Statistical Computing for Scientists and Engineers

Homework 3

Jiale Shi

September/21/2018

1 Problem 1

The problem of interest consists of a scalar output y (the logarithmic of the number of caterpillars colonies in a $500m^2$ area) and $k = 10$ explanatory variables $x_i (i = 1, ..., 10)$ which correlate to features such as altitude, landscape slope, vegetation count, etc. The exact meaning does not matter in this context. Using the caterpillar.mat data, plot the semi-log-y plots of the caterpillar colony (last column of the data file) versus each feature reproducing Figure 3.1 in Bayesian Core.

Answer: we run the code that provided to us. We get the semi-log-y plots are shown below.

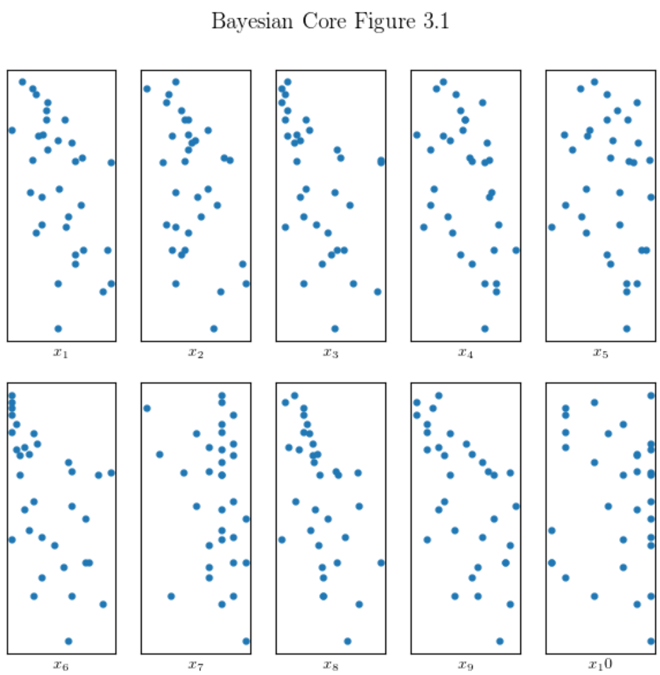

Figure 1: semi-log-y plots

2 Problem 2

From the book, 3.1.1 linear model

The ordinary normal linear regression model is such that

$$y|\beta, \sigma^2, X \sim \mathcal{N}_n(X\beta, \sigma^2 I_n) \tag{1}$$

the likelihood of the ordinary normal linear model is

$$l(\beta, \sigma^2|y, X) = (2\pi\sigma^2)^{n/2} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right] \tag{2}$$

The maximum likelihood estimator of β is then the solution of the (least squares) minimization problem.

$$\min_\beta (y - X\beta)^T(y - X\beta) = \min_\beta \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - ... - \beta k x_{ik})^2 \tag{3}$$

namely,

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{4}$$

which is also the orthogonal projection of y on the linear subspace spanned by the columns of X.

Similarly, an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-k-1}(y - X\hat{\beta})^T(y - X\hat{\beta}) = \frac{s^2}{n-k-1} \tag{5}$$

and $\hat{\sigma}^2(X^T X)^{-1}$ approximates the covariance matrix of $\hat{\beta}$. We can then define the standard t-statistic as

$$T_i = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 \omega_{(i,i)}}} \ \mathcal{J}_{n-k-1} = \tag{6}$$

where $\omega_{(i,i)}$ denotes the (i, i)th element of the matrix $(X^T X)^{-1}$. For this problem $\beta_i = 0$

the p value

$$p_i = P_{H_0}(|T_i| - |t_i|) < \alpha \tag{7}$$

Note that this statistic T_i can also be used to build on the β_is as a (marginal) frequentist confidence interval, of the form.

$$\beta_i; |\beta_i - \hat{\beta}_i| <= \sqrt{\omega_{ii}} F^{-1}_{n-k-1}(1 - \alpha/2) \tag{8}$$

