Bayesian Inference, Shrinkage Estimators, and MSE

M&M's Example: Bias, Variance, Efficiency, and MSE of maximum likelihood estimator and Bayesian posterior mean

Set Up

Problem Reminder

In lab 7, we estimated the proportion of M&M's that are blue, θ, based on a sample of n M&M's. We defined the random variable X, which was the count of how many M&M's were blue in that sample. Our model was $X \overset{\text{i.i.d.}}{\sim} \text{Binomial}(n, \theta)$.

We have developed two approaches to inference for θ:

1. The maximum likelihood estimator $\hat{\theta}_{MLE} = \frac{X}{n}$

2. A Bayesian approach with conjugate prior distribution given by $\Theta \sim Beta(a, b)$. The posterior distribution is given by $\Theta | n, x \sim Beta(a + x, b + n - x)$. From this posterior distribution, we can obtain point estimates, with the most common choice being the posterior mean, which can be written as follows:

$$\hat{\theta}^{Bayes} = \frac{a + x}{(a + x) + (b + n - x)} = \cdots = (1 - w)\frac{a}{a + b} + w\frac{x}{n},$$

where $w = \frac{n}{n+a+b}$.

By considering what the posterior mean would be across different samples, we can view the posterior mean as specifying an estimator,

$$\hat{\theta}^{Bayes} = (1 - w)\frac{a}{a + b} + w\frac{X}{n}.$$

In the previous lab, we considered three different prior specifications for Θ with different level of informativeness, and we saw that for a large sample size it didn't matter which prior we used; the resulting estimates were essentially the same as each other and as the maximum likelihood estimate.

Let's now look more closely at how the Bayesian and Frequentist methods compare for smaller sample sizes. For concreteness, let's work with the intermediate of our three prior specifications, $\Theta \sim \text{Beta}(2, 8)$. Let's also imagine that our sample size is fixed at 10.

In this case, $w = \frac{n}{n+a+b} = \frac{10}{10+2+8} = 0.5$, and the Bayesian estimator is

$$\hat{\theta}^{Bayes} = (1 - w)\frac{a}{a + b} + w\frac{X}{n} = 0.5\frac{2}{10} + 0.5\frac{X}{10}$$

Properties of $\hat{\theta}_{MLE}$

Find the bias of $\hat{\theta}_{MLE}$

Find the variance of $\hat{\theta}_{MLE}$ in terms of θ

Find an expression for the MSE of $\hat{\theta}_{MLE}$ in terms of θ

Find the Fisher information about θ from a sample of size n.

You may use the fact that

$$\frac{d^2}{d\theta^2}\ell(\theta|x) = \frac{-x}{\theta^2} + \frac{-(n-x)}{(1-\theta)^2}$$

Is $\hat{\theta}_{MLE}$ an efficient estimator of θ?

Suppose $\tilde{\theta}$ is an unbiased estimator of θ. Is it possible that $Var(\tilde{\theta}) < Var(\hat{\theta}_{MLE})$?

Properties of $\hat{\theta}^{Bayes}$

Find the bias of $\hat{\theta}_{Bayes}$ in terms of θ. For what value of θ is $\hat{\theta}_{Bayes}$ unbiased? How does that relate to the prior distribution?

Find the variance of $\hat{\theta}_{Bayes}$ in terms of θ

Find an expression for the MSE of $\hat{\theta}_{Bayes}$ in terms of θ

The last part of this lab is on GitHub, Lab 13.