{MoM}} = 1.177}$$

$$\boxed{\hat{\theta}_{\text{MLE}} = 1.038}$$

## 2.5 Log-likelihood plot for different samples

The log-likelihood function $\ell(\theta)$ is:

$$\ell(\theta) = n \ln \theta - n\theta \ln 8 + (\theta - 1) \sum_{i=1}^{n} \ln x_i$$

Additional samples with sample sizes n = 10 and n = 100 are generated. A plot of $\theta$ vs $\ell(\theta)$) was generated for all three samples and the MLE and MOM estimates are indicated on these plots.

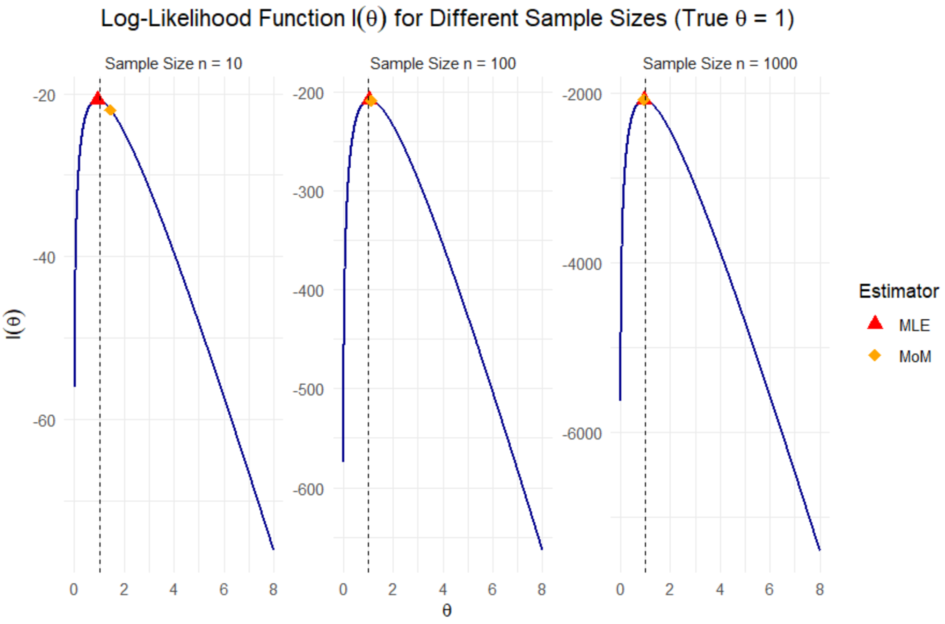


Figure 8: Log-Likelihood Function and Estimators Comparison

## 2.6 Wilks Confidence Interval

The threshold Likelihood for calculating the Wilks CI is given by

$$\ell_{\text{threshold}} = \ell(\hat{\theta}_{\text{MLE}}) - \frac{d}{2}$$

Here d = $\chi^2_{1,0.05}$
In the plots, a horizontal line was drawn at the threshold likelihood values. It meets the log-likelihood function at two points. Two vertical lines are drawn at these points. These lines define the Wilks Confidence Interval in this case.

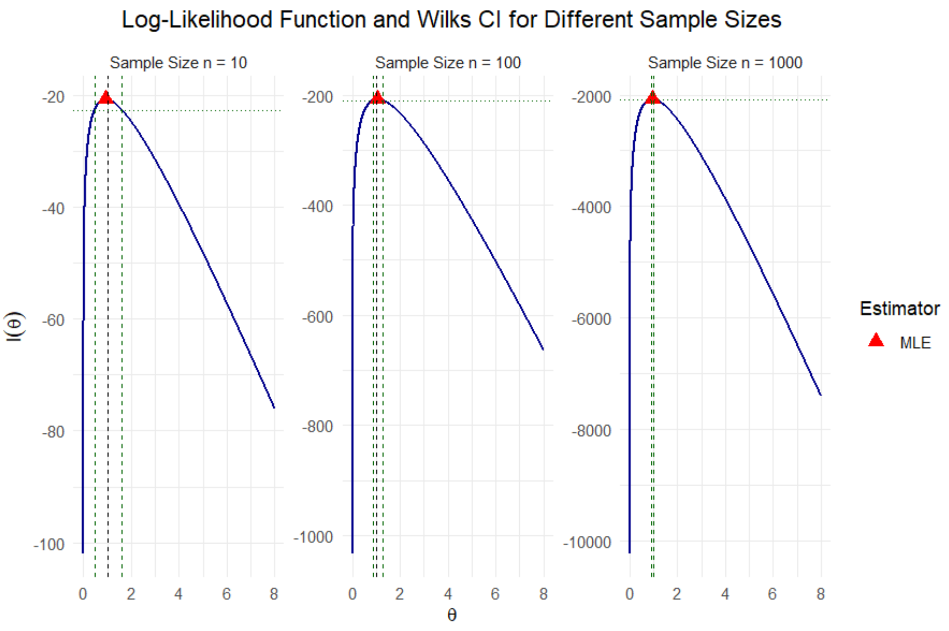
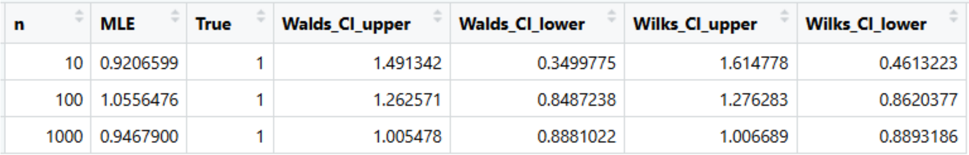


