Chapter 5

Estimation of Confidence Intervals for Nodal Hourly Power Consumption

The aim of this chapter is to estimate a confidence interval for nodal hourly power consumption where it is not measured. The nonparametric bootstrap method developed in the previous chapter is used to carry out this calculation. Once every value of hourly power consumption is known at each node, either by the value measured or by the confidence interval calculated, load flow studies can be executed. More precisely, two load flow studies can be performed for a particular hour of the year: the first one will use the values measured as well as the lower bounds computed for each confidence interval. The second one will use the measured values and the upper bounds calculated for each confidence interval. From these first calculations, confidence intervals for lines power flows as well as for Ohmic losses will be inferred. These intervals will be finally used to find the network configuration which would minimize Ohmic losses or avoid lines overload.

In the following calculations, the mode estimator was chosen as a typical value of the nodal power consumption. It is an interesting estimator for an asymmetric distribution because it is the most likely value for this distribution. It is also the value around which there is the largest data concentration. Among the different estimator of the mode available, a nonparametric estimator was chosen. The method to calculate this estimator is based on the histogram. For each resample, the value of the mode will be estimated as follows

Step 1. Order the resample data by increasing values.

Step 2. Create bins for this ordered resample by dividing its range in small intervals of equal size.

Step 3. Find the bin containing the largest number of data. If this one can still be divided, go back to step 2 in order to create smaller bins.

Step 4. Calculate the midpoint of this interval. It is the mode estimate.

The method to calculate confidence intervals is described through three examples presented next.

5.1-Example 1

In this example, data processed in Example 1 of Paragraph 3.3.1, Chapter 3 is used. Let us recall that the calculation is carried out for a given node and a given hour. The node consists of fifty customers from the *H*eat *R*esidential class, forty from the *N*on-*H*eat *R*esidential class and ten from the *C*ommercial *L*oad class. The hour of the year chosen is the third of January 1. One hundred nodal hourly power consumption values form the basic dataset. A program that calculates 95% confidence intervals is proposed in Appendix F. The nonparametric bootstrap method proceeds as follows. First, two thousand resamples of one hundred data are drawn with replacement from the basic sample. Second, for each resample the mode is calculated. Then, the relative frequency distribution of the mode estimator can be built from these two thousand estimated values. This relative frequency distribution is an estimate of the mode's sampling distribution.

