250C Assignment 5

Team Awesome (aka Steph and Shelley)

5/6/2019

Bootstrap confidence interval for RAP

1. Write out the statistical model that you fit to calculate the RAP in terms of the parameters (e.g. β-coefficients). What is the expression for the RAP in terms of the model parameters? (15 points)

The model that we fit was a cox proportional hazards model for time to incident CVD as a function of BMI, age, sex, education and current smoking status. BMI and education were each categorized into 4 categories (with appropriate indicators.)

$$
\begin{aligned}
log[h(t|\boldsymbol{x})] &= log[h_0(t)] + BMI_{underweight} \times \beta_1 + BMI_{overweight} \times \beta_2 + BMI_{obese} \times \beta_3 + \\
&\quad age \times \beta_4 + sex \times \beta_5 + highschool \times \beta_6 + \\
&\quad somecollege \times \beta_7 + morethancollege \times \beta_8 + currentsmoking \times \beta_9
\end{aligned}
$$

The RAP (rate advancement period) of how much obesity increases the risk of CVD with age is the ratio of the log hazard ratio of obesity on CVD, divided by the log hazard ratio of age on CVD. Thus,

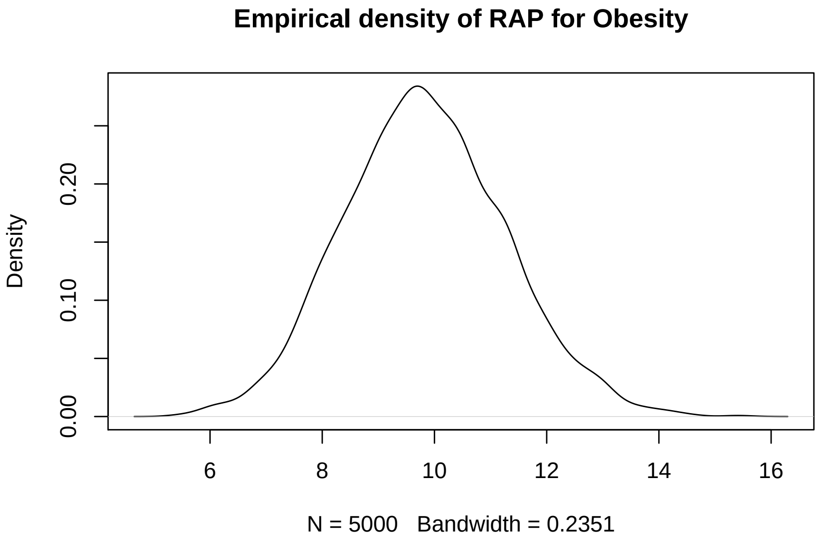
$$RAP_{\text{obesity by age}} = \frac{logHR_{\text{obesity}}}{logHR_{\text{age}}} = \frac{\beta_{\text{obesity}}}{\beta_{\text{age}}} = \frac{\beta_3}{\beta_4}$$

2. What is the value for the RAP for the obese vs. normal weight exposure level and its 95% CI from the delta method? Offer an interpretation of this effect measure. (15 points)

Based on the delta method analysis, the RAP for obese vs. normal weight participants is 9.83 (7, 12.67). This implies that obese subjects reach the same rate of CVD 9.83 years earlier than normal weight subjects assuming that the rate consistently increases with age and conditional on survival to the baseline age. If we were to take many samples of participants from the same underlying population, and calculate a 95% confidence interval from each sample using these methods, 95% of those confidence intervals should contain the true population value for the RAP. (7, 12.67) is one such interval.

3. What is the bootstrapped mean and 95% confidence interval estimate for the RAP for the obesity effect? How does this bootstrapped estimate of the 95% confidence interval compare to the estimate from the delta method (remembering that they are both approximations!)? Turn in the density plot for the bootstrapped RAP. (10 points)

Based on our bootstrap, the RAP is estimated to be 9.83 (7.1, 12.81). This estimate is very similar to the estimate from the delta method. See density plot for the bootstrapped RAP below.



