

Applied Bayesian modeling - HW2 solutions

Score: The maximum number of points in this HW is 12 points, with 3 points extra credit. For calculating a final HW grade, the points will be rescaled to a maximum score of $(12+3)/12 \cdot 100\% = 125\%$.

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.

Part 1 - based on module 4

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

This exercise is about the material in module 4.

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

$$\begin{aligned} y_i|\theta_i, \sigma^2 &\sim N(\theta_i, \sigma^2) \text{ (independent), for } i = 1, 2, \dots, n; \\ \theta_i &= \mu + r_i; \\ \mu &\sim N(m_0, s_{\mu 0}^2); \end{aligned}$$

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).

Answer

When we write $z_i = y_i - r_i$, then with r_i being not random, we get that $p(\mu|\mathbf{y}) = p(\mu|\mathbf{z})$. We can directly apply the derivation from the slides, as follows:

$$\begin{aligned} p(\mu|\mathbf{y}) &= p(\mu|\mathbf{z}), \\ &\propto p(\mu)p(\mathbf{z}|\mu) \\ &\propto \exp\left(\frac{-1}{2s_{\mu 0}^2}(\mu - m_0)^2\right) \cdot \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^n (z_i - \mu)^2\right), \\ &\propto \exp\left(\frac{-1}{2s_{\mu 0}^2}(\mu - m_0)^2\right) \cdot \exp\left(\frac{-n}{2\sigma^2}(\bar{z} - \mu)^2\right) \\ &\propto \exp\left(-\frac{1}{2}f(\mu)\right), \end{aligned}$$

where

$$\begin{aligned} f(\mu) &= 1/s_{\mu 0}^2(\mu^2 - 2 \cdot m_0 \cdot \mu) + n/\sigma^2(\mu^2 - 2 \cdot \bar{z} \cdot \mu) \\ &\propto (1/s_{\mu 0}^2 + n/\sigma^2)\mu^2 - 2 \cdot (1/s_{\mu 0}^2 m_0 + n/\sigma^2 \bar{z}) \cdot \mu, \end{aligned}$$

Let $v = (1/s_{\mu 0}^2 + n/\sigma^2)^{-1}$, then

$$f(\mu) \propto 1/v(\mu^2 - v \cdot 2 \cdot (1/s_{\mu 0}^2 m_0 + n/\sigma^2 \bar{y})\mu).$$

Comparing this to $f(\mu) \propto \frac{1}{v}(\mu^2 - 2m\mu)$, we recognize that

$$\mu|\mathbf{y}, \sigma^2 \sim N\left(\frac{m_0/s_{\mu 0}^2 + n \cdot \bar{y}/\sigma^2}{1/s_{\mu 0}^2 + n/\sigma^2}, \frac{1}{1/s_{\mu 0}^2 + n/\sigma^2}\right).$$

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 $y_1, y_2, \dots, y_n | \mu, \sigma^2 \sim N(\mu, \sigma^2)$ (independent), with $\sigma = 15$. Suppose that $\bar{y} = 113$ and $n = 10$.

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

$$\mu \sim N(m_0, s_{\mu 0}^2),$$

with $m_0 = 100$ and $s_{\mu 0} = \sigma = 15$ (based on the information about IQ scores).

Exercise 2a

- (i) Given the information above, obtain the expression for the Bayesian point estimate of μ , $\hat{\mu}_{Bayes} = E(\mu|\mathbf{y})$, 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 .

Answer

- (i) We get

$$\mu|\mathbf{y}, \sigma^2 \sim N(V(m_0/\sigma_{\mu 0}^2 + n \cdot \bar{y}/\sigma^2), V),$$

where $V = \frac{1}{1/\sigma_{\mu 0}^2 + n/\sigma^2}$. Because $\sigma_{\mu 0} = \sigma$, the expressions simplify a lot: $V = \sigma^2/(n+1)$ and $\hat{\mu}_{Bayes} = \frac{m_0 + n\bar{y}}{n+1}$.

- (ii) Because prior and data variance are equal, here the mean becomes the weighted average of prior mean and \bar{y} with relative weights 1 and n , respectively,
- (iii) The point estimate is given by $\hat{\mu}_{Bayes} = 111.8$ and the 95% CI is given by $\hat{\mu}_{Bayes} \pm 1.96 \cdot \sqrt{V}$, or equivalently, the quantiles of the normal distribution), and ranges from 103.0 to 120.7.