Figure 9: Log-Likelihood Function and Wilks Confidence Interval

## 2.7 Walds Confidence Interval and its comparison with Wilks

For all the three sample sizes, the MLE was calculated using the nlm function, which minimizes the minus log-likelihood function. Using the observed information calculated using nlm (by setting hessian=true), the 95% Walds Confidence Intervals are computed for each sample size. The Walds Confidence Interval and the previously computed Wilks Confidence Interval are compared below,

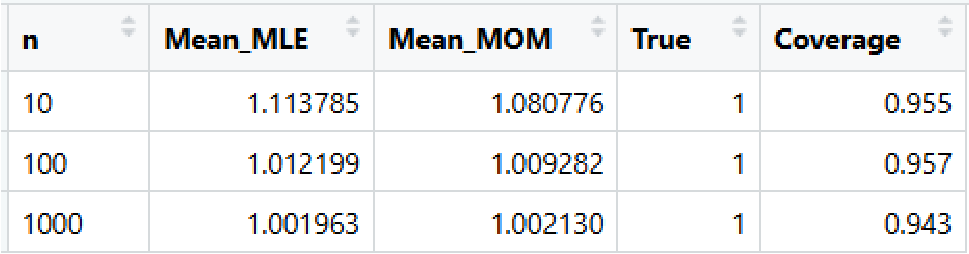


| n | MLE | True | Walds_CI_upper | Walds_CI_lower | Wilks_CI_upper | Wilks_CI_lower |
|---|---|---|---|---|---|---|
| 10 | 0.9206599 | 1 | 1.491342 | 0.3499775 | 1.614778 | 0.4613223 |
| 100 | 1.0556476 | 1 | 1.262571 | 0.8487238 | 1.276283 | 0.8620377 |
| 1000 | 0.9467900 | 1 | 1.005478 | 0.8881022 | 1.006689 | 0.8893186 |

Figure 10: Confidence Interval Comparison

## 2.8 Computing Estimators, Confidence Intervals and Coverage

For each of three sample sizes, the MLE and the Walds CI was computed using nlm for 1000 times. Also, the MOM estimators were calculated for each time. Based on this, the MLE and the MOM Estimators were calculated by taking the mean value. Finally, based on the Walds CI, the coverage was calculated for each sample size. The results are as follows,



| n | Mean_MLE | Mean_MOM | True | Coverage |
|---|---|---|---|---|
| 10 | 1.113785 | 1.080776 | 1 | 0.955 |
| 100 | 1.012199 | 1.009282 | 1 | 0.957 |
| 1000 | 1.001963 | 1.002130 | 1 | 0.943 |

Figure 11: Wald Confidence Interval Coverage

# 3 Question 3 – Bootstrapping

## 3.1 Explanation of Bootstrapping

A bootstrapped dataset is generated by choosing a random set of values and creating a new dataset from the previously chosen dataset by selecting random values (including duplicate values). Now, bootstrapping is the process of preparing a sample of bootstrapped dataset, calculating any statistic for each sample and then keeping track of those statistics to estimate confidence intervals, standard errors or the variance of the estimator.