Linear regression model of BMI

4. Using the R code provided, complete Table 1. (20 points)

Table 1. Posterior means and 95% credible intervals for slope coefficient from linear regression model of BMI on age, sex, and education level.

Variable	Vague Prior	Informative Prior 1*	Informative Prior 2†
Age (per year increase)	0.03 (0.02, 0.05)	0.03 (0.02, 0.05)	1.14 (0.97, 1.3)
Female sex (vs. male)	-0.82 (-1.07, -0.57)	-0.82 (-1.07, -0.58)	18.28 (15.47, 21.14)
HS education (vs. <HS)	-0.94 (-1.23, -0.65)	-0.94 (-1.23, -0.65)	-3.61 (-7.02, -0.18)
Some college (vs. <HS)	-1.39 (-1.74, -1.04)	-1.39 (-1.73, -1.04)	0.29 (-3.82, 4.29)
College+ (vs. <HS)	-1.24 (-1.63, -0.84)	-1.24 (-1.64, -0.85)	2.02 (-2.6, 6.61)
Curr. smoker (vs. non)	-1.37 (-1.61, -1.12)	-1.37 (-1.61, -1.12)	94.57 (93.87, 95.28)

*Prior mean for effect of current smoking = 100, prior variance = 1000
†Prior mean for effect of current smoking = 100, prior variance = 0.1225

5. What seems to be more influential, varying the prior mean, or the prior variance? In one sentence, briefly explain what you think is happening? (10 points)

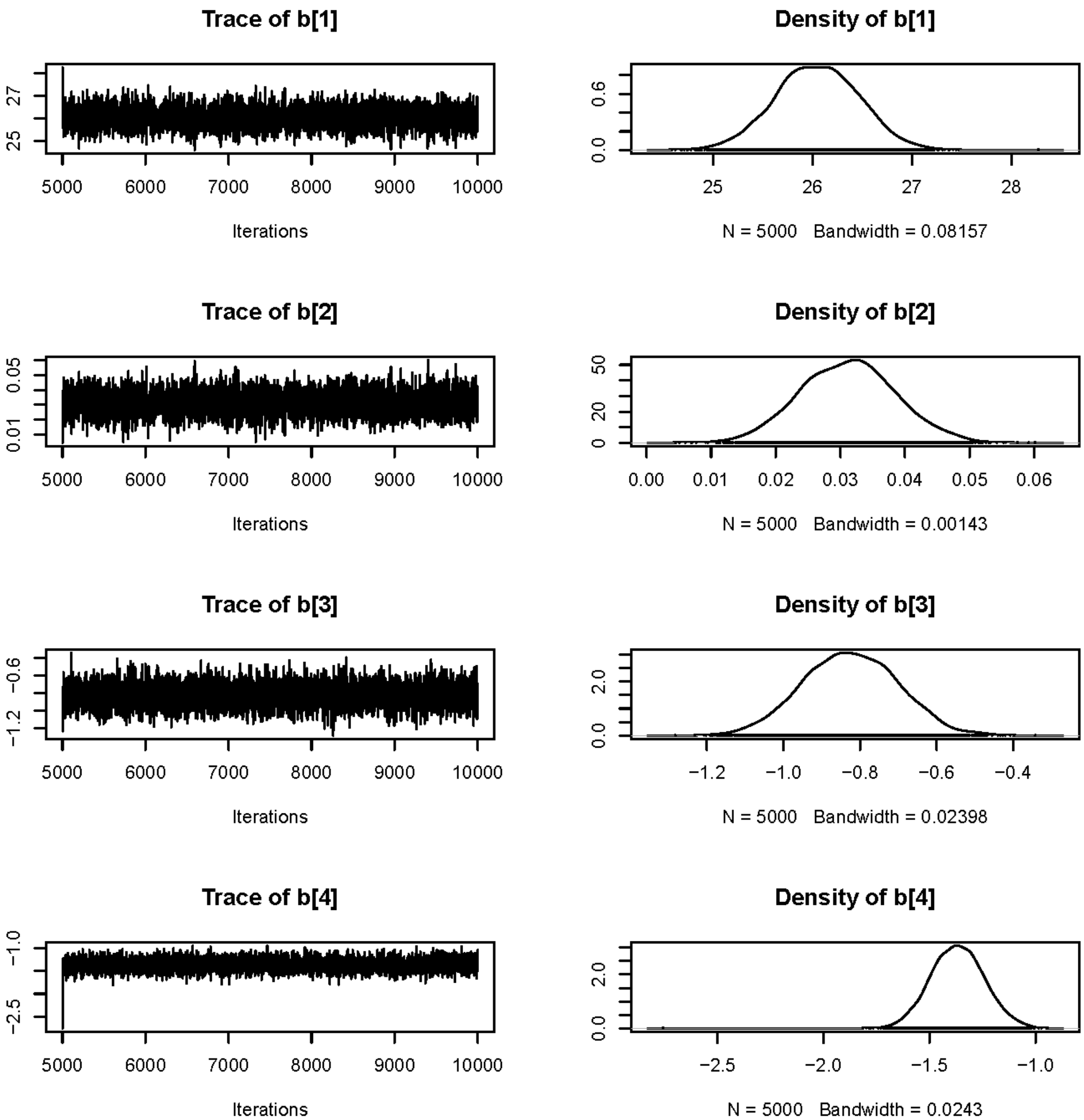
Varying the prior variance was considerably more influential than varying the prior mean. Because the prior and the maximum likelihood from the current data are weighted based on the certainty (i.e. variance in the data), when the variance of the prior is very small it has strong influence, and when the variance is very large it is overwhelmed by the shape of the data itself.