Note: some students used the prior variance here to get the CI, this is not correct because that variance refers to the uncertainty about μ prior to observing the data. After we observe the data, we get the posterior for μ , with updated point estimate and uncertainty relative to the prior.

```
# priors for mean and sd IQ scores
mu0 <- 100
sigma.0 <- 15
sigma.mu0 <- sigma.0

# assume sigma.y known
sigma.y <- 15

# info from town
ybar <- 113
n <- 10

post_var <- (1/sigma.mu0^2 + n/sigma.y^2)^(-1)
post_mean <- post_var*(mu0/sigma.mu0^2 + n*ybar/sigma.y^2)

# estimate and CI
post_mean
```

```
## [1] 111.8182
```

```
qnorm(c(0.025, 0.975), post_mean, sqrt(post_var) )
```

```
## [1] 102.9539 120.6825
```

2b Extra credit [3pts]

We will now compare the sampling properties of the Bayes estimator for μ , $\hat{\mu}_{Bayes} = E(\mu|\mathbf{y})$, 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 \mathbf{y}) and the (unknown) true value:

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

(adding the subscript with the expectation here to make it clear that the expectation is wrt \mathbf{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 $m = E_{\mathbf{y}}[\hat{\mu}|\mu^*]$, the MSE of an estimator $\hat{\mu}$ is

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

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): $E_{\mathbf{y}}(\hat{\mu}_{Bayes}|\mu^*) \neq \hat{\mu}_{Bayes}$, you need to get the expected value w.r.t. the data \mathbf{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.

Answer (leaving out the conditioning on μ^* for notational convenience).

For the Bayes estimator, we find that the bias is $E_{\mathbf{y}}(\hat{\mu}_{Bayes}) - \mu^* = V(m_0/\sigma_{\mu_0}^2 + n \cdot \mu^*/\sigma^2) - \mu^* = \frac{1}{n+1}(m_0 - \mu^*) \approx -1.1$.

For the MLE we find that the bias is $E(\hat{\mu}_{MLE}) - \mu^* = \mu^* - \mu^* = 0$.

The MLE is unbiased, the Bayes estimator has a bias (which goes to zero as the sample size increases).

For the variances (simplifies for the bayes estimate based on settings of prior used but worked out here for general expression):

$$\begin{aligned} Var[\hat{\mu}_{Bayes}] &= Var(V(m_0/\sigma_{\mu_0}^2 + n \cdot \bar{y}/\sigma^2)), \\ &= Var(Vn \cdot \bar{y}/\sigma^2), \\ &= V^2 n^2 / \sigma^4 \cdot \sigma^2 / n, \\ &= n\sigma^2 / (n+1)^2, \\ &\approx 18.6. \end{aligned}$$

$$\begin{aligned} Var[\hat{\mu}_{MLE}] &= \sigma^2 / n, \\ &= 15^2 / 10 = 22.5. \end{aligned}$$

\Rightarrow The variance is smaller for the Bayes estimator as compared to the MLE.

For the MSE:

$$\begin{aligned} MSE[\hat{\mu}_{Bayes}] &= Bias(\hat{\mu}_{Bayes})^2 + Var[\hat{\mu}_{Bayes}], \\ &= -1.1^2 + 18.6 = 19.8. \end{aligned}$$

$$\begin{aligned} MSE[\hat{\mu}_{MLE}] &= Bias(\hat{\mu}_{MLE})^2 + Var[\hat{\mu}_{MLE}], \\ &= 0 + 15^2 / 10 = 22.5. \end{aligned}$$

\Rightarrow The MSE is smaller for the Bayes estimator as compared to the MLE.

Answers to optional questions:

The sampling distribution for both estimators is Normal, given that they each are linear combinations of normal random variables, with means and variances as worked out above. The plot below shows both pdfs to illustrate the small bias-variance trade-off.

