# Handout 16: Bayesian Statistics I

Today, we will once again consider the task of estimating the parameter $p$ from a random variable $X$ taken from a $Bin(n, p)$ distribution with a known value of $n$. The approach, though, will be quite different than our previous attempts.

The core novel idea behind Bayesian statistics can be summarized as this: use probability distributions to model our uncertainty in the unknown parameters of a distribution. So, say that before observering any data, we think that any value of $p$ is equally likely. We could write this by defining our unknown parameter $P$ to be a random variable with a uniform distribution:

$$P \sim Unif(0, 1).$$

This is called the **prior distribution**, because it reflects our knowledge of $P$ prior to observing any data.[1] Now, when describing the random variable $X$, we have to give its distribution conditioned on a specific value of the random variable $p$. That is, we need to write this (which we call the **likelihood**, following the notation from the MLE):

$$X|P \sim Bin(n, P).$$

Now, the important thing is describing our knowledge about $P$ *after* observing the data. That is, we want to know the distribution of $P|X$. Bayes rule tells us that we can calculate this as:

$$f_{P|X}(p|x) = \frac{f_{X|P}(x|p) \times f_P(p)}{f_X(x)}.$$

This quantity is called the **posterior distribution**. Determining the form of the posterior distribution is the key task in generating Bayesian estimators. One simplifying step is to notice that the denominator does not depend on $p$, so we can replace it with a constant, adding it back later (if needed) by whatever number makes the posterior a proper distribution (in other words, it integrates to 1). This gives the following standard form:

$$f_{P|X}(p|x) \propto \frac{f_{X|P}(x|p) \times f_P(p)}{f_X(x)}.$$

I will try to keep the subscripts on the density functions $f$ for clarity in the notes. However, on the board I will almost always drop them. Feel free to do the same in your work.

Now let's actually find the posterior distribution for this specific example. We have the following form of the density function (keep

[1] As I have with some other cases, I am going to introduce the terminology and notation of Bayesian estimation using a specific example. How to extend this to other cases should be clear after noting that the variables called $P$ and $X$ here may have different names in other cases.

in mind that this is a function of p; we can remove any constants that depend only on $x$ and $n$):

$$\begin{aligned} f_{P|X}(p|X) &\propto \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \\ &\propto p^x \cdot (1-p)^{1-x} \\ &= p^{(x+1)-1} \cdot (1-p)^{(n-x+1)-1} \end{aligned}$$

The last step may seem unusual, but if you look at the distribution table it becomes more clear. This is a Beta distribution, with $\alpha = (x+1)$ and $\beta = (n-x+1)$. So, the posterior is given by:

$$P|X \sim Beta(x+1, n-x+1).$$

This new distribution represents our knowledge and uncertainty about the parameter $P$.[2]

While the entire distribution is the clearest picture of our knowledge of $P$, sometimes we need to convert our knowledge into a single best guess point estimator. The **Bayesian point estimator** is the expected value of the posterior distribution. So, using the table, here we have:

$$\hat{p}_{Bayes} = \mathbb{E}[P|X] = \frac{x+1}{(x+1)+(n-x+1)} = \frac{x+1}{n+2}.$$

Notice that this limits the MLE in the limit of large $n$. Similarly, we can represent a version of a confidence interval for Bayesian statistics. A **credible interval** with credibility $1-\alpha$ for the parameter $P$ can be constructed by finding fixed values $l$ and $u$ such that:

$$\mathbb{P}[l \leq P \leq u] = 1-\alpha.$$

We can compute these values using R, as demonstrated in the notebook for today.

There are many methodological and philosophical implications of using Bayesian methods in place of the frequentist techniques that we have so far used. I hope that we can discuss these more in the classes to come after we have some more experience with the mechanics of how to work with them computationally.

[2] I have used a capital $P$ to stress that the parameter is now a random variable. While I think this is more clear, and I will use it in the Binomial case in the worksheet, note that this is not a standard notation. Usually we just use the same letter we have used throughout, with the change from constant to random variable being implicit (and sometimes confusing before you get the hang of it).