Applied Bayesian modeling - HW2

Score: The maximum number of points in this HW is 8 points, with 2 points extra credit. For calculating a final HW grade, the points will be rescaled to a maximum score of (8+2)/8*100% = 125%.

What to hand in: For any exercises using R, we need an Rmd and a knitted pdf. You can add the answers to other exercises in the same markdown file or create them in a different way (as long as it's legible). if you create them outside markdown, please combine all answers into one pdf.

Note that material related to extra credit questions can appear on quizzes and exams.

Exercise 1: Derivation of a posterior using Bayes' rule [3 pts]

This exercise is about the material in module 4.

Obtain $p(\mu|y)$ when prior and data are defined as follows:

$$y_i|\theta_i,\sigma^2 \sim N(\theta_i,\sigma^2)(\text{independent}), \text{ for } i = 1,2,\hdots,n;$$
$$\theta_i = \mu + r_i;$$
$$\mu \sim N(m_0,s_{\mu0}^2);$$

where r_i refers to a known/fixed reporting error (that varies across observations) and σ^2 is known.

Note that you can treat the r_i as fixed, not random.
(A more realistic but somewhat more difficult exercise would be to have r_i be random, and assume that they are normally distributed and independent of the y_i's).

Hint: consider writing $z_i = y_i - r_i$ and then derive $p(\mu|z)$.

Exercise 2: Interpretation of the role of prior information for estimating IQ scores [5 pts]

This exercise is about the material in module 4, based on Hoff 5.4 (but using a different prior on μ).

Background: Scoring on IQ tests is designed to produce a normal distribution with a mean of 100 and a standard deviation of 15 (a variance of 225) when applied to the general population.

Suppose we are to sample n individuals from a particular town in the United States and then estimate μ, the town-specific mean IQ score, based on the sample of size n.
Let y_i denote the IQ score for the i-th person in the town of interest, and assume
$y_1,y_2,\hdots,y_n|\mu,\sigma^2 \sim N(\mu,\sigma^2)$ (independent), with $\sigma = 15$. Suppose that $\bar{y} = 113$ and $n = 10$.

For Bayesian inference about μ, the following prior will be used:

$$\mu \sim N(m_0,s_{\mu0}^2),$$

with $m_0 = 100$ and $s_{\mu0} = \sigma = 15$ (based on the information about IQ scores).

Exercises: (i) Given the information above, obtain the expression for the Bayesian point estimate of μ, $\hat{\mu}_{Bayes} = E(\mu|y)$, in terms of m_0, n, $\bar{y}$, and/or σ. (Note that the expression simplifies relative to the expression we obtained in module 4, using only a subset of these).

    ii. Interpret the Bayesian point estimate as the weighted combination of prior information and data

    iii. Then calculate the value for $\hat{\mu}_{Bayes}$ given the information provided and construct a 95% credible interval for μ.

Hint for (iii): think carefully about what variance term to to use when constructing the CI.

Exercise 3 [Extra credit, 2pts]

We continue with the set up of exercise 2. We will now compare the sampling properties of the Bayes estimator for μ, $\hat{\mu}_{Bayes} = E(\mu|y)$, to the maximum likelihood estimator $\hat{\mu}_{MLE} = \bar{y}$. Sampling properties of estimators refer to their behavior under hypothetically repeatable surveys or experiments, summarized into bias and mean squared error, as explained below with μ^* referring to the (unknown) true value of μ:
- The bias of an estimator is defined as the difference between its expected value (w.r.t. data y) and the (unknown) true value:

$$Bias(\hat{\mu}) = E_y[\hat{\mu}|\mu^*] - \mu^*$$

(adding the subscript with the expectation here to make it clear that the expectation is wrt y, to clarify that the data are what's random here; also adding conditioning on true value μ^*)
- Bias refers to how close the center of mass of the sampling distribution of an estimator is to the true value. An unbiased estimator is an estimator with zero bias, which sounds desirable. However, bias does not tell us how far away an estimate might be from the true value. For example, y_1 is an unbiased estimator of the population mean μ, but will generally be farther away from μ than $\bar{y}$.
- To evaluate how close an estimator $\hat{\mu}$ is likely to be to the true value μ^*, we might use the mean squared error (MSE). Letting $m = E_y[\hat{\mu}|\mu^*]$, the MSE of an estimator $\hat{\mu}$ is

$$
\begin{aligned}
MSE[\hat{\mu}|\mu^*] &= E_y[(\hat{\mu} - \mu^*)^2|\mu^*], \\
&= E_y[(\hat{\mu} - m + m - \mu^*)^2|\mu^*], \\
&= E_y[(\hat{\mu} - m)^2|\mu^*] + 2E_y[(\hat{\mu} - m)(m - \mu^*)] + E[(m - \mu^*)^2|\mu^*], \\
&= E_y[(\hat{\mu} - m)^2|\mu^*] + (m - \mu^*)^2, \\
&= Var[\hat{\mu}|\mu^*] + Bias(\hat{\mu})^2.
\end{aligned}
$$

This means that, before the data are gathered, the expected distance from the estimator to the true value depends on how close μ^* is to the center of the distribution of $\hat{\mu}$ (the bias), as well as how spread out the distribution is (the variance of $\hat{\mu}$).

Exercise

Suppose that (unknown to us) the true mean IQ score μ^* in the town was quite high, $\mu^* = 112$. Calculate the bias, variance and MSE of the Bayes and ML estimators. Which estimator has a larger bias? Which estimator has a larger MSE?
Hint (based on a commonly made mistake): $E_y(\hat{\mu}_{Bayes}|\mu^*) \neq \hat{\mu}_{Bayes}$, you need to get the expected value w.r.t. the data y.

Optional: Consider

(i) plotting the sampling distributions for both the Bayes estimator as well as the MLE.

(ii) obtaining the Bayes and ML MSEs for sample sizes $n = 10$ to 1,000 and plotting the ratio (Bayes MSE)/(ML MSE) against sample size, to then interpret your findings.