

REAL CASE

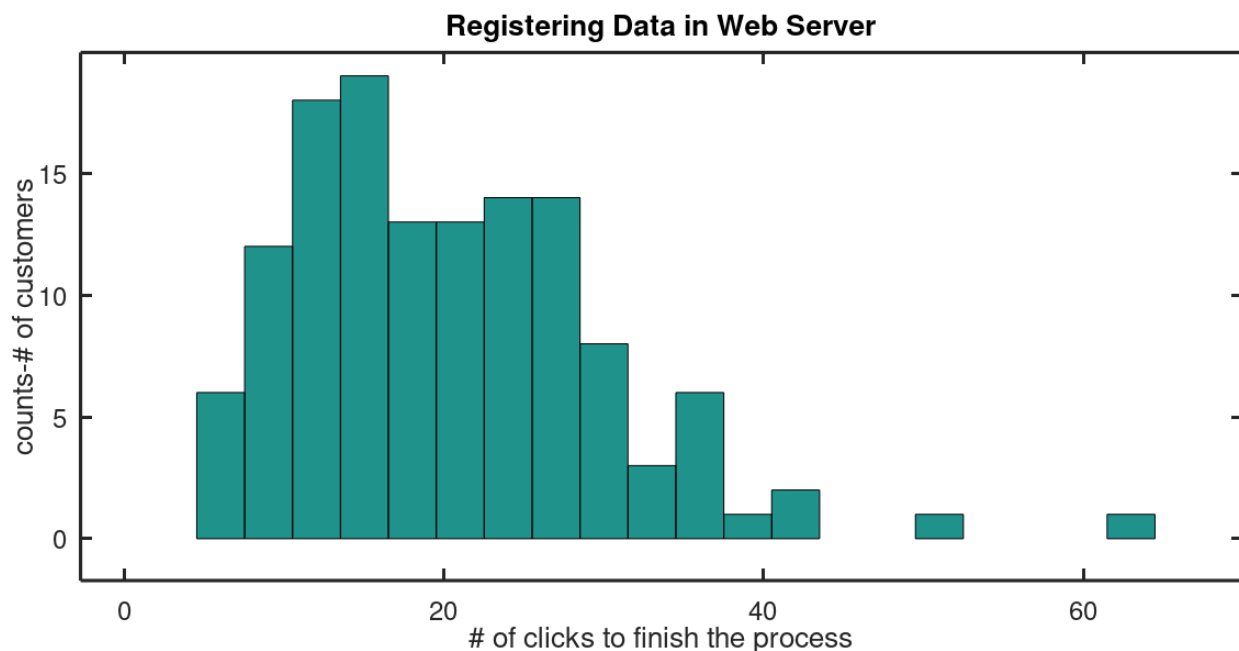
Context: Web server registers data from customers, some already finished the process, others still have missing data.

Goal: Select the list of customers to remind them to finish the process of registering data.

Number of users: more than 150

Input Data: number of times that a user pressed click to save data and finish the process with success. (131 customers = 131 data)

