CSE 312: Foundations of Computing II
Instructor: Alex Tsun
Date: 2/25/22
Lecture Topics: 7.5 Maximum a Posteriori Estimation

[**Tags:** MLE**,** MAP, Beta]
1. In PSet2, you implemented a Naive Bayes classifier as a spam filter. We estimated the quantities $P(Y = y)$ and $P(\text{word} \mid Y = y)$ for $y \in \{\text{spam}, \text{ham}\}$ and all words. Explain in a few sentences which estimation techniques we used (MLE or MAP) for each, and make sure to include which distribution's parameter we were estimating, and the prior distribution if we had one.

[**Tags:** MAP, Beta]
2. Suppose $x = (x_1, \dots, x_n)$ are iid samples from $Geo(\Theta)$ where Θ is a random variable (not fixed).
   a. Using the prior $\Theta \sim Beta(\alpha, \beta)$ (for some arbitrary but known parameters $\alpha, \beta \geq 1$), show that the posterior distribution $\Theta|x$ also follows a Beta distribution and identify its parameters (by computing $\pi_\Theta(\theta|x)$). Then, explain this sentence: "The Beta distribution is the conjugate prior for the rate parameter of the Geometric distribution". Hint: This can be done in just a few lines!
   b. Now derive the MAP estimate for Θ. Recall the mode of $W \sim Beta(\gamma, \eta)$ is $\frac{\gamma - 1}{\gamma - 1 + \eta - 1}$ (pretend you saw $\gamma - 1$ heads and $\eta - 1$ tails). Hint: This should be just one line using your answer to part (a).
   c. Explain how this MAP estimate differs from the MLE/MoM estimate (recall for the Poisson distribution it was just the inverse of the sample mean $\frac{n}{\sum_{i=1}^{n} x_i}$) and provide an interpretation of α and β as to how they affect the estimate.