## 3.2 Bootstrap CI for Exponential Median

An exponential data with sample size 100 and rate 1 was generated. This data was bootstrapped 1000 times and the median value was calculated each of the time. A 95% confidence interval was calculated based on the 0.025 and the 0.975 quantiles of the distribution of the bootstrapped medians.
The confidence intervals are

$$95\%\, CI : (0.54186,\ 0.88641)$$

## 3.3 Performance Evaluation of the Confidence Intervals

The formula for true median:

$$m = \frac{\ln 2}{\lambda}$$

Here, rate $\lambda$ is 1. The true median value is calculated

$$m = 0.6931$$

The bootstrapping was repeated 1000 times and the performance of the confidence intervals (calculated using quantiles) was evaluated.

$$Coverage = 0.931$$

## 3.4 Performance Evaluation of the Wald's Confidence Intervals

An approximate $(1-\alpha)$ level Wald confidence interval for $\lambda$ is

$$\hat{\lambda} \pm z_{\alpha/2}\sqrt{I(\hat{\lambda})^{-1}}.$$

The observed information is given by

$$I(\hat{\lambda}) = \frac{n}{\hat{\lambda}^2}.$$

The Wald's confidence interval for $\lambda$ in this case is given by

$$\left(\hat{\lambda} - z_{\alpha/2}\frac{\hat{\lambda}}{\sqrt{n}},\ \hat{\lambda} + z_{\alpha/2}\frac{\hat{\lambda}}{\sqrt{n}}\right)$$

For 95% confidence interval, the Wald's confidence interval for $\lambda$ is

$$\boxed{\left(\hat{\lambda} - 1.96\frac{\hat{\lambda}}{\sqrt{n}},\ \hat{\lambda} + 1.96\frac{\hat{\lambda}}{\sqrt{n}}\right)}$$

The MLE for $\lambda$ is $\hat{\lambda} = 1/\bar{X}$ and, by standard asymptotic theory,

$$\hat{\lambda} \overset{approx}{\sim} N\left(\lambda,\ \frac{\lambda^2}{n}\right).$$

The median in this case is

$$m = g(\lambda) = \frac{\ln 2}{\lambda}.$$

Then

$$g'(\lambda) = -\frac{\ln 2}{\lambda^2}.$$

By the Delta Method,

$$g(\hat{\lambda}) \overset{approx}{\sim} N\left(g(\lambda),\ [g'(\lambda)]^2 \frac{\lambda^2}{n}\right).$$

Therefore,

$$\mathrm{Var}(g(\hat{\lambda})) \approx \left(-\frac{\ln 2}{\lambda^2}\right)^2 \cdot \frac{\lambda^2}{n} = \frac{(\ln 2)^2}{\lambda^2 n}.$$

Now, the standard error is given by

$$\mathrm{SE}(\hat{m}) \approx \frac{\ln 2}{\lambda\sqrt{n}}.$$

Replacing the unknown $\lambda$ by $\hat{\lambda}$ yields

$$\widehat{\mathrm{SE}}(\hat{m}) = \frac{\ln 2}{\hat{\lambda}\sqrt{n}} = \frac{\hat{m}}{\sqrt{n}}.$$

Therefore an approximate Wald 95% confidence interval for the median is

$$\boxed{\left(\hat{m}\ -\ 1.96\,\frac{\hat{m}}{\sqrt{n}},\ \hat{m}\ +\ 1.96\,\frac{\hat{m}}{\sqrt{n}}\right)}$$

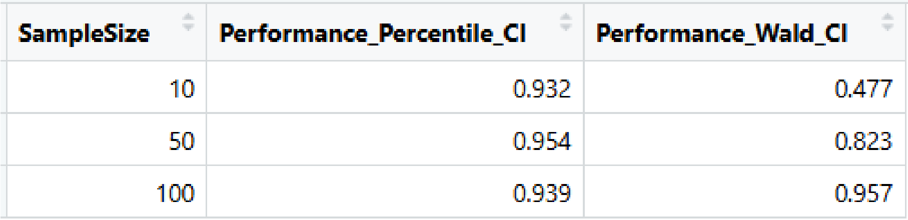
The bootstrapping was repeated 1000 times and the performance of the confidence intervals (calculated using the above formula) was evaluated

$$Coverage\,(Wald'sCI) = 0.957$$

### 3.5 Performance Evaluation of both the Confidence Intervals for different sample sizes

Additionally, the performance evaluation was carried out by repeating the bootstrapping procedure for 1000 times for different sample sizes (n).