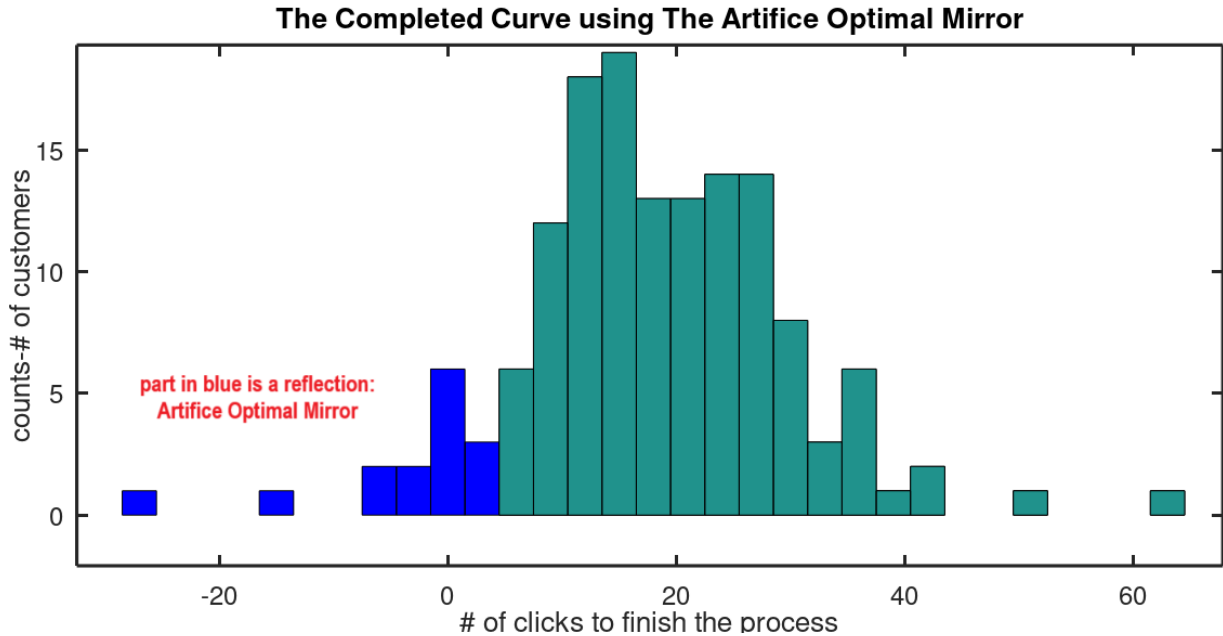
The histogram of the data:



It can be seen that the data is truncated for values lower than 5, it makes sense, because there cannot be negative number of clicks and there is a minimum number of clicks to finish the process.

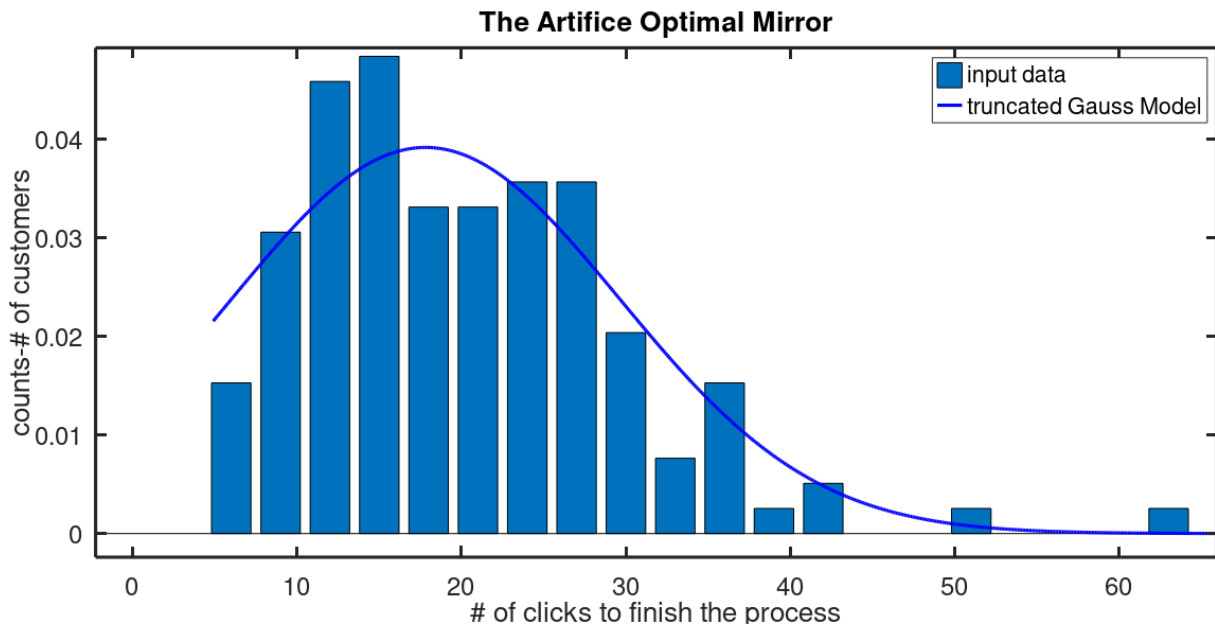
SOLUTION

The Artifice Optimal Mirror [1] can be applied to complete the truncated curve (the data):