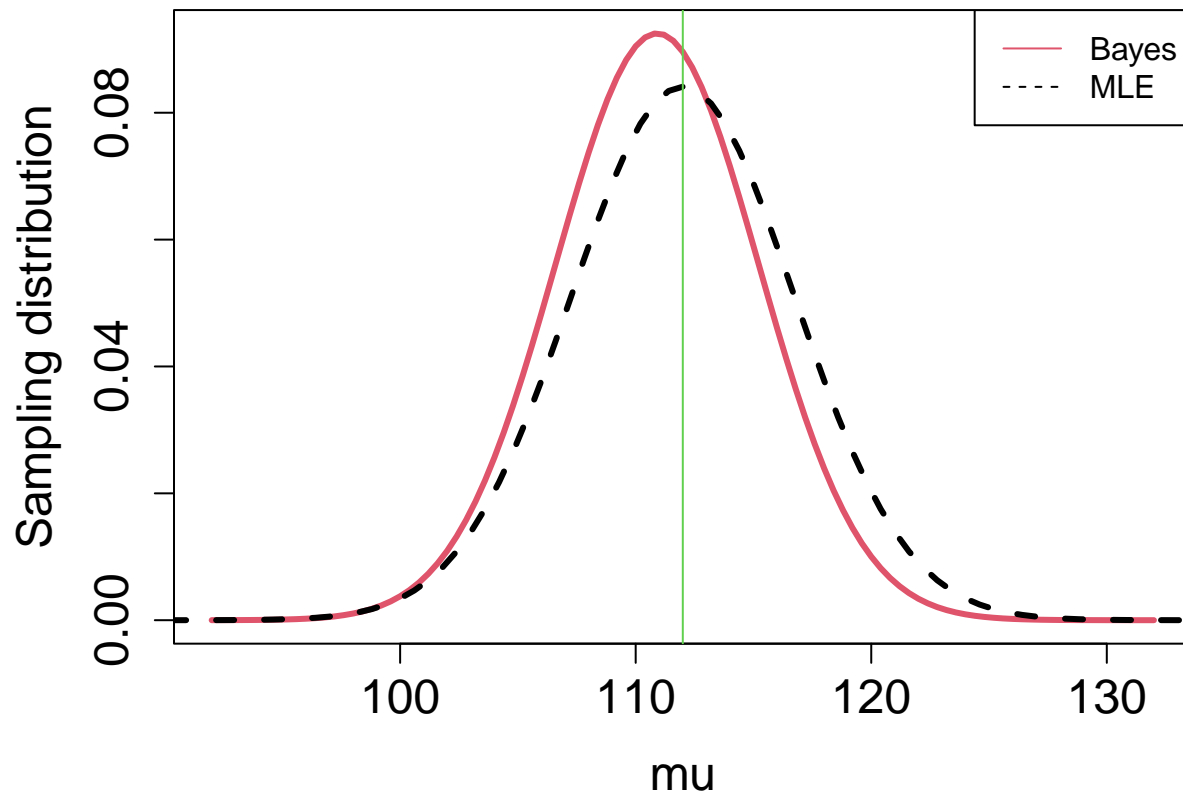
```

# settings
mustar <- 112 # true mu in town

# little function to get info from posterior for mu
# being lazy and not using input arguments, most are defined globally
Bayes <- function(n) {
  V <- (1 / sigma.mu0 ^ 2 + n / sigma.y ^ 2) ^ (-1)
  point <- V * (mu0 / sigma.mu0 ^ 2 + n * mustar / sigma.y ^ 2)
  add <- qnorm(0.975) * sqrt(V)
  samplingvarpointestimate <- V ^ 2 * n / (sigma.y ^ 2)
  return(
    list(
      point = point,
      V = V,
      samplingvarpointestimate = samplingvarpointestimate,
      low = point - add,
      up = point + add
    )
  )
}

# plot sampling distr for n=10
n <- 10
par(
  mfrow = c(1, 1),
  cex.axis = 1.5,
  cex.lab = 1.5,
  mar = c(5, 5, 1, 1)
)
curve(
  dnorm(x, Bayes(n)$point, sqrt(Bayes(n)$samplingvarpointestimate)),
  xlim = c(mustar - 20, mustar + 20),
  col = 2,
  lwd = 3,
  xlab = "mu",
  ylab = "Sampling distribution"
)
curve(
  dnorm(x, mustar, sigma.y / sqrt(n)),
  xlim = c(mustar - 30, mustar + 30),
  add = TRUE,
  col = 1,
  lty = 2,
  lwd = 3
)
abline(v = mustar, col = 3)
legend(
  "topright",
  legend = c("Bayes", "MLE"),
  col = c(2, 1),
  lty = c(1, 2)
)

```



The ratio of MSEs goes to 1 as the sample size increases (because prior info becomes negligible).

```
# simulate bias and MSE, with a little function again
BiasEtc <-
  function(est = "Bayes", # choice of Bayes or else defaults to MLE
           n = n) {
    if (est == "Bayes") {
      V <- Bayes(n)$V
      bias <- V * (mu0 / (sigma.mu0 ^ 2) + n * mustar / (sigma.y ^ 2)) - mustar
      var <- Bayes(n)$samplingvarpointestimate
    } else {
      bias <- 0
      var <- sigma.y ^ 2 / n
    }
    return(list(
      var = var,
      bias = bias,
      MSE = bias ^ 2 + var
    ))
  }
BiasEtc(n = 10)
```

