Statistics 321, Statistical Computing, Home Work 3

Posted: Sept. 3, 1993 Due: Sept. 12, 1993

The main idea of this exercise is a simulation investigation into "heavy tailed" distributions (recall this is where the CLT and even Laws of Large Numbers can break down). A convenient parametric family for this is the Pareto. Given $\alpha > 1$, this distribution has density:

$$f_\alpha(x) = (\alpha - 1)x^{-\alpha} 1_{(1,\infty)}(x).$$

1. Write a program which generates a sample of $n = 1000$ Pareto(α) random variables, for the values

   $$\alpha = 1.5, \ 2, \ 2.5, \ 3, \ 4, \ 10.$$

   Check that your algorithm is correct by both theoretically and empirically calculating

   $$P[X \le x], \qquad \text{for} \quad x = 1.1, \ 1.5, \ 2, \ 3, \ 5.$$

   Present the answer by printing tables, one for each x, that show, for each α, the theoretical value, and the left and right endpoints of 95% C.I.'s based on the empirical value (just use the simplest CLT based intervals, and the sample standard deviation for the scale). Explain why it is "not very nice" (but possible) that some of these intervals extend outside $[0,1]$.

ONLY TURN IN THE OUTPUT AND NOT THE PROGRAM LISTING

2. Investigate when the Law of Large Numbers holds (and does not hold!) here, by constructing a table whose entries are sample averages, for one sample each of size

   $$n = 10, \ 100, \ 1000, \ 10000,$$

   for each of the above choices of α. Add comments for when convergence in n seems to occur, and relate to sufficient conditions from the LLN (note: organize your tables carefully so convergence in n is EASY to see). In addition construct a table of sample standard deviations and comment on this.

3. Calculate the theoretical means and standard deviations for the above choices of α, and print the numerical values in a table (Caution: think carefully about which ones exist, and which don't). Comment on how these can be used to understand the behavior observed in part 2.

4. Investigate the effect of these calculations on standard normal theory C.I.'s, by generating 500 data sets of i.i.d. Pareto(α) r. v. 's, each with sample size $n = 100$, for each of the above α's for which the mean exists. For each data set calculate the left and right endpoints of the simple Normal theory 95% Confidence Intervals for the mean of the distribution. Print a table which shows the empirical coverage probability of these C.I.'s (i.e. the proportion of times the C.I. covers the true mean). Investigate whether or not these coverage probabilities are significantly different from 95% by computing Binomial theory standard errors for them.