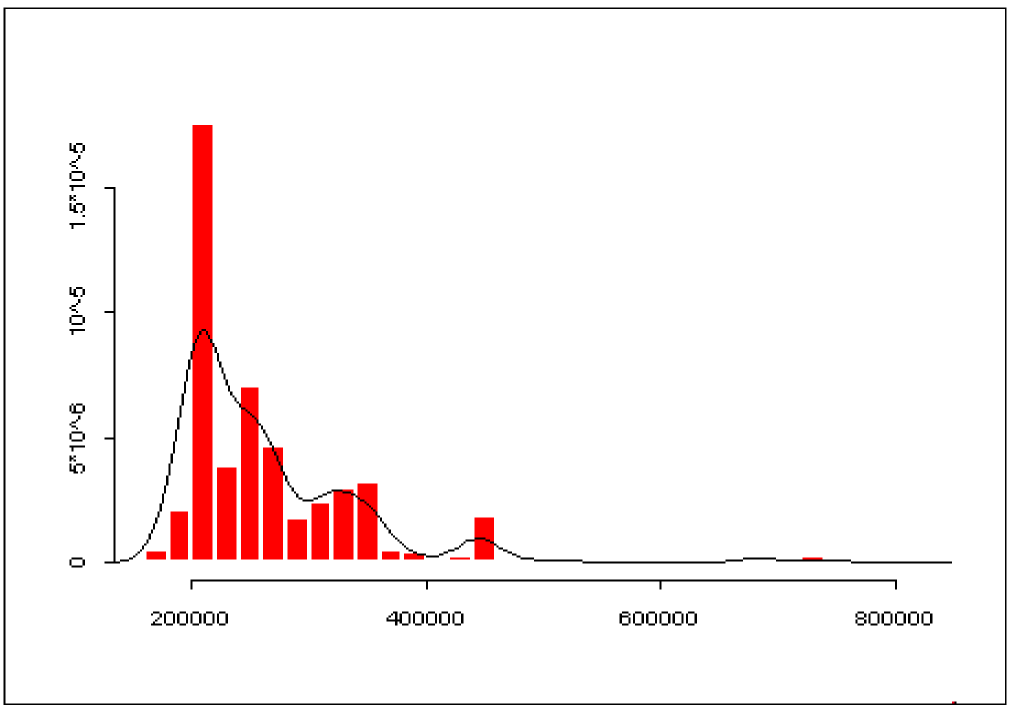

Fig 5.1 Histogram and Density Estimate of the Mode's Sampling Distribution for Example 1

Now, this sampling distribution estimate is used to calculate a 95% confidence interval for the mode. This interval is the confidence interval for the nodal hourly power consumption. The percentile method is utilized to infer the interval. The basic approach is very simple: the endpoints of a 95% confidence interval for the mode are the values of the mode at the 2.5th and 97.5th percentile of the bootstrap sampling distribution. To compute the interval, the 2000 values calculated of the mode must be arranged in increasing order. The 50th and 1950th values are the endpoints of the interval. They are respectively equal to 2.5%*2000 and 97.5%*2000. In this example, the 95% confidence interval is found to be **[167,315; 445,313]**.

5.2-Example 2

In this example, data processed in Example 2, Paragraph 3.3.1, Chapter 3 is used. The demand point consists of one hundred and fifty customers: seventy five customers from the (*HR*) class, fifty from the (*NHR*) class and twenty five from the (*CL*) class. The chosen hour is the forth of May 1. One hundred nodal hourly power consumption values were calculated as explained in Chapter 3. They form the basic set of measurements. The 95% confidence interval is found to be **[163,719; 897,631]**.

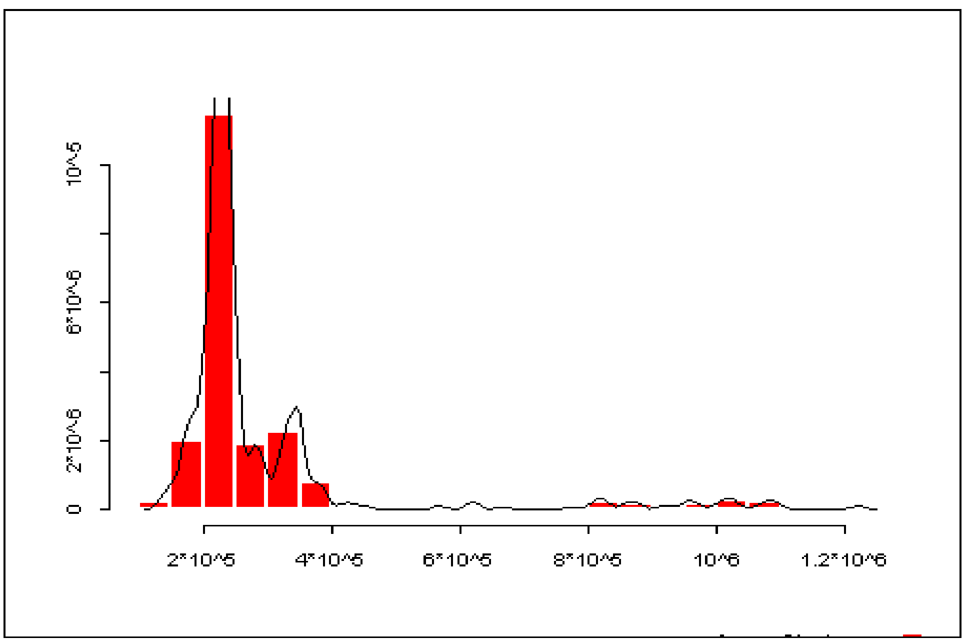

Fig 5.2 Histogram and Density Estimate of the Mode's Sampling Distribution for Example 2

5.3-Example 3

For this last example, the sample created in Example 3, Paragraph 3.3.1, Chapter 3 constitutes the original basic sample. Let us recall that the composition of the load is a little bit different in that case. It consists of thirty customers from the (*HR*) class and thirty from the (*NHR*) class. No customer from the (*CL*) class was considered. The hour of the year chosen is the 7^{th} of January 25. The density estimate of the mode's sampling distribution is given in Figure 5.3.

