CS109 Lecture 19 Worksheet: Maximum Likelihood Estimation

Parameters, Likelihood, Log-Likelihood, and the MLE Recipe

Summer 2026

Lecture 19 Worksheet: Maximum Likelihood Estimation

Topics: parametric models and θ, the likelihood of i.i.d. data, log-likelihood, the calculus recipe for the MLE, cases where calculus fails, and MLE point estimates vs. Bayesian belief distributions.

How to use this sheet. Work in small groups. A TA and the instructor will circulate to help. We will go over selected solutions on the board. You do not need to finish every part; aim for understanding over speed. Keep fractions exact where you can.

Problem 1 (Review of Lecture 18): Quick Entropy

(a) Compute the entropy of a uniform distribution over 4 values.

(b) A yes/no question splits those 4 equally likely values into two groups of 2. What is the expected remaining entropy after asking it, and how many bits of information did the question gain?

Problem 2: MLE for a Geometric

Your CI system re-runs a flaky deploy until it succeeds. Each attempt independently succeeds with unknown probability p, so the number of attempts for one deploy is $X \sim \text{Geo}(p)$ with PMF $P(X = x) = (1-p)^{x-1}p$ for $x = 1, 2, \ldots$. You observe 5 independent deploys needing

$$[2,\ 1,\ 4,\ 1,\ 2]$$

attempts.

(a) Write the likelihood $L(p)$ of the data as a product, then simplify it into the form $p^a(1-p)^b$.

(b) Write the log-likelihood $LL(p)$.

(c) Differentiate, set to 0, and solve for $\hat{p}$ in terms of n and $\sum x_i$. Then compute $\hat{p}$ for this data.

(d) Your formula should look like a very natural estimator for a geometric. In one sentence: what is it doing?

Problem 3: MLE for an Exponential

The time between requests at your API is modeled as $X \sim \text{Exp}(\lambda)$ with density $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. You observe 4 independent gaps (in hours):

$$[0.8,\ 1.4,\ 0.2,\ 0.6]$$

(a) Write the likelihood and log-likelihood of the data.

(b) Solve for $\hat{\lambda}$, and compute it for this data.

(c) How does $\hat{\lambda}$ relate to the sample mean? Why does that make sense for a rate parameter?

Problem 4: When Calculus Fails

You have i.i.d. samples [1.2, 3.7, 2.1] from $X \sim \text{Uni}(0, \beta)$ with unknown β; the density is $f(x) = \frac{1}{\beta}$ for $0 \leq x \leq \beta$ and 0 otherwise.

(a) Write the likelihood $L(\beta)$. Be careful: for which values of β is it nonzero?

(b) Try the calculus recipe (differentiate the log-likelihood and set it to 0). What goes wrong?

(c) Find $\hat{\beta}$ by reasoning directly about where $L(\beta)$ is maximized.

(d) Is $\hat{\beta}$ an overestimate or underestimate of the true β? Answer in one sentence.

Problem 5: Wind Farm (pset6: mle_wind)

The speed of the wind at a windfarm is a random variable that varies as a *Rayleigh distribution*, parameterized by a single scale parameter θ with probability density function

$$f_X(x) = \begin{cases} \frac{x}{\theta} e^{-x^2/2\theta} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

We collect N independent measurements of wind speeds $w_1, w_2, \ldots, w_N$. Derive an equation for the maximum likelihood estimate of θ. Then use your equation to estimate θ for 10 observed speeds:

```
[7.55, 8.15, 8.91, 1.17, 6.77, 3.03, 8.43, 5.56, 3.26, 2.55]
```

Give your answer to three decimal places.

Problem 6: Point Estimate vs. Belief Distribution

You ship a new feature flag to 9 users; it works for 6 and fails for 3. Model each user as an i.i.d. $\text{Bern}(p)$.

(a) What is the MLE $\hat{p}$? (You may use the result derived in lecture.)

(b) A Bayesian classmate starts from the Laplace prior $p \sim \text{Beta}(2, 2)$. What is their posterior, and what is its mean?

(c) The two answers disagree. In a sentence or two: what does each number *mean*, and why does the Bayesian answer sit closer to $\frac{1}{2}$?

(d) With 600 successes out of 900 users, what happens to the disagreement? Why?

Challenge Problem (optional): Negative Binomial (pset6: mle_neg_bin)

The Negative Binomial random variable is a generalization of the Geometric random variable and counts **the number of experiments until r successes**, where each experiment is independent with probability of success p. The PMF is

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}.$$

Given that $r = 5$ and 100 samples of X, use MLE to estimate the parameter p.

```
X = [9, 23, 19, 10, 21, 27, 6, 10, 12, 16, 10, 20, 18,
     12, 13, 9, 7, 23, 17, 23, 7, 11, 19, 15, 16, 14,
     20, 10, 23, 9, 18, 10, 10, 11, 14, 19, 13, 7, 15,
     17, 10, 19, 14, 10, 14, 16, 17, 15, 19, 14, 17,
     16, 11, 17, 22, 12, 10, 9, 22, 11, 6, 14, 12, 17,
     15, 15, 7, 23, 9, 12, 17, 21, 10, 17, 18, 7, 15,
     23, 8, 6, 12, 18, 7, 25, 11, 16, 9, 11, 13, 30, 28,
     14, 10, 11, 15, 11, 11, 14, 16, 18]
```