SMM069: Modelling and Data Analysis

Group coursework No. 1: Applications in EXCEL

There are two parts in this coursework. The first involves some data you are asked to analyse, the second requires simulation. Your submission should be **a single `EXCEL` file** that contains the numerical calculations, and **a single `PDF` file** that contains a report with fully worked solutions to each part, including graphics. The `EXCEL` file must be self contained, not linked to external files and contain sufficient comments that it is readily comprehensible. Marks will be awarded on the basis of the content of both files.

NOTE: the `EXCEL` work must not use VBA or any external packages.

Part 1: Parameter Estimation [25 Marks]

The data set given in '`SMM069_T3_2018-19-CW1-data.csv`' file values represent a sample of 1,000 claims sizes by independent policyholders. It is assumed that the claims sizes are Pareto distributed, with two parameters $\alpha > 0$ and $\lambda > 0$, and density function given by:

$$f(x) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}, \text{ with } x \geq 0$$

1.1. Estimate the parameters of the distribution by the method of moments.

1.2. Estimate the parameters of the distribution by the method of maximum likelihood using the SOLVER ADD-IN.

1.3. Using the estimates from part 1.2, produce both a QQ-plot and a probability plot to assess the assumption of the Pareto distribution for the claim sizes[1]. Comment on the result.

1.4. By taking (natural) logarithms of the data, produce both a QQ-plot and a probability plot for the (standard) Normal distribution. From the plot(s) discuss whether a LogNormal distribution would be a good fit for the data.

1.5. Interpret the plots from part 1.3 and 1.4. Which of the two models (i.e. Pareto or Lognormal) seems to be more appropriate for the claim sizes data?

[1] The construction of these diagnostic plots in `EXCEL` should be quite similar to those we saw in week 1, but using the inverse of the Pareto *c.d.f.* instead of `NORM.INV`. Note that there isn't a function in `EXCEL` giving the inverse of the Pareto *c.d.f.* so you have to invert it by hand and make use of the resulting formula.

Part 2: Monte Carlo simulation [25 Marks]

Consider the following hypothetical situation:

An insurer estimates that individual losses (X) to an insured follow a Pareto distribution with mean £2,500 and standard deviation £10,000.[2] The insurer pays 80% of individual losses in excess of £2,000 with a maximum payment of £100,000. It re-insures that portion of any payments in excess of £12,000.

2.1. Simulate a sample of 5,000 losses with no insurance (X) using the inverse transform method. Using the simulated sample, estimate the probability $(\hat{p})$ of individual losses (with no insurance) above £10,000. Check your result calculating the actual value of p from the theoretical distribution. Then calculate a 90% confidence interval for p based on your sample.

2.2. Without performing additional simulations, find how many losses would you need to simulate to be 90% confident that the estimate for p in part 2.1 is within $\pm 3\%$ of the actual value.[3]

2.3. The individual payments by the insurer (Y) to claimant, before recovery of reinsurance, are given by:

$$Y = \begin{cases} 0, & X \leq 2{,}000 \\ 0.80(X - 2{,}000), & 2{,}000 < X \leq 127{,}000 \\ 100{,}000, & X > 127{,}000 \end{cases}$$

Use the 5,000 simulations from part 2.1 to simulate the payments Y, and calculate a 95% confidence interval for the expected value, $E[Y]$.

2.4. Without performing additional simulations, find how many losses would you need to simulate to reduce the 95% confidence interval to a quarter of the range estimated in part 2.3.

2.5. The net costs to the insurer (Z) after recovery of reinsurance payments are given by:

$$Z = \begin{cases} 0, & X \leq 2{,}000 \\ 0.80(X - 2{,}000), & 2{,}000 < X \leq 17{,}000 \\ 12{,}000, & X > 17{,}000 \end{cases}$$

Use the 5,000 simulations from part 2.1 to simulate the net costs Z and plot its estimated distribution function. Further, find the median and the third quartile of the net cost Z.

2.6. Consider now that the individual losses (X) follow a logNormal distribution instead of a Pareto, but with the same mean and standard deviation as above. Repeat parts 2.1 to 2.5 using the logNormal. Discuss your results compared to those for the Pareto distribution.

[2] Use the same definition of the Pareto given in Part 1.

[3] The actual value of p is the probability of X above £10,000 derived from the theoretical distribution.