Then we run the code that is offered to us. We reproduce Figure 3.2

Bayesian Core Figure 3.2

	Estimate	Std. Error	t-value	$Pr(>\|t\|)$
intercept	10.998412	3.060272	3.593933	0.001615
XV0	-0.004431	0.001557	-2.846347	0.009391
XV1	-0.053830	0.021900	-2.458008	0.022317
XV2	0.067939	0.099472	0.682998	0.501738
XV3	-1.293636	0.563811	-2.294452	0.031678
XV4	0.231637	0.104378	2.219207	0.037095
XV5	-0.356800	1.566464	-0.227774	0.821926
XV6	-0.237469	1.006006	-0.236051	0.815577
XV7	0.181060	0.236724	0.764858	0.452483
XV8	-1.285316	0.864847	-1.486177	0.151423
XV9	-0.433106	0.734869	-0.589364	0.561621

Figure 2: Dataset caterpillar: output providing the maximum likelihood estimates of the regression coefficients and their standard significance analysis.

3 Problem 3

We mainly focus on the section 3.2.1 Conjugate Priors.

We use the conjugate prior:

$$\begin{aligned} &\beta|\sigma^2, X \sim \mathcal{N}_{k+1}(\tilde{\beta}, \sigma^2 M^{-1}) \\ &\sigma^2|X \sim \mathcal{IG}(a, b) \end{aligned} \tag{9}$$

The prior is conjugate, leading us to the following normal-inverse-gamma posterior distribution.

$$\begin{aligned} &\beta|\sigma^2, y, X \sim \mathcal{N}_{k+1}((M + X^T X)^{-1} X^T X \tilde{\beta}, \sigma^2 [M^{-1} + (X^T X)^{-1}]^{-1}) \\ &\sigma^2|y, X \sim \mathcal{IG}(\frac{n}{2}, b + \frac{s^2}{2} + \frac{\hat{\beta}^T [M^{-1} + (X^T X)^{-1}]^{-1} \hat{\beta}}{2}) \end{aligned} \tag{10}$$

In this setting, the Bayes estimators of β and σ^2 associated with squared error losses, namely the posterior means of β and σ^2, can be computed in closed form.

In fact, a simple computation shows that they are given by

$$\begin{aligned}\mathbb{E}^{\pi}[\beta|y,X] &= \mathbb{E}^{\pi}[\mathbb{E}^{\pi}(\beta|\sigma^2,y,X)|y,X] \\ &= (M+X^TX)^{-1}(X^TX)\hat{\beta}+M\hat{\beta}\end{aligned} \tag{11}$$

and for $n \geq 2$

$$\mathbb{E}^{\pi}[\beta|y,X] = \frac{2b+s^2+(\tilde{\beta}-\hat{\beta})^T M^{-1}+{(X^TX)^{-1}}^{-1}(\tilde{\beta}-\hat{\beta})}{n+2a-2} \tag{12}$$

Integrating Eq.10 in σ^2 leads to a multivariate t marginal posterior distribution on β since.

$$\begin{aligned}\pi(\beta|y,X) \propto [(\beta - {M+X^TX}^{-1}[X^TX\hat{\beta}+M\tilde{\beta}])^T(M+X^TX) \\ \times (\beta - {M+X^TX}^{-1}[X^TX\hat{\beta}+M\tilde{\beta}])+2b+s^2 \\ +(\tilde{\beta}-\hat{\beta})^T(M^{-1}+(X^TX)^{-1})^{-1}(\tilde{\beta}-\hat{\beta})]^{-(n/2+k/2+a)}\end{aligned} \tag{13}$$

We thus have that, marginally and a posteriori,

$$\beta|y,X \sim \mathcal{J}_{k+1}(n+2a,\hat{\mu},\hat{\Sigma}) \tag{14}$$

with

$$\begin{aligned}\hat{\mu} &= (M+X^TX)^{-1}((X^TX)\hat{\beta}+M\tilde{\beta}) \\ \hat{\Sigma} &= \frac{2b+s^2+(\tilde{\beta}-\hat{\beta})^T(M^{-1}+(X^TX)^{-1})^{-1}(\tilde{\beta}-\hat{\beta})}{n+2a}(M+X^TX)^{-1}\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbb{V}^{\pi} &= \frac{n+2a}{n+2a-4}\hat{\Sigma} \\ &= \frac{2b+s^2+(\tilde{\beta}-\hat{\beta})^T(M^{-1}+(X^TX)^{-1})^{-1}(\tilde{\beta}-\hat{\beta})}{n+2a-4}(M+X^TX)^{-1}\end{aligned} \tag{16}$$