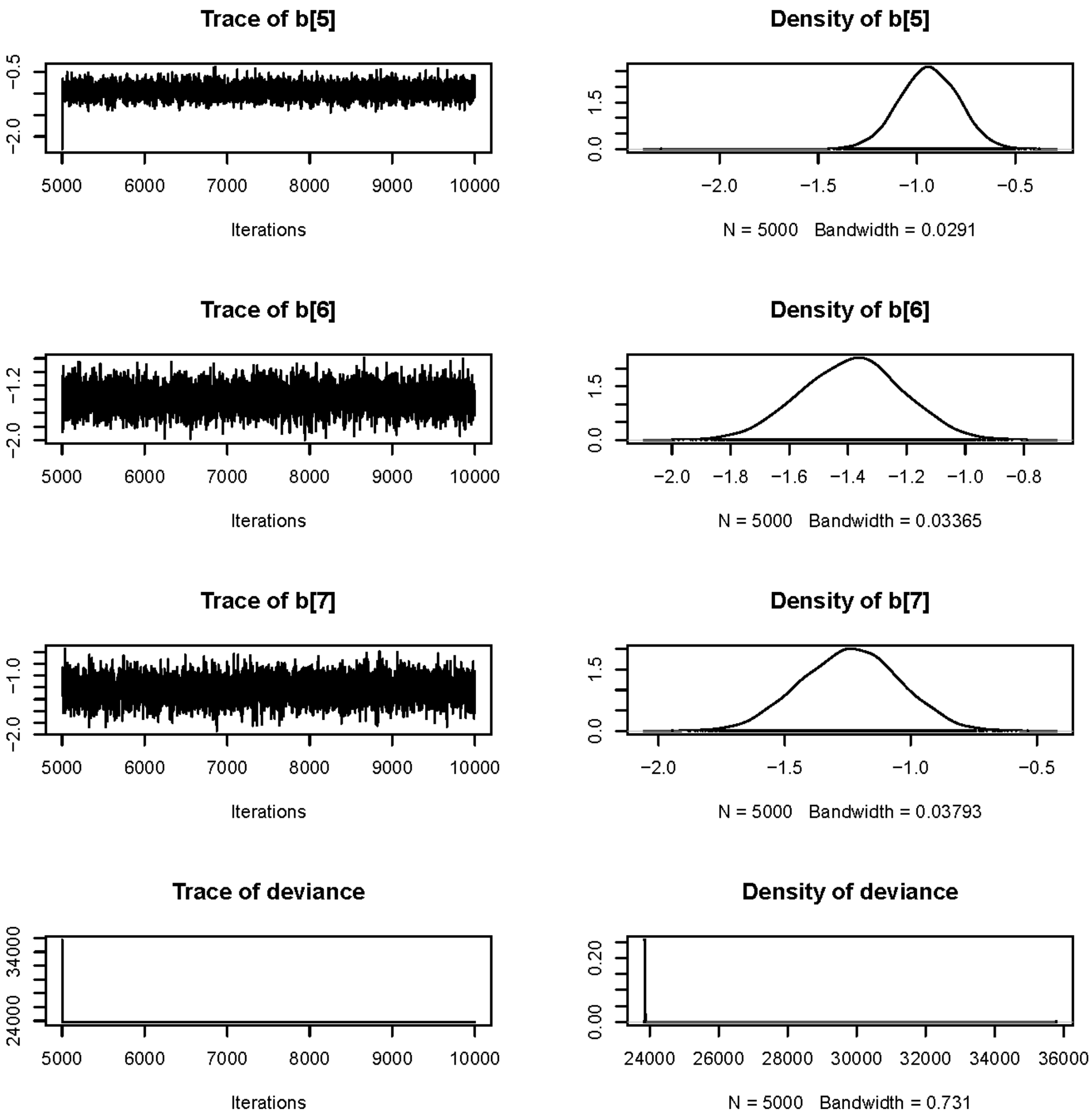
6. Using the trace plots, density plots and autocorrelation plots (focus on 1st chain) from the diagnostics for the first model ("Vague prior"), briefly describe any evidence of convergence (or lack of convergence) that you see. Attach these plots (2 pages for trace/density plots; 1 page for autocorrelation plots). (20 points)

