1. Explain why we prefer to use DIC or WAIC to compare models instead of in-sample deviance.

5

In-sample deviance summarizes how well the model fits the observed data, and always shows "improvement" as models increase in complexity, which would lead to overfitting. DIC and WAIC both attempt to approximate the out-of-sample deviance, which helps guard against over-fitting.

2. Explain what overfitting is and describe one strategy to avoid it.

Overfitting occurs when your statistical model is too closely tuned to the observed data set. When this happens you start modeling features unique to that sample ("peculiarities") rather than the overarching trend. Possible strategies include:

5

1) Regularizing priors
2) Cross validation

3. Suppose $y_1, \ldots, y_n$ form a random sample from $\mathcal{N}(\mu, \sigma^2)$. The joint posterior distribution that results from the reference prior is

5

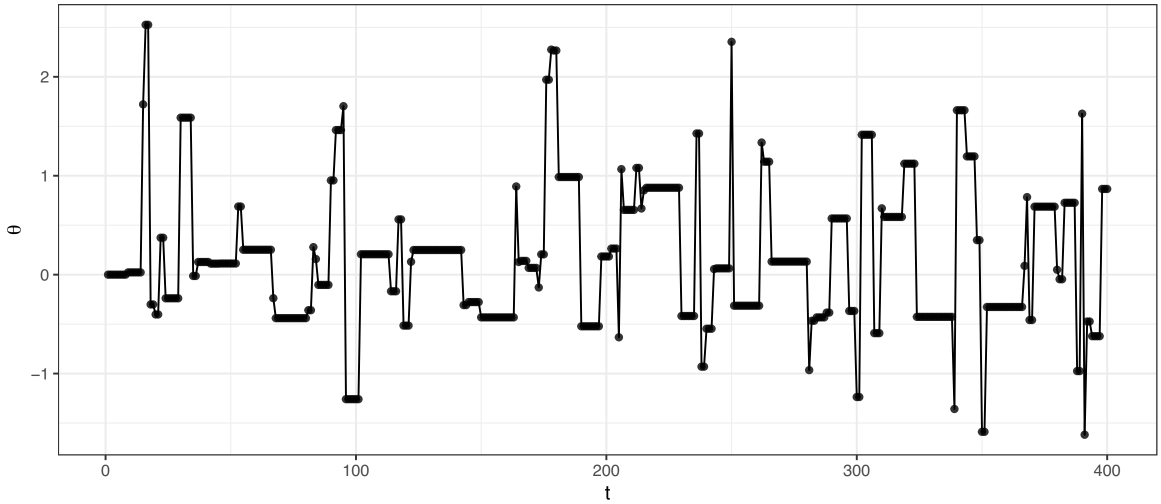
$$p(\mu, \sigma^2 | y_1, \ldots, y_n) \propto \left(\sigma^2\right)^{-n/2-1} \exp\left\{ \sum_{i=1}^{n} -\frac{1}{2\sigma^2}(y_i - \mu)^2 \right\}$$

Find $p(\sigma^2 | \mu, y_1, \ldots, y_n)$, the conditional posterior of σ^2 given μ and the data. If it is a member of a named family of distributions, be sure to specify this, along with its parameter values.

$$p(\sigma^2 | \mu, \vec{y}) \propto (\sigma^2)^{-n/2-1} \exp\left\{ -\frac{1}{2} \sum (y_i - \mu)^2 \Big/ \sigma^2 \right\}$$

This is the kernel of the InvGamma$\left(\frac{n}{2}, \frac{1}{2} \sum (y_i - \mu)^2 \right)$

4. The figure below is a trace plot from 400 steps of an MCMC (Metropolis) run.



(a) Is the acceptance rate: too high, too low, or just right? Briefly explain your reasoning.

The acceptance rate is too low, as can be seen by the many "mini plateaus" in the traceplot. (i.e. the chain is "getting stuck" too often)

5

(b) If the acceptance rate for a random walk Metropolis algorithm using a normal proposal (jump) density is too high, how should the standard deviation be adjusted?

2pts

To decrease the acceptance rate you increase the standard deviation of the proposal distribution.

5. Suppose that you have a random sample, $x_1, \ldots, x_n$, from a Galenshore distribution with PDF

$$f(x_i|\theta) = \frac{2}{\Gamma(a)}\theta^{2a}y^{2a-1}e^{-\theta^2 y^2}$$

where $x_i, \theta > 0$ and a is a known constant. Further, suppose that you put a Gamma(3, 1) prior on θ.

(a) Derive the unnormalized posterior distribution for θ.