Bayesian Core Table 3.1

c	$E^\pi(\sigma^2 \mid Y, X)$	$E^\pi(\beta_0 \mid Y, X)$	$V^\pi(\beta_0 \mid Y, X)$
0.1000	0.5746	9.6626	6.8355
1.0000	0.5746	9.6626	6.8355
10.0000	0.5746	9.6626	6.8355
100.0000	0.5746	9.6626	6.8355
1000.0000	0.5746	9.6626	6.8355

Figure 3: Dataset caterpillar: Influence of the prior scale c on the Bayes estimates of σ^2 and β_0.

Bayesian Core Table 3.2

β_i	$E^\pi(\beta_i \mid Y, X)$	$V^\pi(\beta_i \mid Y, X)$
β_0	9.6626	6.8355
β_1	-0.0040	0.0000
β_2	-0.0516	0.0004
β_3	0.0418	0.0077
β_4	-1.2633	0.2615
β_5	0.2307	0.0090
β_6	-0.0832	1.9310
β_7	-0.1917	0.8254
β_8	0.1608	0.0462
β_9	-1.2069	0.6127
β_{10}	-0.2567	0.4267

Figure 4: Dataset caterpillar: Bayes estimates of β for $c = 100$.

4 Problem 4

The posterior distribution can be derived as

$$\begin{aligned} \pi(\beta, \sigma^2 \mid y, X) &\propto f(y \mid \beta, \sigma^2, X)\pi(\beta, \sigma^2 \mid X) \\ &\propto (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}(y - X\hat{\beta})^T(y - X\hat{\beta}) - \frac{1}{2\sigma^2}(\beta - \hat{\beta})^T(X^T X)(\beta - \hat{\beta})\right] \\ &\cdot (\sigma^2)^{-k/2} \exp\left[-\frac{1}{2c\sigma^2}(\beta^T X^T X\beta)\right] \end{aligned} \tag{17}$$

Conduct marginalization and derive the conditional and marginal posteriors on β and σ^2:

$$\begin{aligned}\beta|\sigma^2, y, X &\sim \mathcal{N}_{k+1}(\frac{c}{c+1}\hat{\beta}, \frac{\sigma^2 c}{c+1}(X^T X)^{-1}) \\ \sigma^2|y, X &\sim \mathcal{IG}(\frac{n}{2}, \frac{s^2}{2} + \frac{1}{2(c+1)}\hat{\beta}^T X^T X\hat{\beta})\end{aligned} \tag{18}$$

By integrating out σ^2 on the conditional posterior β, we can show that

$$\beta|y, X \sim \mathbb{T}_{k+1}(n, \frac{c}{c+1}\hat{\beta}, \frac{c(s^2 + \frac{\hat{\beta}^T X^T X\hat{\beta}}{c+1})}{n(c+1)}(X^T X)^{-1}) \tag{19}$$

$$\begin{aligned}\mathcal{E}[\beta|y, X] &= \frac{c}{c+1}\hat{\beta} \\ \mathcal{V}[\beta|y, X] &= \frac{c(s^2 + \frac{\hat{\beta}^T X^T X\hat{\beta}}{c+1})}{(n-2)(c+1)}(X^T X)^{-1}\end{aligned} \tag{20}$$

And the expression of Bayes factor has been given.

Bayesian Core Table 3.3, C = 100

β_i	$E^\pi(\beta_i\|Y, X)$	$V^\pi(\beta_i\|Y, X)$
β_0	10.8895	6.8229
β_1	-0.0044	0.0000
β_2	-0.0533	0.0003
β_3	0.0673	0.0072
β_4	-1.2808	0.2316
β_5	0.2293	0.0079
β_6	-0.3533	1.7877
β_7	-0.2351	0.7373
β_8	0.1793	0.0408
β_9	-1.2726	0.5449
β_{10}	-0.4288	0.3934

Figure 5: Dataset caterpillar: Posterior mean and variance of β for $c = 100$ using Zellar's G-prior.