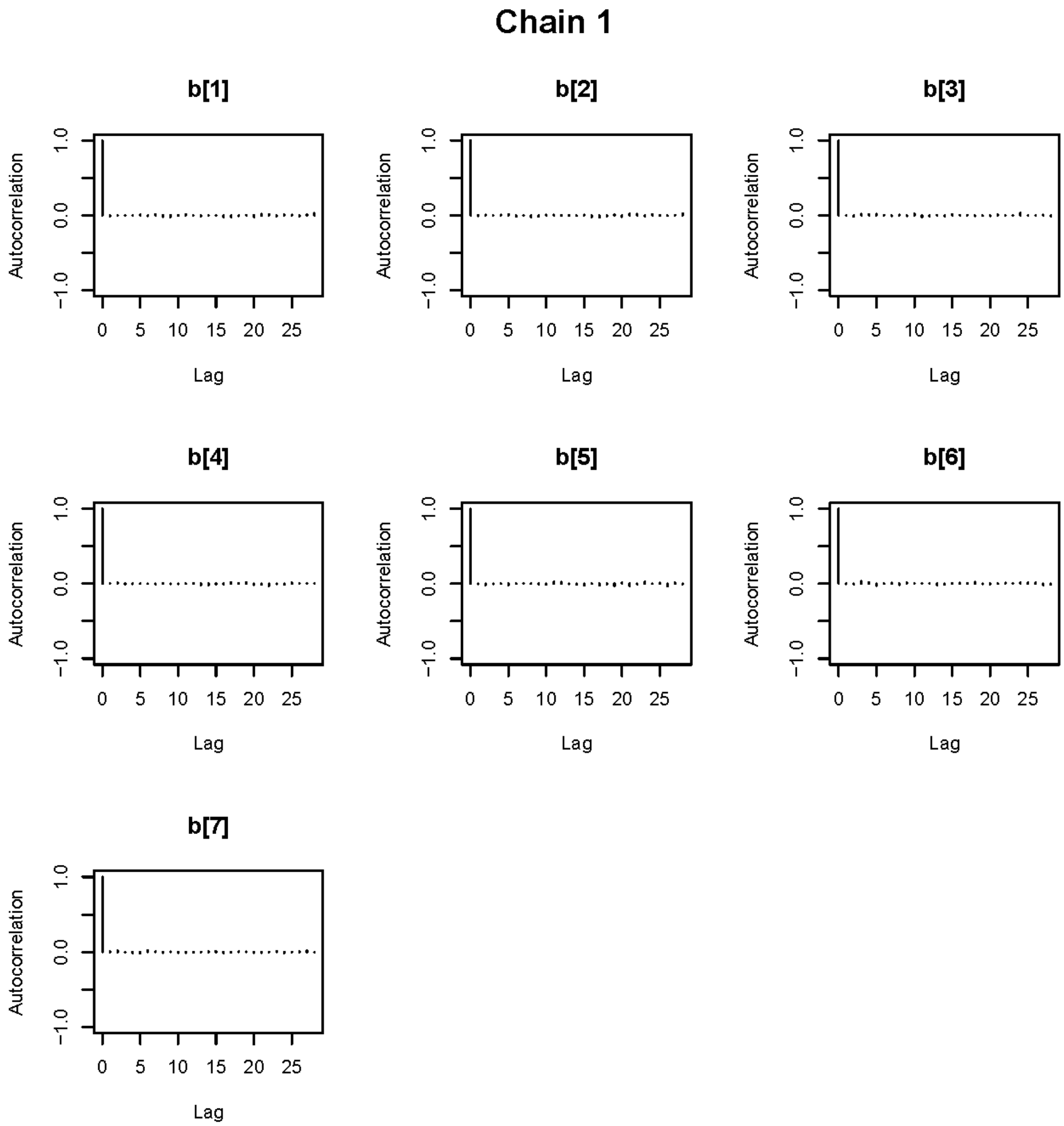
According to the trace plots, density plots, and 1st chain autocorrelation plots, there is strong evidence of convergence. The trace plots show "fuzzy caterpillars" that are straight and essentially display noise along a horizontal line with no sharp jumps, periodicity or trending. The density plots look fairly normally distributed, with only minor bumps and distortions (i.e. like with β_2); nothing concerning. The autocorrelation plots drop nearly to zero starting with k = 1, which is also a sign that the thinning was sufficient. (*See the following pages for the plots referenced in this answer.*)

