Assignment 2

2017-05-14

Question 1:

Consider data on Foot-and-Mouth disease (FMDdata.xls) with the foal of assessing the difference in incidence between the eastern and western regions of Turkey, based on data from 1998:

(a) Apply classical Poisson regression to the FMD counts, with region and size of cattle population (standardised) as covariates. Check model diagnostics. [10 marks]

Firstly let's import the data

Table 1: Raw FMD dataset

Province	EasternTurkey	FMD1998	Cattle
Balikesir	0	4	258450
Bilecik	0	1	39330
Canakkale	0	0	118880
Edirne	0	0	116210
Eskisehir	0	0	78130
Istanbul	0	0	90040

Let's remove the unnecessary variables (**Province**) and standardise the **Cattle** variable

Table 2: FMD dataset with standardised Cattle values

EasternTurkey	FMD1998	StCattle
0	4	0.8543942
0	1	-1.1001966
0	0	-0.3905959
0	0	-0.4144128
0	0	-0.7540935
0	0	-0.6478541

Next, we fit a basic GLM using the poisson family option in R: *fm_pois <- glm(FMD1998 ~ .,data = FMDData2, family = poisson)*

Table 3: Model Output

term	estimate	std.error	statistic	p.value
(Intercept)	0.0307282	0.1537837	0.1998142	0.8416259
EasternTurkey	-0.1313546	0.2518113	-0.5216391	0.6019217
StCattle	0.4566755	0.0832757	5.4838978	0.0000000

Table 4: Model Summary

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual
186.0122	65	-112.5923	231.1846	237.7535	160.228	63

It seems as if the only significant covariate is the (standardised) **Cattle** variable; the **EasternTurkey** region variable does not return a signifcant ($\alpha \leq 0.05$) p value.

Just to check that we aren't missing anything, let's also look at the model summaries for the model including an interaction term **StCattle x EasternTurkey** and the model with only **StCattle** as a covariate.

Table 5: Model (with Interaction term) Summary

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual
186.0122	65	-109.7224	227.4447	236.2034	154.4882	62

Table 6: Model (only StCattle) Summary

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual
186.0122	65	-112.7297	229.4593	233.8386	160.5028	64

The original model and the model with only the **StCattle** covariate are – unsurprisingly – the best performing models, given that we found the **EasternTurkey** covariate to be insignificant at the beginnning. Whilst the origial model proves the superior one when comparing **deviance, AIC, BIC** and **logLik** scores, the **StCattle** model does have an extra degree of freedom for the residuals, which might be desirable.

(b) Repeat (a), using a Bayesian approach, and compare results. [10 marks]

```
## Loading required package: survival

##
## Attaching package: 'Zelig'

## The following object is masked from 'package:purrr':
##
##     reduce
```

(c) Use a Bayesian zero-inflated Poisson (ZIP) model instead. [10 marks]

Density function:

$$f(y \mid \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} E(Y) = Var(Y) = \lambda$$

where λ is the mean.

Likelihood:

$$\mathcal{L}(\lambda \mid y) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$
$$= \frac{\lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}}{y_1! y_2! \cdots y_n!}$$

Log-likelihood:

$$\begin{aligned}
\mathcal{LL}(\lambda \mid y) &= \sum_{i=1}^{n} log(\lambda^{y_i} e^{-\lambda_i}) - log(y_i!) \\
\mathcal{LL}(\lambda \mid y) &= \sum_{i=1}^{n} log(\lambda^{y_i}) + log(e^{-\lambda_i}) - log(y_i!) \\
\mathcal{LL}(\lambda \mid y) &= \sum_{i=1}^{n} y_i log(\lambda_i) - \lambda_i - log(y_i!)
\end{aligned}$$

(d) Compare the above two Bayesian models and, for your model of choice, report on any difference in mean FMD occurrence for eastern vs western Turkey. [10 marks]

Question 2:

Consider the AR(1) model (see Phillips, 1991 and Berger & Yang, 1994):

Assuming that $|\rho| \leq 1$, compare posteriors (and intervals) arising from the uniform prior and the reference prior for some generated data sets, based on *e.g* $\rho = 0.1$ and $\rho = 0.99$, and sample sizes $n = 50$ and $n = 200$. Also compare with confidence intervals.

The Gaussian likelihood is given by Phillips as

$$f(y|\rho,\sigma,y_0) \ = (2\pi)^{-T/2}\sigma^{-T}exp\left\{-(1/2)\sigma^{-2}\textstyle\sum_1^T(y_t-\rho y_{t-1})\right\},$$

which – assuming a flat prior for $(\rho, log(\sigma))$ – leads to the uninformative prior for $(\rho,\sigma) : \pi(\rho,\sigma) \propto 1/\sigma$.

Reference Prior

Berger & Yang (1994) derive the reference prior – see equation **(18)** – strictly for AR(1) models with $|\rho| \leq 1$ (as $T \to \infty$) as:

$$\pi_{_{Ref}}(\rho) = (1-\rho^2)^{-1/2}$$

References

Berger, J., & Yang, R. (1991). Noninformative Priors and Bayesian Testing for the AR (1) Model. Econometric Theory, 10(3/4), 461–482. Retrieved from http://www.jstor.org/stable/3532546

Phillips, P. (1991). To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends. Journal of Applied Econometrics, 6(4), 333-364. Retrieved from http://www.jstor.org/stable/2096686