CS109 Lecture 19 Worksheet: ANSWER KEY

Maximum Likelihood Estimation

Summer 2026

Lecture 19 Worksheet: Answer Key

Instructor / TA copy. Problem-set solutions are hidden in the standard build; run `./convert-to-pdf.sh 19 solutions` *for the full instructor/TA edition.*

Problem 1 (Review): Quick Entropy

(a) $H = \log_2 4 = 2$ bits.

(b) Each answer has probability $\frac{1}{2}$ and leaves 2 equally likely values ($H = 1$ bit), so expected remaining entropy $= \frac{1}{2}(1) + \frac{1}{2}(1) = 1$ bit. Information gained: $2 - 1 = 1$ bit.

Problem 2: MLE for a Geometric

(a) With $n = 5$ and $\sum x_i = 10$:

$$L(p) = \prod_{i=1}^{5} (1-p)^{x_i - 1} p = p^5 (1-p)^{\sum x_i - 5} = p^5 (1-p)^5.$$

(b)

$$LL(p) = 5 \ln p + 5 \ln(1-p) \qquad \text{(in general, } n \ln p + (\textstyle\sum x_i - n) \ln(1-p)\text{)}.$$

(c)

$$\frac{d\,LL}{dp} = \frac{n}{p} - \frac{\sum x_i - n}{1-p} = 0 \implies n(1-p) = p\,(\textstyle\sum x_i - n) \implies \hat{p} = \frac{n}{\sum x_i}.$$

Here $\hat{p} = \frac{5}{10} = 0.5$.

(d) $\hat{p}$ is (total successes)/(total attempts): every deploy is exactly one success, and $\sum x_i$ counts all attempts — the natural frequency estimate.

Problem 3: MLE for an Exponential

(a)

$$L(\lambda) = \prod_{i=1}^{4} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}, \qquad LL(\lambda) = n \ln \lambda - \lambda \sum x_i.$$

(b)

$$\frac{d\,LL}{d\lambda} = \frac{n}{\lambda} - \sum x_i = 0 \implies \hat{\lambda} = \frac{n}{\sum x_i} = \frac{4}{3.0} \approx 1.33 \text{ requests/hour.}$$

(c) $\hat{\lambda} = 1/\bar{x}$: the rate is the reciprocal of the average gap. If requests are 0.75 hours apart on average, they arrive at $\frac{4}{3}$ per hour — and indeed $E[X] = 1/\lambda$ for an exponential.

Problem 4: When Calculus Fails

(a) Each sample contributes $\frac{1}{\beta}$ *provided* $\beta \geq x_i$ (otherwise the density is 0). So

$$L(\beta) = \begin{cases} \frac{1}{\beta^3} & \beta \geq \max_i x_i = 3.7 \\ 0 & \text{otherwise.} \end{cases}$$

(b) $LL(\beta) = -3 \ln \beta$ gives $\frac{d\,LL}{d\beta} = -\frac{3}{\beta}$, which is never 0. The likelihood has no interior critical point — it is strictly decreasing in β — and the real action is at the *boundary* of the allowed region, which calculus alone doesn't see.

(c) $L(\beta)$ is 0 for $\beta < 3.7$ and strictly decreasing for $\beta \geq 3.7$, so it is maximized at the smallest allowed value:

$$\hat{\beta} = \max_i x_i = 3.7.$$

(d) Underestimate: every sample is at most β, so $\max_i x_i \leq \beta$ always (with equality having probability 0 for continuous data).

Problem 5: Wind Farm (pset6: mle_wind)

(Problem set problem)

Problem 6: Point Estimate vs. Belief Distribution

(a) MLE for a Bernoulli is the sample mean: $\hat{p} = \frac{6}{9} = \frac{2}{3} \approx 0.667$.

(b) $p \mid \text{data} \sim \text{Beta}(2+6,\ 2+3) = \text{Beta}(8, 5)$, with mean $\frac{8}{13} \approx 0.615$.

(c) The MLE is a single *point* — the parameter value that makes the observed data most probable — with no statement of confidence. The posterior is a full *distribution over beliefs* about p; the Laplace prior acts like 2 imaginary successes and 2 imaginary failures, pulling the estimate toward $\frac{1}{2}$ when data is scarce.

(d) MLE: $\frac{600}{900} = 0.667$; posterior mean: $\frac{602}{904} \approx 0.666$. With lots of data the likelihood swamps the prior and the two approaches agree — the prior's 4 pseudo-observations are a drop in the bucket.

Challenge Problem: Negative Binomial (pset6: mle_neg_bin)

(Problem set problem)