The performance of both types of Confidence Intervals for the median was evaluated for each sample size.



| SampleSize | Performance_Percentile_CI | Performance_Wald_CI |
|---:|---:|---:|
| 10 | 0.932 | 0.477 |
| 50 | 0.954 | 0.823 |
| 100 | 0.939 | 0.957 |

The Percentile CI performed consistently across different sample sizes. The coverage was around the 95% level even for smaller sample sizes. This is because it adapts well to the actual sampling distribution of the estimator.

The Wald (Delta-method) confidence interval performed very poorly for small samples: at n=10, where the coverage is around 0.48, indicating severe under-coverage due to the strong skewness

of the sampling distribution of the median. With large sample sizes, the Wald CI improved. This is also due to the fact that the Wald CI for median was calculated using Delta method and it works in case of larger sample sizes.
These results highlight that the Wald CI is unreliable for skewed estimators in small or moderate samples, whereas the percentile bootstrap CI provides more robust performance.

# 4 Question 4 – Statistical Neural Network Regression

## 4.1 Log-likelihood Derivation

We assume the regression model

$$Y_i \mid X_i \sim \mathcal{N}(\mu_i, \phi),$$

where $\mu_i = \mu(X_i)$ is the prediction from the neural network and $\phi = \mathrm{Var}(Y_i)$.

The density of $Y_i$ is therefore

$$f(y_i \mid \mu_i, \phi) = \frac{1}{\sqrt{2\pi\phi}} \exp\left( -\frac{(y_i - \mu_i)^2}{2\phi} \right).$$

The log-likelihood for a sample of size $n$ is

$$\ell(\phi) = \sum_{i=1}^{n} \log f(y_i \mid \mu_i, \phi) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\phi) - \frac{1}{2\phi}\sum_{i=1}^{n}(y_i - \mu_i)^2.$$

To find the MLE of $\phi$, we differentiate:

$$\frac{\partial \ell}{\partial \phi} = -\frac{n}{2\phi} + \frac{1}{2\phi^2}\sum_{i=1}^{n}(y_i - \mu_i)^2.$$

Setting this to zero,

$$-\frac{n}{2\phi} + \frac{1}{2\phi^2}\sum_{i=1}^{n}(y_i - \mu_i)^2 = 0$$

$$\Longrightarrow \quad \hat{\phi} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_i)^2 = \mathrm{MSE},$$

that is, the MLE of $\phi$ equals the mean squared prediction error.

## 4.2 Log-likelihood After Substituting MSE

Substituting $\hat{\phi} = \mathrm{MSE}$ into the log-likelihood expression gives:

$$\ell = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\mathrm{MSE}) - \frac{1}{2\,\mathrm{MSE}}\sum_{i=1}^{n}(y_i - \mu_i)^2.$$

But since

$$\sum_{i=1}^{n}(y_i - \mu_i)^2 = n \cdot \mathrm{MSE},$$

the last term simplifies to

$$-\frac{1}{2\,\mathrm{MSE}}\, n \cdot \mathrm{MSE} = -\frac{n}{2}.$$

Thus the log-likelihood becomes

$$\ell = -\frac{n}{2}\big( \log(2\pi) + \log(\mathrm{MSE}) + 1 \big).$$

This is the simplified form used to compute the BIC in neural network regression.

## 4.3 Dataset Description

The dataset selected for this analysis is the *Productivity Prediction of Garment Employees* dataset. It contains 1197 instances and 14 features, including a mixture of numerical, categorical, date-based, and continuous variables. The goal of this dataset is to model and understand employee productivity based on a wide range of operational factors such as work schedules, team organisation, and production workload.

### Numerical Variables

The dataset includes several numerical predictors describing various aspects of production activity: `targeted_productivity` (daily productivity target), `smv` (task complexity in standard minutes), `wip` (work in progress), `over_time` (overtime worked), `incentive` (bonus payments), `idle_time` and `idle_men` (measures of downtime), `no_of_style_change` (number of style changes), and `no_of_workers` (team size).

### Categorical Variables

The categorical predictors are: `day` (weekday), `quarter` (period within the month), `department` (production department), `team` (team identifier), and `date`, which is encoded into dummy variables using `model.matrix`.