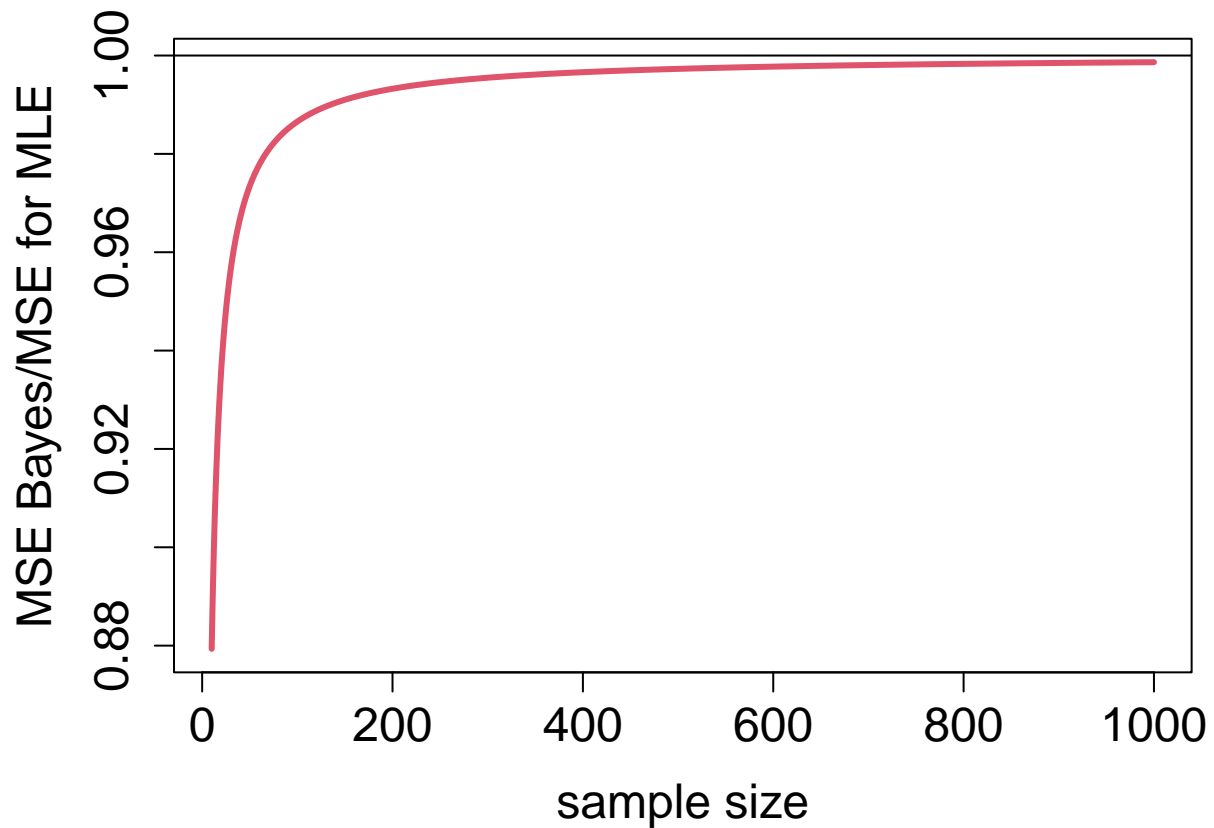
```
## $var
## [1] 18.59504
##
## $bias
## [1] -1.090909
##
## $MSE
```

```
## [1] 19.78512
```

```
BiasEtc(est = "MLE", n = 10)
```

```
## $var  
## [1] 22.5  
##  
## $bias  
## [1] 0  
##  
## $MSE  
## [1] 22.5
```

```
samplesizes <- seq(10, 1000)  
#pdf(paste(figdir, "hw3_samplingMSEratios.pdf", sep = ""))  
par(  
  mfrow = c(1, 1),  
  cex.axis = 1.5,  
  cex.lab = 1.5,  
  mar = c(5, 5, 1, 1)  
)  
plot(  
  BiasEtc(est = "Bayes", n = samplesizes)$MSE /  
    BiasEtc(est = "MLE", n = samplesizes)$MSE ~ samplesizes,  
  type = "l",  
  col = 2,  
  lwd = 3,  
  ylab = "MSE Bayes/MSE for MLE",  
  xlab = "sample size"  
)  
abline(h = 1)
```



```
#dev.off()
```

Part 2 - based on module 5

Exercise 3: get stan going on your laptop [4pts]

The goal of this exercise is just to make sure you have stan (specifically, through rstan and brms) working on your laptop. To do so, please work through the Rmd file with module5, module5_sampling.Rmd.

For this HW Rmd, please add an example model fit using brms. This can be a model fit to radon data, copied from the module5 example, or a model fit from the brms examples. Also let us know if you successfully installed rstan and got it to work. If not, please explain the issue.

Solutions

Please see module5_sampling.