

CS109 Lecture 18 Worksheet: ANSWER KEY

Information Theory

Summer 2026

Lecture 18 Worksheet: Answer Key

Instructor / TA copy. Problem-set solutions are hidden in the standard build; run `./convert-to-pdf.sh 18 solutions` for the full instructor/TA edition.

Problem 1 (Review): A Recursive Expectation

Condition on the Bernoulli:

$$E[T] = \frac{1}{2}(4) + \frac{1}{2}(2 + E[T]) = 3 + \frac{1}{2}E[T] \implies E[T] = 6.$$

Problem 2: Surprise

- (a) $I = \log_2 52 \approx 5.70$ bits. (The second identical draw adds another $\log_2 52$; see part (c).)
- (b) $I = \log_2 \frac{1}{1} = 0$ bits: a certain event carries no surprise, exactly as a measure of surprise should.
- (c) By independence, $P(E \cap F) = P(E)P(F)$, so

$$I(E \cap F) = \log_2 \frac{1}{P(E)P(F)} = \log_2 \frac{1}{P(E)} + \log_2 \frac{1}{P(F)} = I(E) + I(F).$$

The log is the (essentially unique) monotonic choice that makes surprise *add* across independent events, which matches intuition: two independent shocks are twice the news.

Problem 3: Entropy

- (a) $H(\text{Bern}(0.5)) = 0.5 \log_2 2 + 0.5 \log_2 2 = 1$ bit.

$$H(\text{Bern}(0.9)) = -(0.9 \log_2 0.9 + 0.1 \log_2 0.1) \approx 0.469 \text{ bits.}$$

The fair coin is more uncertain — matching intuition, since the biased coin's outcome is fairly predictable.

- (b)

$$H = \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + \frac{1}{8} \log_2 8 + \frac{1}{8} \log_2 8 = \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = 1.75 \text{ bits.}$$

- (c) The uniform distribution: $H = \log_2 8 = 3$ bits. Uncertainty is maximized when no outcome is favored — every observation is equally surprising, so expected surprise is as large as possible.

Problem 4: Which Diagnostic Test Is Better?

- (a) **Test A.** $P(\text{yes}) = \frac{1}{2}$: cause identified, remaining entropy 0. $P(\text{no}) = \frac{1}{2}$: conditional distribution over {backend, db, network} is $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, with entropy $\frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{4}(2) = 1.5$ bits.

Expected remaining entropy = $\frac{1}{2}(0) + \frac{1}{2}(1.5) = 0.75$ bits (gain $1.75 - 0.75 = 1.0$ bit).

- (b) **Test B.** $P(\text{yes}) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$: conditional distribution $(\frac{2}{3}, \frac{1}{3})$, entropy ≈ 0.918 bits. $P(\text{no}) = \frac{1}{4}$: conditional distribution $(\frac{1}{2}, \frac{1}{2})$, entropy 1 bit.

Expected remaining entropy = $\frac{3}{4}(0.918) + \frac{1}{4}(1) \approx 0.939$ bits (gain ≈ 0.811 bits).

- (c) Test A is better (expected remaining entropy $0.75 < 0.939$), even though its “no” answer leaves more uncertainty than nothing — note the contrast with the tempting “split the outcomes in half” heuristic, which ignores the prior. Before running the test the answer is random, so the remaining entropy is a random variable; the principled score for a test is its *expectation*.

Problem 5: Two Dice, Two Hints (pset5: info_theory_dice)

(Problem set problem)