7. From the results of the Geweke test, is there evidence for lack of convergence? Justify your answer. (10 points)

The largest test statistic (in absolute number) from the Geweke test is -1.82775 in the 1st chain, 1.8920 in the second chain, and -1.41083 in the 3rd chain. Since none of those test statistics are $>$ 1.96 (the threshold for non-convergence), we can feel relatively comfortable we have reached convergence (noting that the test statistics do come fairly close to the threshold).


Trace of b[1]
Iterations
Density of b[1]
N = 5000 Bandwidth = 0.08157
Trace of b[2]
Iterations
Density of b[2]
N = 5000 Bandwidth = 0.00143
Trace of b[3]
Iterations
Density of b[3]
N = 5000 Bandwidth = 0.02398
Trace of b[4]
Iterations
Density of b[4]
N = 5000 Bandwidth = 0.0243


Trace of b[5]
−2.0
−0.5
5000
6000
7000
8000
9000
10000
Iterations
Density of b[5]
0.0
1.5
−2.0
−1.5
−1.0
−0.5
N = 5000 Bandwidth = 0.0291
Trace of b[6]
−2.0
−1.2
5000
6000
7000
8000
9000
10000
Iterations
Density of b[6]
0.0
1.5
−2.0
−1.8
−1.6
−1.4
−1.2
−1.0
−0.8
N = 5000 Bandwidth = 0.03365
Trace of b[7]
−2.0
−1.0
5000
6000
7000
8000
9000
10000
Iterations
Density of b[7]
0.0
1.5
−2.0
−1.5
−1.0
−0.5
N = 5000 Bandwidth = 0.03793
Trace of deviance
24000
34000
5000
6000
7000
8000
9000
10000
Iterations
Density of deviance
0.00
0.20
24000
26000
28000
30000
32000
34000
36000
N = 5000 Bandwidth = 0.731


Chain 1
b[1]
b[2]
b[3]
b[4]
b[5]
b[6]
b[7]
Autocorrelation
−1.0
0.0
1.0
0
5
10
15
20
25
Lag