CS109 Lecture 17 Worksheet: ANSWER KEY

Algorithmic Analysis

Summer 2026

Lecture 17 Worksheet: Answer Key

Instructor / TA copy. Problem-set solutions are hidden in the standard build; run `./convert-to-pdf.sh 17 solutions` for the full instructor/TA edition.

Problem 1 (Review): Reading a Bootstrap

(a) $p = \frac{140}{10{,}000} = 0.014$.

(b) If the UI change made no difference, only about 1.4% of experiments would show a gap this large by chance — reasonably strong evidence against the null hypothesis (at the conventional 0.05 bar, we would reject it).

Problem 2: Conditional Expectation from a Joint Table

(a) $P(Y = 1) = 0.30$ and $P(Y = 2) = 0.70$. Conditioning renormalizes each row:

$$E[X \mid Y = 1] = \frac{0(0.15) + 1(0.10) + 2(0.05)}{0.30} = \frac{0.20}{0.30} = \frac{2}{3},$$

$$E[X \mid Y = 2] = \frac{0(0.10) + 1(0.30) + 2(0.30)}{0.70} = \frac{0.90}{0.70} = \frac{9}{7} \approx 1.29.$$

(b) $E[X] = E[X \mid Y = 1]P(Y = 1) + E[X \mid Y = 2]P(Y = 2) = \frac{2}{3}(0.3) + \frac{9}{7}(0.7) = 0.2 + 0.9 = 1.1$.

(c) Marginal: $P(X = 0) = 0.25$, $P(X = 1) = 0.40$, $P(X = 2) = 0.35$, so $E[X] = 0.40 + 2(0.35) = 1.1$. ✓

Problem 3: Expected Runtime with a Cache Hierarchy

(a) Let L be the location. By the law of total expectation,

$$E[T] = 0.5(1) + 0.35(40) + 0.15(300) = 0.5 + 14 + 45 = 59.5 \text{ ms.}$$

(b) Given a browser-cache miss (probability 0.5), renormalize:

$$E[T \mid \text{miss}] = \frac{0.35(40) + 0.15(300)}{0.5} = \frac{59}{0.5} = 118 \text{ ms.}$$

(c) The distribution is highly bimodal: half of requests take 1 ms, but 15% of users wait 300 ms — five times the "average." Expectation hides the tail, and the tail is what users complain about. (Any answer noting the spread/tail of the distribution is fine.)

Problem 4: Roll Until Big (pset5: code_analysis_dice)

(Problem set problem)

Problem 5: Analyzing Recursive Code

Condition on x:

$$E[Y] = \tfrac{1}{4}(2) + \tfrac{1}{4}\big(1 + E[Y]\big) + \tfrac{2}{4}\big(3 + E[Y]\big)$$
$$E[Y] = 0.5 + 0.25 + 0.25E[Y] + 1.5 + 0.5E[Y] = 2.25 + 0.75E[Y]$$
$$0.25E[Y] = 2.25 \implies E[Y] = 9.$$

Problem 6: Hash Table Collisions

(a) Let B_i indicate that bucket i is empty. A bucket is empty iff all 20 keys avoid it:

$$E[B_i] = P(B_i) = \left(\frac{9}{10}\right)^{20}, \qquad E[X] = \sum_{i=1}^{10} E[B_i] = 10\left(\frac{9}{10}\right)^{20} \approx 1.22.$$

(b) Let I_{jk} indicate that keys j and k collide; $P(I_{jk}) = \frac{1}{10}$ (whatever bucket j lands in, k matches it with probability $\frac{1}{10}$). There are $\binom{20}{2} = 190$ pairs:

$$E[C] = \binom{20}{2} \cdot \frac{1}{10} = 19.$$

(c) Linearity of expectation holds *regardless of dependence*: $E[\sum X_i] = \sum E[X_i]$ needs no independence. (Independence would only matter for, e.g., the variance.)

Problem 7: Complete the Set

(a) Once you own i heroes, each pack is a "success" (new hero) with probability $p_i = \frac{8-i}{8}$, independently. So $X_i \sim \text{Geo}\big(p = \frac{8-i}{8}\big)$ and $E[X_i] = \frac{8}{8-i}$.

(b)

$$E[X] = \sum_{i=0}^{7} \frac{8}{8-i} = 8\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{8}\right) = 8 \cdot \frac{761}{280} = \frac{761}{35} \approx 21.7 \text{ packs.}$$

(c)

$$\sum_{i=0}^{3} \frac{8}{8-i} = 1 + \frac{8}{7} + \frac{8}{6} + \frac{8}{5} \approx 5.08 \text{ packs.}$$

Early on, nearly every pack is new, so the first half of the collection is cheap; the expected cost is dominated by the *last few* heroes (the final missing hero alone costs 8 packs in expectation).

Challenge Problem: Llama-Flu (pset5: code_analysis_flu)

(Problem set problem)