The dependence of the Bayes factor on the pair (c, c_0) cannot be bypassed in the sense that the Bayes factor varies between 0 and ∞ when c_0/c goes from 0 to ∞.

Although these Bayes factors should not be used simultaneously, an informal conclusion is that the most important variables besides the intercept seem to be X_1, X_2, X_4 and X_5.

Bayesian Core Table 3.4, C = 1000

β_i	$E^\pi(\beta_i\|Y,X)$	$V^\pi(\beta_i\|Y,X)$
β_0	10.9874	6.6644
β_1	-0.0044	0.0000
β_2	-0.0538	0.0003
β_3	0.0679	0.0070
β_4	-1.2923	0.2262
β_5	0.2314	0.0078
β_6	-0.3564	1.7461
β_7	-0.2372	0.7202
β_8	0.1809	0.0399
β_9	-1.2840	0.5323
β_{10}	-0.4327	0.3843

Figure 6: Dataset caterpillar: Same legend as Table 3.3 for c=1000

5 Problem 5

The marginal posterior distribution from Problem 4:

$$\beta|y,X \sim \mathbb{T}_{k+1}(n, \frac{c}{c+1}\hat{\beta}, \frac{c(s^2+\frac{\hat{\beta}^T X^T X\hat{\beta}}{c+1})}{n(c+1)}(X^TX)^{-1}) \tag{21}$$

As the standard student-t is one symmetrical distribution, we can analytically calculate its 90% HPD region as the two points whose respective CDF is 5% and 95%. Denote them as δ_5 and δ_{95}. The corresponding HPD of β would than be

$$\begin{aligned}
\beta_5 &= \frac{c}{c+1}\hat{\beta} + \delta_5 \cdot \frac{c(s^2+\frac{\hat{\beta}^T X^T X\hat{\beta}}{c+1})}{n(c+1)}(X^TX)^{-1}) \\
\beta_{95} &= \frac{c}{c+1}\hat{\beta} + \delta_{95} \cdot \frac{c(s^2+\frac{\hat{\beta}^T X^T X\hat{\beta}}{c+1})}{n(c+1)}(X^TX)^{-1})
\end{aligned} \tag{22}$$

Bayesian Core Table 3.5

β_i	HPD Interval
β_0	[5.7435, 16.2533]
β_1	[-0.0071, -0.0018]
β_2	[-0.0914, -0.0162]
β_3	[-0.1029, 0.2387]
β_4	[-2.2618, -0.3255]
β_5	[0.0524, 0.4109]
β_6	[-3.0466, 2.3330]
β_7	[-1.9649, 1.4900]
β_8	[-0.2254, 0.5875]
β_9	[-2.7704, 0.1998]
β_{10}	[-1.6950, 0.8288]

Figure 7: Dataset caterpillar:95% HPD intervals for the components of β for $c = 100$

6 Problem 6

We now consider c an unknown hyper-parameter, that is we use the same G-prior distribution, and we now introduce a diffuse prior distribution on c.

$$\pi(c) = c^{-1}\mathcal{I}_{\mathcal{N}*}(c) \tag{23}$$

the corresponding marginal posterior on the parameters of interest is then

$$\pi(\beta, \sigma^2|y, X) \sum_{c=1}^{\infty} \pi(\beta, \sigma^2|y, X, c) f(y|X, c) c^{-1} \tag{24}$$

$$\begin{aligned} f(y|X) &\propto_c f(y|X, c)\pi(c) \\ &\sum_c c^{-1}(c+1)^{-\frac{k+1}{2}}[y^T y - \frac{c}{c+1} y^T X (X^T X)^{-1} X^T y]^{-n/2} \end{aligned} \tag{25}$$

