STSB6816 Special Test 2024

Mathematical Statistics and Actuarial Science; University of the Free State

2024/06/06

## Time: 170 minutes; Marks: 45

# Instructions

* Answer all questions in a single R Markdown document. Please knit to PDF or Word at the end and submit both the PDF/Word document and the “.Rmd” file for assessment, in that order.
* Label questions clearly, as it is done on this question paper.
* All results accurate to about 3 decimal places.
* Show all derivations, formulas, code, sources, and reasoning.
* Intervals should cover 95% probability unless stated otherwise.
* No communication software, devices, or communication capable websites may be accessed prior to submission. You may not (nor even appear to) attempt to communicate or pass information to another student.
* **Use of AI tools must be disclosed and summarised.**

# Introduction

The data is 0, 0, 1, 0, 1, 3, 0, 2, 1, 2. It is a time series of annual claim numbers for a small insurer.

They wish to model the data as a count process with different underlying forms. You are requested to fit Negative Binomial (Type II) regressions (with log link functions).

# Question 1

**1.1)** You should fit the following forms and in each case plot the resulting density of the time dependence coefficient $\left(β\right)$:

1. a linear regression on time $\left(t\_{i}=i=1…10\right)$: $η\_{i}=logE\left(Y\_{i}\right)=logμ\_{i}=α+βt\_{i}, i=1…10$
2. an AR(1) process, where $η\_{i}=α+βy\_{i-1}, i=1…10$
3. an underlying moving average of form MA(1): $η\_{i}=α+β η\_{i-1}, i=1…10$

They suggest a $U\left(0,0.9\right)$ prior on the time dependence coefficients. Additionally, you are requested to use the prior $ϕ \~Exp\left(0.01\right)$ for any scale or precision parameters. For the AR(1) form, assume that $y\_{0}=0$. Lastly, for the MA(1) model only, let $logμ\_{1}=η\_{1}∼N\left(-3,1\right)$. **[26]**

*[Hint: fit the first two forms to start, as that is sufficient to answer the rest of the questions (also, they can be fit with a single model specification). The MA(1) is more challenging and should only be attempted once the rest of the test is completed.]*

**1.2)** Compare the forms to determine which best fits the data, then explain why it was necessary to drop Observation 1 in the first form in order to be able to compare them. **[7]**

**For the AR(1) model:**

**1.3)** Illustrate the posterior predictive distributions for $t\_{11}$ and $t\_{12}$ taking the full parameter uncertainty into account (say using a bar plot). What is the probability that the company fails at Time 11 or Time 12 given that more than 5 claims in a month will cause them to fail? **[9]**

**1.4)** Suppose that for the last observation (Year 10) there may be claims incurred but not reported (IBNR), so the number is at least 2 (censored). Explain, in detail, everything you would change in order to incorporate this into your model. **[3]**

*[NB: Do not actually fit the model with censoring, the simulations will slow down too much for a test.]*