

Homework 6: Frequentists vs Bayesians

AUTHORS: Jed Pulley

DO NOT POLLUTE! AVOID PRINTING, OR PRINT 2-SIDED MULTIPAGE.**Problem 6.1. Frequentist (MLE)**

To find the MLE of p^* , we first start with the likelihood function:

$$\mathbb{P}(p^*) = \prod_{i=1}^n (p^{x_i} (1-p)^{1-x_i})$$

Then we take the log of the likelihood function:

$$\log(\mathbb{P}(p^*)) = \log p \sum_{i=1}^n x_i + \log(1-p) \sum_{i=1}^n (1-x_i)$$

Using our optimization 101 technique, we get the derivative and set it equal to zero:

$$p_{MLE} = \frac{1}{n} \sum_{i=1}^n x_i$$

Which we recognize to just be the mean.

Problem 6.2. Bayesian (MAP)

We use our MAP formula $p_{MAP} = \arg \max_p \mathbb{P}(X|p)$ as our starting point. Using Bayes Rule, we can rearrange it as such:

$$p_{MAP} = \arg \max_p \frac{\mathbb{P}(X|p)\mathbb{P}(p)}{\mathbb{P}(X)}$$

Since we're maximizing for p , we can ignore the $\mathbb{P}(X)$ term as it doesn't depend on p :

$$p_{MAP} = \arg \max_p \mathbb{P}(X|p)\mathbb{P}(p)$$

This is very similar to our MLE statement above, with the exception of our prior term $\mathbb{P}(p)$ which we're assuming is information gathered about a previous event. Notably, if there is no prior information, our MAP estimate is equal to our MLE.

Given our prior $\mathbb{P}(p)$ being modeled as $Beta(\alpha, \beta)$, we can show our prior below:

$$\mathbb{P}(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

Using the prior information, the likelihood of our data (we'll say $X = [x_1, x_2, \dots, x_N]^T$) can be written as:

$$\mathbb{P}(X|p) = p^{\sum_{i=1}^N x_i} (1-p)^{N-\sum_{i=1}^N x_i} = p^{\mathbf{1}^T X} (1-p)^{N-\mathbf{1}^T X}$$

Simplifying and matching with the form of a Beta distribution, we find our MAP estimator to be:

$$\hat{p}_{MAP} = \frac{\sum_{i=1}^N x_i + \alpha - 1}{N + \alpha + \beta - 2}$$

Problem 6.3. (Correct prior)

- (a) To find the number of samples for the MLE, we need to start with the variance of the MLE when it's distributed bernoulli, which in our case is $\frac{p^*(1-p^*)}{N}$. We want this variance to be $\leq 0.01^2$ for our estimator to be within 0.01 of p^* . So now we plug and chug our numbers:

$$\frac{0.99(1-0.99)}{N} \leq 0.01^2 = \frac{0.99 * 0.01}{N} \leq 0.0001 = N \geq 99$$

So we would need at least 99 samples to be within 0.01 of p^* .

- (b) While the variance of MAP is somewhat trickier to analyze directly, we can use the Central Limit Theorem to assume a normal distribution with mean p^* and variance $\frac{p^*(1-p^*)}{N}$, which happens to be the same as above. Therefore, in both cases, we require at least 99 samples to be within 0.01 of the true value $p^* = 0.99$.

Problem 6.4

- (a) Going through the same process as before, we plug and chug with different values:

$$\frac{0.01(1-0.01)}{N} \leq 0.01^2 = \frac{0.01 * 0.99}{N} \leq 0.0001 = N \geq 99$$

Oddly enough, this shows the same answer as the previous problem of $N \geq 99$.

- (b) Using the same logic as 6.3 and 6.4a, we can also say $N \geq 99$.

Problem 6.5

Based on my results, the answers to 6.3 and 6.4 were exactly the same. So in my opinion, it comes down to user preference. I personally prefer the MLE approach since it doesn't rely on a prior, which in many instances is a naive assumption that you would have one to begin with. That being said, when you actually do have a prior, it's incredibly helpful. So what I would prefer to use largely depends on the situation (i.e. given an actual prior or not).

The advantage of MLE is its simplicity and ease of use. Especially since in a lot of cases the MLE just turns out to be the mean. The disadvantage is the complement of the above: it's too simple and often just winds up being the mean. Using MLE can constrain you depending on the situation.

The advantage of MAP lies entirely in the assumption of a prior, especially given that MLE and MAP are exactly the same thing if there is no prior at all. Having the prior allows us to affectively leverage previous information to influence our decision. The disadvantage is that often times we don't have a prior or we need to make naive assumptions to use one.