### Exploratory Data Analysis

We begin with a histogram of the response variable, productivity:

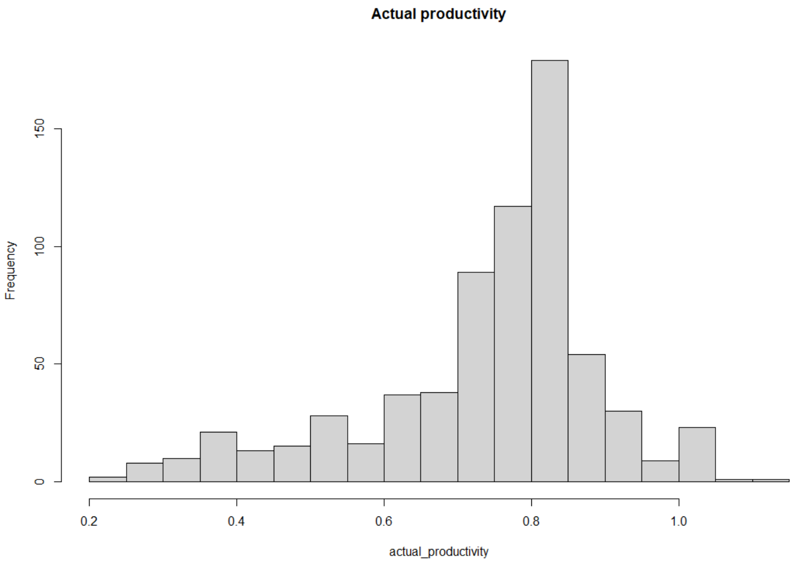
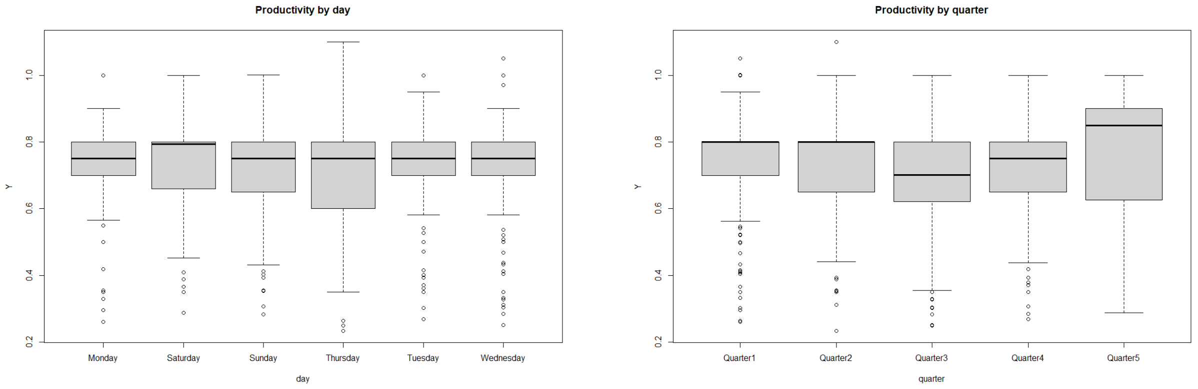


Figure 12: Histogram of employee productivity.

The histogram is centred slightly to the right, indicating that employees typically achieve productivity levels around 70–80%.

We then examine productivity across different days of the week and across the monthly quarters:



(a) Productivity by day of the week

(b) Productivity by quarter

Figure 13: Productivity patterns across days and quarters.

Two patterns emerge. First, productivity tends to be lower during weekends. This could be due to a reduced workforce, lighter workload, or differences in employee motivation. Second, Thursdays show the largest spread in productivity, as indicated by the wider whiskers, suggesting that this day has greater operational variability. Regarding the monthly quarters, Quarter 5 displays the highest variability, which may reflect fluctuations in workload intensity or team availability.

