BIOSTAT/STAT 570: Coursework 3

To be submitted to the course canvas site by 11:59pm Monday 25th October, 2021.

1. Consider the Poisson-gamma random effects model given by

$$\begin{aligned} Y_i \mid \mu_i, \theta_i &\sim \text{Poisson}(\mu_i \theta_i) \qquad (1) \\ \theta_i &\sim \text{Ga}(b, b), \qquad (2) \end{aligned}$$

   which leads to a negative binomial marginal model with the variance a quadratic function of the mean. Design a simulation study, along the lines of that which produced Table 2.3 in the book (overdispersed Poisson example) to investigate the efficiency and robustness under:

   - A Poisson model.
   - A negative binomial model[1].
   - Quasi-likelihood with $\text{E}[Y] = \mu$, $\text{var}(Y) = \alpha\mu$.
   - Sandwich estimation.

   Use a loglinear model

$$\log \mu_i = \beta_0 + \beta_1 x_i,$$

   with $x_i \sim_{iid} \text{N}(0, 1)$, for $i = 1, \ldots, n$, and $\beta_0 = 0$ and $\beta_1 = \log 2$. You should carry out the simulation for:

   - $b = 0.1, 1, 10, 1000.$
   - $n = 10, 20, 50, 100, 250.$

   Summarize what your take-away message is, after carrying out these simulations.

2. The data in Table 1 contain data on a typical reliability experiment and give the failure stresses (in GPa) of four samples of carbon fibers of lengths 1, 10, 20 and 50mm.

Length (mm)	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2.247	2.640	2.842	2.908	3.099	3.126	3.245	3.328	3.355	3.383	3.572	3.581	3.681
10	1.901	2.132	2.203	2.228	2.257	2.35	2.361	2.396	2.397	2.445	2.454	2.454	2.474
20	1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958	1.966	1.997	2.006
50	1.339	1.434	1.549	1.574	1.589	1.613	1.746	1.753	1.764	1.807	1.812	1.840	1.852

Table 1: Failure stress data for four groups of fibers.

[1] The `glm.nb` function in the `MASS` library can be used to fit this model

(a) The exponential distribution $Y \mid \lambda \sim_{iid}$ Exponential(λ), is a simple model for reliability data:

$$p(y|\lambda) = \lambda \exp(-\lambda y),$$

with $\lambda, y > 0$. The hazard function is the probability of imminent failure and is given by

$$h(y|\lambda) = \frac{p(y|\lambda)}{S(y|\lambda)},$$

where $S(y|\lambda) = \Pr(Y > y|\lambda)$ is the probability of failure beyond y. Derive the hazard function for the exponential distribution. Suppose we have a sample $y_1, \ldots, y_n$, of size n from an exponential distribution. Find the form of the MLE of λ and the asymptotic variance.

(b) For each of the four groups in Table 1 estimate a separate λ, with an associated standard error. Examine the appropriateness of the exponential model via QQ plots.

(c) Consider a quasi-likelihood approach to inference for λ under the model with

$$\mathsf{E}[Y|\lambda] = \lambda^{-1}, \qquad \mathsf{var}(Y|\lambda) = \alpha\lambda^{-2}$$

with $\alpha > 0$. Suggest an estimator for α. Estimate λ, α and the standard errors, separately for each of the four groups in Table 1. What do the results suggest to you about the fit of the exponential model?

(d) Obtain the form of the sandwich estimate for the variance of $\widehat{\lambda}$. Numerically evaluate sandwich standard errors for the estimate of λ in each of the four groups.

(e) The Weibull distribution is a common model for survival or reliability data: $Y \mid \eta, \alpha \sim_{iid}$ Weibull(η, α), with $\eta > 0$, and $\alpha > 0$. The random variable Y has a Weibull distribution if its density can be written in the form

$$p(y|\eta, \alpha) = \eta\alpha^{-\eta}y^{\eta-1} \exp\left[-\left(\frac{y}{\alpha}\right)^{\eta}\right].$$

Find the mean, variance and hazard function of a Weibull distribution. For what value of the parameters does the exponential distribution result?

(f) Is the Weibull distribution with unknown parameters η, α a member of the exponential family? What are the implications for inference?

(g) For the Weibull model and a random sample of size n obtain: the log-likelihood, the score and the observed information matrix.

(h) Solve the score equations in order to obtain the maximum likelihood estimators (MLEs). You should obtain a single equation that needs to be numerically solved.

(i) Obtain the MLEs and standard errors for the parameters of the Weibull model, for each of the groups in Table 1.