STA 6352, Report 8.7

Carson Slater *Baylor University*

Problem: Processionary Caterpillars

The pine processionary caterpillar (*Thaumetopoea pityocampa*), so named because of their head-to-tail processions when moving from an old nest to establish a new one, have larvae that are a major pest in pine forests. Photos of these caterpillars appear below.[1]

Data from a 1973 study of processionary caterpillars was published and analyzed in Tomassone et al. (1993).[2]

The response variable is the log of the average number of caterpillar nests per tree in a 500m^2 area. The authors considered $k = 10$ potential explanatory variables, for $n = 33$ areas as listed below:

- x_1: altitude
- x_2: slope (in degrees)
- x_3: number of pines in the area
- x_4: height of the central tree
- x_5: diameter of the central tree
- x_6: index of the settlement density
- x_7: orientation of the area (from 1 [southbound] to 2)
- x_8: height of the dominant tree
- x_9: number of vegetation strata
- x_{10}: mix settlement index (from 1 if not mixed to 2 if mixed)

The goal is to choose a subset of these variables as predictors. Variable selection is a classic—perhaps *the* classic—model choice problem.

[1] Adapted from Marin and Robert, Chapter 3. Online resources available at: `https://www.ceremade.dauphine.fr/~xian/BCS/`
[2] Tomassone, R., Dervin, C., and Masson, J. (1993). *Biométrie: Modélisation de Phénomènes Biologiques*. Masson, Paris.

Here is the traditional (non-Bayesian analysis):

Characteristic	**Beta**	**95% CI**[1]	**p-value**
x1	0.00	0.00, 0.00	0.015
x2	-0.04	-0.07, 0.00	0.025
x3	0.04	-0.01, 0.10	0.11
x4	-0.03	-0.42, 0.37	0.9
x5	-0.63	-1.8, 0.55	0.3
x6	0.01	-0.29, 0.31	>0.9
x7	-1.1	-2.2, -0.06	0.040
x8	-0.23	-1.1, 0.65	0.6

[1]CI = Confidence Interval

Now we apply the conjugate prior analysis, but assuming little or no prior information about $\tilde{\boldsymbol{\beta}}$, $\mathbf{M}$, a, and b. For the sake of illustration, take $a = 2.1$ and $b = 2$, yielding a prior mean and variance of 33.06 and 1.82, respectively, for σ^2. Let $\tilde{\boldsymbol{\beta}} = \mathbf{0}_{k+1}$. There remains the problem of choosing $\mathbf{M}$. Suppose we take $\mathbf{M} = \frac{1}{g}\mathbf{I}_{k+1}$. Here are the results when we take $g = 1000$:

	Parameter	Posterior_Mean	Posterior_Variance
beta[1]	x1	6.9001	3.2429
beta[2]	x2	-0.0024	0.0000
beta[3]	x3	-0.0330	0.0002
beta[4]	x4	0.0059	0.0033
beta[5]	x5	-0.4694	0.1194
beta[6]	x6	0.1063	0.0041
beta[7]	x7	0.2710	0.7700
beta[8]	x8	-0.1817	0.4264
beta[9]	x9	0.0206	0.0220
beta[10]	x10	-0.8238	0.2982
beta[11]	x11	-0.1540	0.2022

	Parameter	Lower_HPD	Upper_HPD	Contains_Zero
beta[1]	x1	3.2786	10.8210	FALSE
beta[2]	x2	-0.0047	-0.0004	FALSE
beta[3]	x3	-0.0604	-0.0055	FALSE
beta[4]	x4	-0.1044	0.1331	TRUE
beta[5]	x5	-1.1139	0.0573	TRUE
beta[6]	x6	0.0178	0.2476	FALSE
beta[7]	x7	-1.7724	1.5128	TRUE
beta[8]	x8	-1.5882	1.0310	TRUE
beta[9]	x9	-0.2017	0.3092	TRUE
beta[10]	x10	-1.6726	0.2560	TRUE

	Parameter	Lower_HPD	Upper_HPD	Contains_Zero
beta[11]	x11	-1.0552	0.7197	TRUE