Finally, we analyse the relationship between targeted productivity and actual productivity:

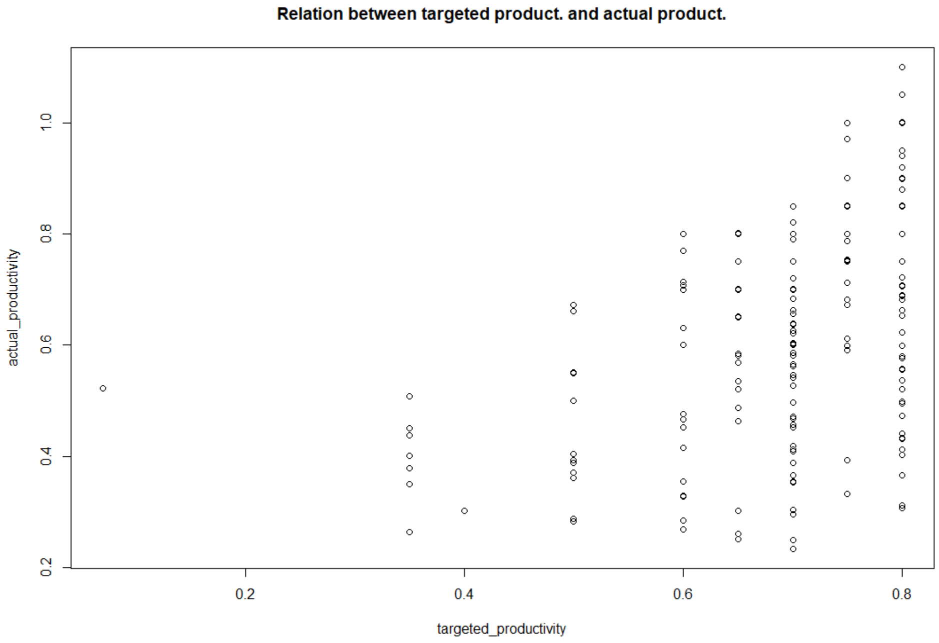
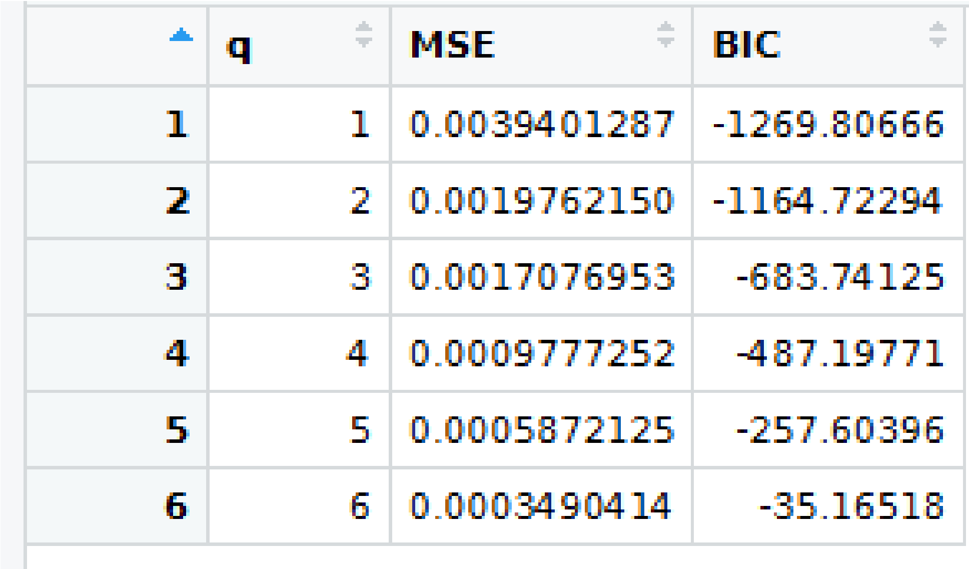


Figure 14: Actual vs. targeted productivity.

We observe that actual productivity tends to increase with the targeted productivity, although the relationship is not strictly linear. The substantial vertical spread for each targeted value indicates that the target alone cannot fully explain actual productivity levels; additional factors must play a significant role.

### 4.4 Model Selection Using BIC

The goal of this section is to fit a neural network to the dataset in order to predict employee productivity using all available features. We evaluate models with different numbers of hidden neurons and identify the value of $q$ that provides the best performance according to the Bayesian Information Criterion (BIC). After selecting the optimal neural network, we compare its BIC to that of a standard linear regression model.



|  | q | MSE | BIC |
|---|---|---|---|
| 1 | 1 | 0.0039401287 | -1269.80666 |
| 2 | 2 | 0.0019762150 | -1164.72294 |
| 3 | 3 | 0.0017076953 | -683.74125 |
| 4 | 4 | 0.0009777252 | -487.19771 |
| 5 | 5 | 0.0005872125 | -257.60396 |
| 6 | 6 | 0.0003490414 | -35.16518 |

Figure 15: BIC values for neural networks with different numbers of hidden neurons.