6

$$p(\theta|x_1,\ldots,x_n) \propto \prod_{i=1}^{n} \frac{2}{\Gamma(a)} \theta^{2a} x_i^{2a-1} e^{-\theta^2 x_i^2} \cdot e^{-\theta}\theta^2$$

$$= \left(\frac{2}{\Gamma(a)}\right)^n \theta^{2na} \left[\prod x_i^{2a-1}\right] e^{-\theta^2 \sum x_i^2} e^{-\theta}\theta^2$$

$$\propto \theta^{2(na+1)} e^{-\theta\left[1+\theta\sum x_i^2\right]}$$

(b) Describe a method for obtaining draws, $\theta^{(1)}, \ldots, \theta^{(m)}$, from the posterior distribution. If helpful, you may use R function names, but you need to also describe the process.

6

Since we don't know the form of this distribution we could sample from it via the Metropolis algorithm:

① Choose starting value $\theta^{(0)}$.

For $i = 1, \ldots, m$

② Propose $\theta^* \sim N(\theta^{(i-1)}, V)$

③ Calc. $r = \min\left\{\frac{p(\theta^*|\vec{x})}{p(\theta^{(i-1)}|\vec{x})}, 1\right\}$

④ Accept θ^* with probability r.
If accepted, set $\theta^{(i)} = \theta^*$.
Otherwise set $\theta^{(i)} = \theta^{(i-1)}$.

6. Twelve healthy men who did not exercise regularly were recruited to take part in a study of the effects of two different exercise regimen on oxygen uptake. Six of the twelve men were randomly assigned to a 12-week flat-terrain running program, and the remaining six were assigned to a 12-week step aerobics program. The maximum oxygen uptake of each subject was measured (in liters per minute) while running on an inclined treadmill, both before and after the 12-week program. Of interest is how a subjects change in maximal oxygen uptake may depend on which program they were assigned to. However, other factors, such as age, are expected to affect the change in maximal uptake as well.

The researchers considered the following five models:

Model	μ_i
m1	$\mu_i = \alpha$
m2	$\mu_i = \alpha + \beta_1 \texttt{group}_i$
m3	$\mu_i = \alpha + \beta_2 \texttt{age}_i$
m4	$\mu_i = \alpha + \beta_1 \texttt{group}_i + \beta_2 \texttt{age}_i$
m5	$\mu_i = \alpha + \beta_1 \texttt{group}_i + \beta_2 \texttt{age}_i + \beta_3 \texttt{group}_i \times \texttt{age}_i$

(a) Below is the output from the `compare(m1, m2, m3, m4, m5)`. Based on this information, which model would your chose to predict the change in maximal oxygen uptake? Why?

	WAIC	pWAIC	dWAIC	weight	SE	dSE
m4	70.78	7.48	0.00	0.89	11.99	NA
m3	75.20	5.48	4.42	0.10	9.99	6.12
m5	78.72	9.27	7.94	0.02	12.68	8.30
m2	89.23	6.51	18.46	0.00	7.96	12.90
m1	97.41	6.96	26.63	0.00	9.63	13.48

5

We would choose m4 to predict the maximal model uptake, since it has the smallest WAIC value, and a far larger Akaike weight

3. **Regression I.** Suppose that you have been recruited to create a regression model to predict the price of dinner in New York City to help set prices at a new restaurant. Using data from a recent Zagat survey, you fit a multiple linear regression model with the following mean function:

$$\mu(\texttt{price}_i|\texttt{X}_i) = \alpha + \beta_1\texttt{food} + \beta_2\texttt{decor} + \beta_3\texttt{service},$$

where the `food`, `decor`, and `service` variables are average customer ratings out of 30 points, and `price` is recorded in dollars. You fit this multiple linear regression model using the reference prior distribution for multiple linear regression and obtain the following results:

Parameter	Mean	StdDev	5.5%	94.5%
α	-24.642	4.697	-32.148	-17.136
β_1	1.556	0.369	0.967	2.145
β_2	1.847	0.215	1.504	2.191
β_3	0.135	0.391	-0.490	0.760
σ	5.734	0.313	5.234	6.234

(a) Give a careful interpretation of the *maximum a posteriori* estimate of β_3, in the context of the problem.

6 pts

A one-point increase in service score is associated with a \$0.135 increase in the cost of dinner, holding all other variables constant.

(b) Does `service` appear to be an important predictor of price, after controlling for `food` and `decor`? Justify your answer.

3 pts

No. The credible interval for β_3 contains 0, so it is very likely that service is not an important predictor after accounting for food and decor.