These can be calculated: $\mu = 17.82$ (mean) $\sigma = 11.82$ (standard deviation)

The truncated Gauss Model can be plotted:

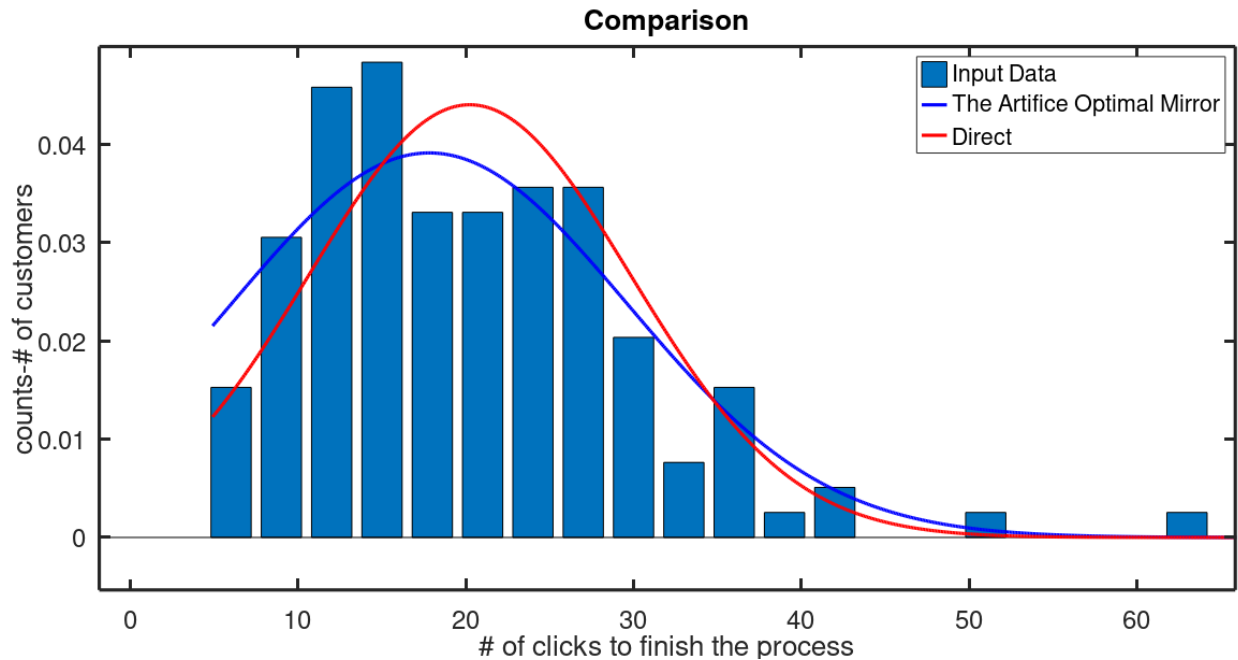


With the model already defined, many criteria can be chosen. One criterion can be to select the list of customers which have lower clicks than the average (17.82) and send a communication to tell them to finish the process soon.

What would happen if the Artifice Optimal Mirror is not used, and the direct value of parameters of the input data are used instead?

The direct parameters of the input data are:

$\mu=20.23$ (mean) $\sigma=9.59$ (standard deviation)



The error for the mean = $\frac{20.23-17.82}{17.82} * 100\% = 13.52\%$

The error for the standard deviation = $\frac{11.82-9.59}{11.82} * 100\% = 18.87\%$

The errors are significant. The direct standard deviation is lower than the standard deviation from the Artifice Optimal Mirror. In the other hand, the direct mean is higher than the mean from the Artifice Optimal Mirror.

Note: The calculations were performed in Excel, the graphs were generated in Octave.

References

- [1] C. San Roman. The Practical Guide To Truncated Probability Distributions. 2020.
<https://sudamericaciencia.org/PracGuideTrunc.html>