From the figure, we observe that the neural network with a single hidden neuron achieves the lowest BIC. This model strikes the best balance between predictive accuracy and complexity: although networks with more neurons may yield slightly lower MSE values, the increase in the number of parameters is not justified from a BIC perspective. In other words, increasing complexity does not provide a sufficient improvement in model fit to compensate for the penalty imposed by BIC.

For comparison, we also fitted a standard linear regression model, obtaining a BIC of

$$\mathrm{BIC}_{\mathrm{LM}} = -1256.454, \qquad \mathrm{MSE}_{\mathrm{LM}} = 0.004093744.$$

The neural network with one hidden neuron performs slightly better than the linear regression model, achieving a lower BIC while maintaining a comparable prediction error. This suggests that a small amount of nonlinearity captured by a single neuron is beneficial for modelling productivity, whereas adding additional neurons does not meaningfully improve performance.

### 4.5 Variable Importance via BIC

In this section, we aim to identify the most important predictors using the BIC criterion. Following the instructions of the assignment, the idea is to remove each variable one at a time and examine how the BIC changes. If removing a variable worsens the BIC (i.e., produces a much lower BIC value), then that variable is considered important.

It is worth noting that many features in this dataset (such as individual date dummies or team identifiers) do not have meaningful standalone interpretations. For this reason, we focus our analysis on the variables that are more interpretable from a production-management perspective.

| **Variable** | **BIC without variable** | ΔBIC |
|---|---|---|
| *Most important interpretable variables (largest effect on BIC)* | | |
| incentive | -862.05 | -60.98 |
| targeted_productivity | -982.19 | -181.12 |
| smv | -1205.13 | -404.06 |
| idle_men | -1205.61 | -404.54 |
| no_of_workers | -1257.07 | -456.00 |
| wip | -1270.13 | -469.06 |
| idle_time | -1270.21 | -469.14 |
| over_time | -1271.17 | -470.10 |
| no_of_style_change | -1276.32 | -475.25 |

Table 1: Most influential interpretable predictors based on change in BIC.

These variables are the most meaningful predictors because they correspond to actual operational factors that influence productivity—unlike date, team, or quarter dummy variables, which are numerous and do not carry inherent interpretability. The strong negative ΔBIC values indicate that removing these variables significantly worsens model performance, confirming their importance for predicting productivity.

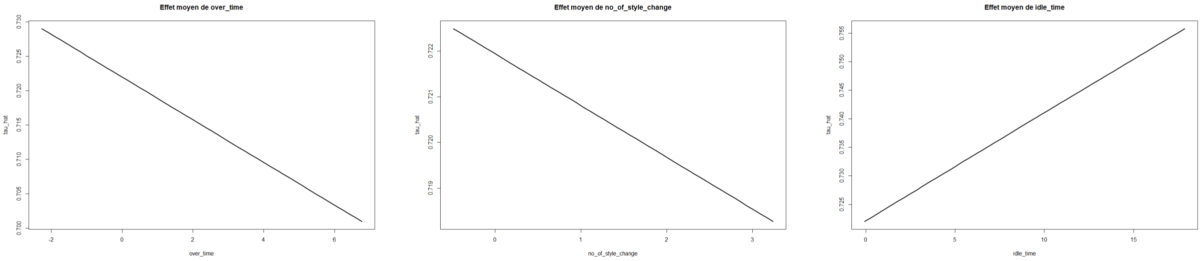
### 4.6 $\hat{\tau}$ Interpretation

Based on the BIC analysis, the three most influential and interpretable predictors are `idle_time`, `over_time`, and `no_of_style_change`. These variables correspond to operational factors that logically affect productivity, and they were selected because removing them produced the largest deterioration in BIC.

Before visualising their effects, we may anticipate the direction of influence:

- `over_time` is expected to reduce productivity, as extended working hours often decrease efficiency.
- `idle_time` may also lower productivity, as it represents periods during which production is halted.
- `no_of_style_change` (the number of style switches) is expected to be negatively associated with productivity, because frequent changeovers interrupt workflow.

To confirm these hypotheses, we compute the $\hat{\tau}$ effect measure for each variable and visualise their relationship with productivity:



(a) $\hat{\tau}$ effect of idle_time (b) $\hat{\tau}$ effect of over_time (c) $\hat{\tau}$ effect of no_of_style_change

Figure 16: Estimated $\hat{\tau}$ effects for the three most influential variables.

