ACST 3061 ACTUARIAL STATISTICS

Assignment 2 – Semester I, 2022

Total Marks Available: 10

The Norwegian fire insurance dataset `norauto` (provided in `norauto.csv`) is considered. The dataset contains 183,999 observations of automobile insurance policies losses in one year. It was obtained from the R package "CASdatasets" `http://cas.uqam.ca/pub/web/CASdatasets-manual.pdf`

Using Bayesian statistics and Bayesian credibility theory, we first analyze the data for the numbers of claims, the variable `NbClaim` in the dataset, while take into account the exposure variable `Expo` (and ignoring all other variables). Specifically, we first consider a parametric Bayesian model under which it is supposed that the numbers of claims N_i from the ith policy is from a Poisson distribution, $Poisson(V_i\lambda)$ with an unknown rate parameter λ. N_i corresponds to variable `NbClaim` and V_i corresponds to variable `Expo`. It is assumed that, conditional on λ, N_i, $i = 1, 2, ..., n$ are independent; $n = 183,999$. The unknown rate parameter λ is assumed to be a random variable which follows a Gamma distribution. It is given that the prior mean and variance of λ are 0.03 and 0.05, respectively. Answer Questions 1-3 as follows:

(1) Derive the formula to calculate the mean and standard deviation of the Bayesian posterior density for the unknown rate parameter λ and write R code to evaluate these for the given dataset and in the case if the dataset contains only first 100 policies from `norauto`. Compare the results.

(2) Plot the Bayesian posterior density for λ using R for the full dataset and in the case if the dataset contains only first 100 policies from `norauto`. Also calculate corresponding 90% Bayesian confidence intervals for λ. Compare the results.

(3) Derive formula for the credibility estimator of λ, i.e. formula representing credibility estimator as the weighted sum of the prior estimator for λ and the maximum likelihood estimator of λ (formula for the corresponding credibility factor should be derived too). Using R, evaluate the credibility estimator of λ and credibility factor for the given dataset and in the case if the dataset contains only first 100 policies from `norauto`. Compare the results.

Now, we analyze the data for the amounts/severities of claims, which are contained in the variable `ClaimAmount` of the dataset (i.e. policies with no claims can be ignored in this analysis). Specifically, we consider another parametric Bayesian model under which it is supposed that the log-amounts of claims are independent and identically distributed random variables from a Normal distribution with an unknown mean Θ and a known standard deviation $\sigma = 1.2$ (if policy has more than one claim then `ClaimAmount` represents the average claim amount and you can assume the same severity for each claim in this policy). The unknown mean parameter Θ is assumed to be a random variable which follows another Normal distribution. It is given that the prior mean and standard deviation of Θ are 6.0 and 4.0, respectively. Answer Questions 4-5 as follows:

(4) Use R to compute the posterior mean and standard deviation of the unknown mean parameter Θ. Calculate these in the case of the full dataset and in the case if the dataset contains only first 100 policies from `norauto`. Compare the results.

(5) Use R to plot the posterior and to compute a 90% Bayesian confidence interval for Θ, for the given dataset and in the case if the dataset contains only first 100 policies from `norauto`. Compare the results.

(6) Derive formula for the credibility estimator of Θ, i.e. formula representing credibility estimator as the weighted sum of the prior estimator for Θ and the maximum likelihood estimator of Θ (formula for the corresponding credibility factor should be derived too). Using R, evaluate the credibility estimator of Θ and credibility factor for the given dataset and in the case if the dataset contains only first 100 policies from `norauto`. Compare the results.

Important remarks:

- Answer **ALL** questions (**Questions 1-6**) by providing a typed report in PDF.
- All students must submit an assignment of their individual **o**wn work.
- For the PDF report, please describe and demonstrate your working steps and thought process as clearly as possible apart from showing important numerical answers/tables/graphs.
- Solution should be delivered as `PDF` report summarizing your modelling, reasoning and results. This file will be used for marking. It should be accompanied by `html` file obtained by knitting `R`-notebook `RMD` file used to calculate results presented in `PDF`. This html file be used to check the code if there are some concerns/doubts about the results/solutions presented in pdf. The submission files should be labelled as follows:
  - StudentIDLastNameFirstNameAssignment1.pdf
  - StudentIDLastNameFirstNameAssignment1.html
- **Page limit: please keep pdf file within 10 pages and a size of 10 MB**.
- **Due date:** 11:59pm, 27 May 2022

∼ End ∼