Lisa Yan and Jerry Cain
CS 109

Section #7
November 3-4, 2020

Section #7

Based on the work of many CS109 staffs

1 Warmups

1.1 Parameters and MLE

Suppose $x_1, \ldots, x_n$ are i.i.d. (independent and identically distributed) values sampled from some distribution with density function $f(x|\theta)$, where θ is unknown. Recall that the likelihood of the data is

$$L(\theta) = f(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Recall we solve an optimization problem to find $\hat{\theta}$ which maximizes $L(\theta)$, i.e., $\hat{\theta} = \arg\max_{\theta} L(\theta)$.

1. Write an expression for the log-likelihood, $LL(\theta) = \log L(\theta)$.

2. Why can we optimize $LL(\theta)$ rather than $L(\theta)$?

3. Why do we optimize $LL(\theta)$ rather than $L(\theta)$?

1.2 Beta

1. Suppose you have a coin where you have no prior belief on its true probability of heads p. How can you model this belief as a Beta distribution?

2. Suppose you have a coin which you believe is fair, with "strength" α. That is, pretend you've seen α heads and α tails. How can you model this belief as a Beta distribution?

3. Now suppose you take the coin from the previous part and flip it 10 times. You see 8 heads and 2 tails. How can you model your posterior belief of the coin's probability of heads?

1.3 Maximum A Posteriori

1. Intuitively, what is MAP? What problem is it trying to solve? How does it differ from MLE?

2. Given a 6-sided die (possibly unfair), you roll the die N times and observe the counts for each of the 6 outcomes as $n_1, ..., n_6$. What is the maximum a posteriori estimate of this distribution, using Laplace smoothing? Recall that the die rolls themselves follow a multinomial distribution.

2 The Honor Code

We have decided that automated tools should be used to identify if two submissions are suspiciously similar. (N.B. these tools are never used as a basis for community standards cases — those always require professional human opinion.) However, automated tools are never perfect. As active CS109 students, we would like to explore the chance of a false positive in automated tools. For this task, we will consider Breakout, a CS106A assignment where students implement Breakout.

This problem combines our knowledge of Beta, the Central Limit Theorem, and Maximum Likelihood Estimation. Exciting!

2.1 Beta Sum

What is the distribution of the sum of 100 i.i.d. Betas? Let X be the sum

$$X = \sum_{i=1}^{100} X_i \qquad \text{where each } X_i \sim \text{Beta}(a = 3, b = 4)$$

Note the variance of a Beta:

$$\text{Var}(X_i) = \frac{ab}{(a+b)^2(a+b+1)} \qquad \text{where } X_i \sim \text{Beta}(a, b)$$

2.2 Single Match

Say there are 1000 decision points when writing Breakout. Assume: Each decision point is binary. Each student makes all 1000 decisions. For each decision there is a probability p that a student takes the more popular choice. What is the probability distribution for the number of matching decisions (we are going to refer to this as the “score”)? If possible, could you approximate this probability?

2.3 Maximum Match

When we look at two assignments, the probability of a false match is exceedingly small. What would the max similarity score look like when we compare one student to 5000 historical breakout submissions? Let X_i be the similarity score between a student who worked on their own and student i. Let Y be the highest match score between the student and all other submissions:

$$Y = \max_i X_i$$

The Central Limit Theorem tells us about the distribution of the sum of IID random variables. A more obscure theorem, the Fisher-Tippett-Gnedenko theorem, tells us about the *max* of IID random variables. It says that the max of IID exponential or normal random variables will be a “Gumbel” random variable.

$$Y \sim \text{Gumbel}(\mu, \beta) \qquad \text{The max of IID vars}$$

$$f(Y = k) = \frac{1}{\beta} e^{-(z + e^{-z})} \text{ where } z = \frac{k - \mu}{\beta} \qquad \text{The Gumbel PDF}$$

You are given data of 1300 students’ max scores from three quarters (we believe they all worked independently): $y^{(1)} \dots y^{(1300)}$. Set up (but do not solve) simultaneous equations we could solve to find the values of μ and β.