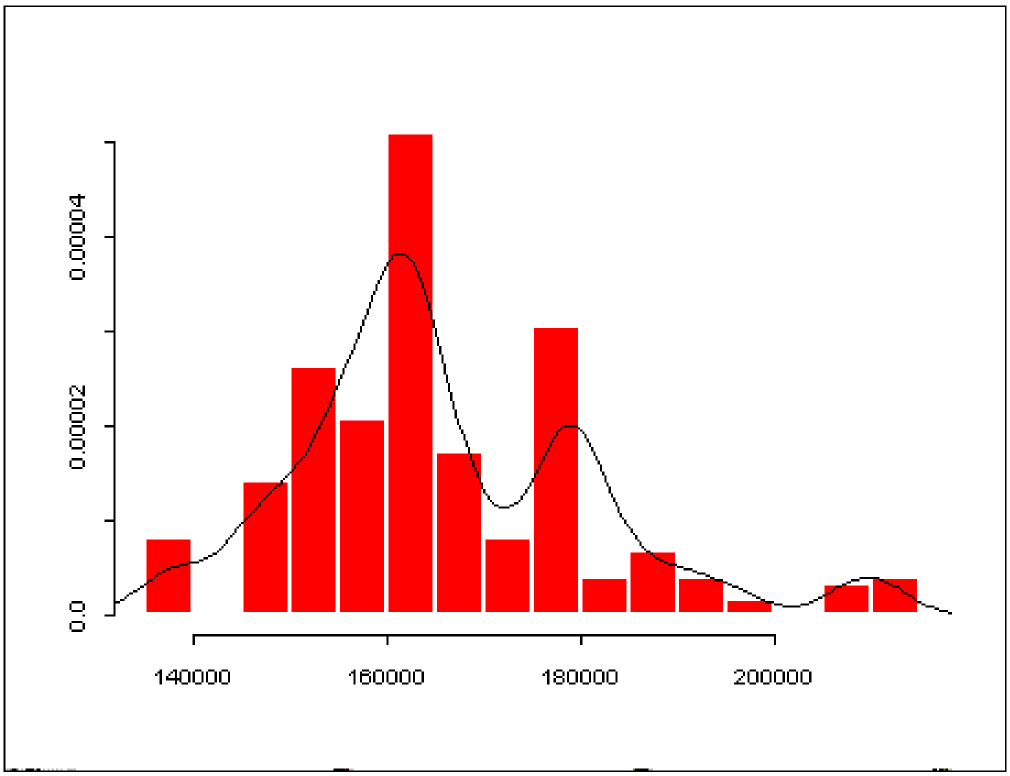

Fig 5.3 Histogram and Density Estimate of the Mode's Sampling Distribution for Example 3

By employing the program presented in Appendix F, the 95% confidence interval for the nodal hourly power consumption is found to be **[137,824 ; 209,088]**.

5.4-Conclusion

Three examples based on different initial conditions were presented in this chapter. The characteristics of each basic samples were described in Chapter 3. As shown in the histograms above, the bootstrap distributions of the mode are far from any parametric shape, especially from a Gaussian distribution. Furthermore, they are very different from one case to another. In the two first examples, the distributions are discrete, very asymmetric with long tails. The confidence interval associated with each one is very large. In the third example, the confidence interval is much smaller. The distribution, however, stays very asymmetric. Thus, the use of a nonparametric method to infer confidence intervals was justified in any case.

From the results obtained above, we can conclude that the calculation of confidence intervals for the nodal hourly power consumption is closely related to the composition of the load considered. For each different composition, the distribution of the mode is different and the confidence interval associated is more or less large. In fact, the more customers from the *C*ommercial *L*oad class are added, the larger the range of the 95% confidence interval will be. Let us recall that a 95% confidence level was chosen in order to find very accurate results. An interval less accurate, that is, with a lower confidence level, may permit to infer intervals with a variability not as large. For instance, in the second case, if this level is only decreased by 10%, the 85% confidence interval that is obtained is **[182,553; 393,557]**. It is much smaller than the 95% confidence interval which is equal to **[163,719; 897,631]**.