The results are quite insightful. As expected, both `no_of_style_change` and `over_time` exhibit a clear negative relationship with productivity: the more style changes or overtime hours occur, the lower the predicted productivity becomes. This aligns with practical intuition, since both factors disrupt the workflow or reflect increased strain on workers.

Interestingly, `idle_time` shows a strong positive relationship with productivity. This suggests that brief periods of idleness—possibly corresponding to breaks or pauses in production—may actually improve performance. This interpretation is plausible: short breaks can help employees recover and therefore work more efficiently afterwards, leading to higher overall productivity.

Overall, the $\hat{\tau}$ visualisations highlight that not all sources of downtime negatively affect productivity: structured or beneficial idle periods can have a positive impact, whereas operational disruptions such as overtime and style changes clearly reduce efficiency.

### 4.7 Prediction Intervals

| **Observed** $Y$ | $\hat{\mu}_{\text{NN}}$ | **Lower 95% PI** | **Upper 95% PI** |
|---|---|---|---|
| 0.9407 | 0.9288 | 0.7561 | 1.1015 |
| 0.8006 | 0.7583 | 0.5856 | 0.9311 |
| 0.8006 | 0.7921 | 0.6194 | 0.9648 |
| 0.8004 | 0.8261 | 0.6534 | 0.9988 |
| 0.8001 | 0.7720 | 0.5993 | 0.9447 |
| 0.7537 | 0.7299 | 0.5572 | 0.9026 |

Table 2: Observed productivity values, neural network predictions, and 95% prediction intervals.

| **LM Prediction** | **Lower 95% PI** | **Upper 95% PI** |
|---|---|---|
| 0.9493 | 0.8083 | 1.0902 |
| 0.7521 | 0.6097 | 0.8945 |
| 0.7920 | 0.6520 | 0.9320 |
| 0.8219 | 0.6789 | 0.9648 |
| 0.7689 | 0.6288 | 0.9090 |
| 0.7307 | 0.5905 | 0.8708 |

Table 3: Linear regression predictions and 95% prediction intervals.

As we can see, the predicted values from the neural network and the linear model are very close to each other, indicating that both approaches capture the main structure of the data. However, the prediction intervals from the linear model are noticeably wider than those produced by the neural network. This suggests that the neural network provides more precise estimates of productivity in most cases, reflecting its slightly better fit.

Nevertheless, the linear model remains an attractive option due to its simplicity and its strong overall performance. Despite being less flexible, it achieves predictions that are almost as accurate as those of the neural network, making it a reasonable choice when interpretability or model simplicity is a priority.

# 5 Question 5 – Significance Article Summary

## Summary of the Article: *Climate Change Attribution: An Explainer*

*Climate Change Attribution: An Explainer*, written by Dario Domingo, Andrew Parnell, and David B. Stephenson, discusses how we can statistically determine the influence of human activity on climate change. To achieve this, statisticians use observational data, statistical models, and climate model simulations.

The first part of the article talks about attribution of climate change, which could be defined as the process of using statistical and climate models to separate the effects of human actions from natural causes on climate events.

The next part of the article focuses on the attribution of rare extreme events. The objective is to estimate how human influence increases the probability or intensity of events such as heatwaves, storms, or floods. Statisticians model the distribution of these extreme events as a function of global warming and quantify how climate change affects their likelihood. The type of statistical modelling employed here is generally known as Extreme value Distribution.

In the next part, the authors explain how to interpret statements such as "the heatwave was twice as likely because of climate change". These conclusions rely on counterfactual scenarios, in which scientists simulate how an event would evolve with and without human influence. However, this type of modelling involves substantial uncertainty because extreme events are rare even if they become twice as likely. Therefore, statisticians use confidence intervals and estimate which scenarios are most consistent with human-induced climate change.

The last part focuses on the attribution of long-term climate trends using historical records and climate simulations. This includes detection, attribution, and the identification of optimal indices. Detection is essentially a hypothesis test used to assess whether observed trends exceed natural climate variability. Attribution aims to determine whether these trends are due to human or natural factors. For example, if a model without human influence cannot reproduce the observed warming, the warming is statistically attributed to human activity. Finally, the authors describe how scientists create special indicators, known as optimal indices, to clearly isolate the human impact on the climate by focusing on spatial or temporal patterns predicted by climate models.