For the posterior mean and variance of β, we integrate out c by summation:

$$\mathcal{E}[\beta|y, X] = \sum_c \mathcal{E}[\beta|y, X, c]\pi(c|y, X) \tag{26}$$

because $\pi(c|y, X) \propto f(y|X, c)\pi(c)$

$$\mathcal{E}[\beta|y,X] = \sum_c \frac{c}{c+1}\hat{\beta}\pi(c|y,X) = \hat{\beta}\sum_c [\frac{c}{c+1}\frac{f(y|X,c)\pi(c)}{\sum_c f(y|X,c)\pi(c)}] \tag{27}$$

$$\mathcal{V}[\beta|y,X] = \mathcal{E}[\mathcal{V}(\beta|y,X,c)|y,X] + \mathcal{V}(\mathcal{E}[\beta,X,c]|y,X) \tag{28}$$

Bayesian Core Table 3.6

β_i	$E^\pi(\beta_i\|Y,X)$	$V^\pi(\beta_i\|Y,X)$
β_0	9.2714	9.6424
β_1	-0.0037	0.0000
β_2	-0.0454	0.0005
β_3	0.0573	0.0092
β_4	-1.0905	0.3079
β_5	0.1953	0.0105
β_6	-0.3008	2.2750
β_7	-0.2002	0.9383
β_8	0.1526	0.0522
β_9	-1.0835	0.7063
β_{10}	-0.3651	0.5020

Figure 8: Table of Problem 6

An important point with this approach is that the marginal distribution of the dataset is available in closed form. It is therefore possible to produce a Bayes regression output, just as in the first-level prior. This is one additional reason why this non-informative prior is introduced.

7 Problem 7

The algorithm is stochastic search for the most likely model. When k is large, it becomes computationally intractable to compute the posterior probabilites of the 2^k models. Need of a tailored algorithm that samples from $p(\gamma|y, X)$ and selects the most likely models. This can be done by Gibbs sampling, given the availability of the full conditional posterior probabilities of the γ_i's. If the π with the that model parameter is much larger than that without that model parameter, that model parameter should be turned on. If the π with the that model parameter is much smaller than that without that model parameter, that model parameter should be turned off.

Bayesian Core Table 3.7, C = 100

$t_1(\gamma)$	$\pi(\gamma\|Y, X)$
0, 1, 2, 4, 5	0.23154
0, 1, 2, 4, 5, 9	0.03736
0, 1, 9	0.03444
0, 1, 2, 4, 5, 10	0.03297
0, 1, 4, 5	0.03061
0, 1, 2, 9	0.02502
0, 1, 2, 4, 5, 7	0.02414
0, 1, 2, 4, 5, 8	0.02378
0, 1, 2, 4, 5, 6	0.02374
0, 1, 2, 3, 4, 5	0.02321
0, 1, 6, 9	0.01459
0, 1, 2, 3, 9	0.01449
0, 9	0.01428
0, 1, 2, 6, 9	0.01355
0, 1, 4, 5, 9	0.01276
0, 1, 3, 9	0.01171
0, 1, 2, 8	0.01148
0, 1, 8	0.00952
0, 1, 2, 3, 4, 5, 9	0.00904
0, 1, 2, 4, 5, 6, 9	0.00903

Figure 9: Table of Problem 7(a)

Bayesian Core Table 3.8

$t_1(\gamma)$	$\pi(\gamma\|Y,X)$
0, 1, 2, 4, 5	0.09293
0, 1, 2, 4, 5, 9	0.03254
0, 1, 2, 4, 5, 10	0.02950
0, 1, 2, 4, 5, 7	0.02311
0, 1, 2, 4, 5, 8	0.02284
0, 1, 2, 4, 5, 6	0.02280
0, 1, 2, 3, 4, 5	0.02240
0, 1, 2, 3, 4, 5, 9	0.01673
0, 1, 2, 4, 5, 6, 9	0.01672
0, 1, 2, 4, 5, 8, 9	0.01372
0, 1, 4, 5	0.01104
0, 1, 2, 4, 5, 9, 10	0.00993
0, 1, 2, 3, 9	0.00970
0, 1, 2, 9	0.00932
0, 1, 2, 4, 5, 7, 9	0.00925
0, 1, 2, 6, 9	0.00919
0, 1, 4, 5, 9	0.00876
0, 1, 2, 3, 4, 5, 10	0.00793
0, 1, 2, 4, 5, 8, 10	0.00790
0, 1, 2, 4, 5, 7, 10	0.00789

