STOR 665 CWE 2022

Student Name:

Theory Part

1. GLM basics.

(a) Use the two Bartlett equations to derive the expected value and variance function of a random variable in the exponential family as a function of the cumulant generating function $b(\cdot)$.

(b) Derive the variance function for the Gamma distribution as a function of μ when the second parameter ν is fixed. The Gamma(μ, ν) density is given by

$$f_{\mu,\nu}(y) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu y}{\mu}\right)^{\nu} \exp\left(-\frac{\nu y}{\mu}\right) \frac{1}{y}, \quad y \geq 0.$$

2. Assume that in a randomized clinical trial patients with epilepsy are given either drug A or drug B. Here B is a new drug and the goal is to determine whether it is better in reducing the number of seizures compared to an established drug A. We count the number of epileptic seizures in a six month period while patients are taking either drug A or drug B. Age and time since onset of the disease are also recorded as additional predictors.

(a) Provide all components of a suitable GLM, where it is decided to use a Poisson regression with canonical link.

(b) Write the null hypothesis and alternative of interest to address the main question of the study. Which test would you apply (no details needed)?

(c) Write the information matrix on which the approximate inference for the model is based; it suffices to specify a general element of the matrix.

(d) How is approximate inference obtained using this information matrix?

(e) Further checks revealed that in the recorded number of seizures there is a disproportionate number of six seizures, which was found out to be due to sloppy recording where the actual count of seizures was not recorded and then later on 6 was substituted based on the expectation of one seizure per month. It is not possible to identify the patients for whom this happened. How can you address this in terms of modeling? Provide details and the components of a suitably modified model.