Figure 10: Table of Problem 7(b)

Bayesian Core Table 3.9

$t_1(\gamma)$	$\pi(\gamma\|Y,X)$	$\hat{\pi}(\gamma\|Y,X)$
0, 1, 2, 4, 5	0.23154	0.23020
0, 1, 2, 4, 5, 9	0.03736	0.03755
0, 1, 9	0.03444	0.03444
0, 1, 2, 4, 5, 10	0.03297	0.03200
0, 1, 4, 5	0.03061	0.02922
0, 1, 2, 9	0.02502	0.02600
0, 1, 2, 4, 5, 7	0.02414	0.02366
0, 1, 2, 4, 5, 8	0.02378	0.02589
0, 1, 2, 4, 5, 6	0.02374	0.02633
0, 1, 2, 3, 4, 5	0.02321	0.02611
0, 1, 6, 9	0.01459	0.01489
0, 1, 2, 3, 9	0.01449	0.01200
0, 9	0.01428	0.01289
0, 1, 2, 6, 9	0.01355	0.01544
0, 1, 4, 5, 9	0.01276	0.01211
0, 1, 3, 9	0.01171	0.00967
0, 1, 2, 8	0.01148	0.01211
0, 1, 8	0.00952	0.01000
0, 1, 2, 3, 4, 5, 9	0.00904	0.00922
0, 1, 2, 4, 5, 6, 9	0.00903	0.00878

Figure 11: Table of Problem 7(c)

Bayesian Core Table 3.10

$t_1(\gamma)$	$\pi(\gamma \vert Y, X)$	$\hat{\pi}(\gamma \vert Y, X)$
0, 1, 2, 4, 5	0.09293	0.09148
0, 1, 2, 4, 5, 9	0.03254	0.03149
0, 1, 2, 4, 5, 10	0.02950	0.02874
0, 1, 2, 4, 5, 7	0.02311	0.02424
0, 1, 2, 4, 5, 8	0.02284	0.02174
0, 1, 2, 4, 5, 6	0.02280	0.02599
0, 1, 2, 3, 4, 5	0.02240	0.02224
0, 1, 2, 3, 4, 5, 9	0.01673	0.01600
0, 1, 2, 4, 5, 6, 9	0.01672	0.01800
0, 1, 2, 4, 5, 8, 9	0.01372	0.01225
0, 1, 4, 5	0.01104	0.01175
0, 1, 2, 4, 5, 9, 10	0.00993	0.01425
0, 1, 2, 3, 9	0.00970	0.01025
0, 1, 2, 9	0.00932	0.00850
0, 1, 2, 4, 5, 7, 9	0.00925	0.00900
0, 1, 2, 6, 9	0.00919	0.00875
0, 1, 4, 5, 9	0.00876	0.00950
0, 1, 2, 3, 4, 5, 10	0.00793	0.00900
0, 1, 2, 4, 5, 8, 10	0.00790	0.00575
0, 1, 2, 4, 5, 7, 10	0.00789	0.00850

Figure 12: Table of Problem 7(d)

Bayesian Core Table 3.11, C=100

Left: Informative, Right: Non-Informative

γ_i	$\hat{P}^\pi(\gamma_i = 1\|Y, X)$	$\hat{P}^\pi(\gamma_i = 1\|Y, X)$
γ_0	0.86579	0.88078
γ_1	0.70892	0.76456
γ_2	0.14854	0.29393
γ_3	0.66370	0.73382
γ_4	0.64915	0.71332
γ_5	0.16565	0.29768
γ_6	0.13243	0.28968
γ_7	0.16743	0.27743
γ_8	0.38962	0.50412
γ_9	0.11188	0.25419

Figure 13: Table of Problem 7(e)