

CS 574: Randomized Algorithms, Fall 2026

Homework 3

Released: Sunday, October 18, 2026

Due: Tuesday, November 3, 2026 at 11:59pm

Instructions & Course Policies:

- **Collaboration:** You may work in groups of up to three, but each student must write their solutions independently, in their own words. List all collaborators and external resources on the first page of your submission.
- **AI and other resources:** You are encouraged to use AI agents to help format your solutions, and may also consult them about the mathematics. Declare all textbooks, papers, websites, and AI tools consulted, and explain how you used them, in a footnote or bibliography.
- **Submission:** Submit a single typeset L^AT_EX PDF on Gradescope. Start each problem you submit on a new page.
- **Clarity and rigor:** Define your random variables and justify each use of independence, conditioning, and a tail inequality. You may use a preceding subpart's stated conclusion even if you have not proved it. Hints are suggestions, not additional requirements.
- **Review parts:** A part marked **Review** revisits an algorithm, result, or proof already developed in the lecture notes, possibly with minor variations. Its purpose is to check understanding: reconstruct the argument in your own words and justify the requested details; a new approach is not expected. You may consult and cite the notes, but citing a result alone does not replace a requested proof. Review parts carry the stated points and are required within the problem you choose. A **Review and extension** label identifies a part that also asks for an additional application or example.
- **Grading:** There are four problems, each worth 100 points. **You have to solve only one problem of your choice**, for a total of 100 points. Think about all four: they are fair game for the midterm.

Conventions. Write $[r] = \{1, \dots, r\}$. All logarithms are natural unless a base is displayed. Fresh random choices are independent of previous choices unless stated otherwise. A stream is fixed independently of the algorithm's random choices; repeated items occupy distinct stream positions. A word can store an item identifier, a stream position, or a counter of the size used in the problem.

Lecture coverage. This assignment continues after Homework 2, covering lectures 13–16 (October 7–16): martingales and Azuma's inequality, the power of two choices, streaming algorithms, and frequency estimation. Background is in the corresponding course notes and in the standard texts [MR95, MU17]; the lecture schedule is <https://courses.grainger.illinois.edu/cs574/fa2026/lectures/>.

Problem 1: Exposure Martingales

[100 points]

You may use the following form of Azuma's inequality. If $(M_i)_{i=0}^r$ is a martingale with respect to the revealed history $\mathcal{F}_i$, M_0 is constant, and $|M_i - M_{i-1}| \leq c_i$ for deterministic constants c_i , then

$$\mathbb{P}[|M_r - M_0| \geq u] \leq 2 \exp \left(-\frac{u^2}{2 \sum_{i=1}^r c_i^2} \right) \quad (u > 0).$$

Treat a deterministic random variable separately when the denominator is zero.

(A) [20 points] **Review: A useful exposure principle.**

Let $X_1, \dots, X_r$ be independent random variables with finite ranges, and let $Z = f(X_1, \dots, X_r)$. Suppose changing only coordinate i changes f by at most c_i . Define $M_i = \mathbb{E}[Z \mid X_1, \dots, X_i]$, with $M_0 = \mathbb{E}[Z]$. Prove that (M_i) is a martingale and that $|M_i - M_{i-1}| \leq c_i$. Deduce a concentration bound for Z . Identify the role of independence in the increment bound.

Hint. Fix a prefix of length $i - 1$. Compare two possible values of X_i by giving both conditional completions the same random suffix.

(B) [25 points] **A better exposure for matchings.** Let G be a random simple graph on $[n]$, $n \geq 2$, in which each edge $\{u, v\}$ is present independently with a prescribed probability p_{uv} . Let $\nu(G)$ be the maximum number of edges in a matching. First expose all $\binom{n}{2}$ possible edges separately and obtain a tail bound using (A). Next partition the edge indicators into n blocks, where block i contains the edges $\{j, i\}$ with $j < i$. Prove the bound

$$\mathbb{P}[|\nu(G) - \mathbb{E}[\nu(G)]| \geq u] \leq 2e^{-u^2/(2n)}.$$

Explain why changing an entire block changes ν by at most one, although the block can contain many edges. Compare the deviation scales obtained by the two exposures for failure probability δ .

Hint. For two graphs differing only on edges incident to one vertex, delete that vertex from both graphs.

(C) [20 points] **Dependent samples.** Fix numbers $a_1, \dots, a_N \in [0, 1]$, where $N \geq 2$, and let $\mu = N^{-1} \sum_j a_j$. Reveal a uniformly random permutation of these numbers as $X_1, \dots, X_N$; the underlying positions are distinct even if some values coincide. For a fixed $1 \leq k < N$, set

$S_i = \sum_{j=1}^i X_j$ and $M_i = \mathbb{E}[S_k \mid X_1, \dots, X_i]$ for $0 \leq i \leq k$. Derive an explicit formula for M_i and show that

$$M_i - M_{i-1} = \frac{N-k}{N-i} \left(X_i - \frac{N\mu - S_{i-1}}{N-i+1} \right).$$

Why can we not simply apply (A) to $X_1, \dots, X_k$?

- (D) **[20 points] Use the smaller side of a sample.** Prove that for $1 \leq k < N$ and $u > 0$,

$$\mathbb{P}[|S_k - k\mu| \geq u] \leq 2 \exp \left(-\frac{u^2}{2 \min\{k, N-k\}} \right).$$

State the resulting bound for the sample mean S_k/k and explain what happens when $k = N$. No independence of the draws may be assumed.

Hint. First use the increments in (C), each of absolute value at most one. If k is large, expose the omitted positions instead.

- (E) **[15 points] Stopping does not preserve every estimate.** For any integrable martingale $(Z_i)_{i=0}^r$, let $\tau \in \{0, \dots, r\}$ be a stopping time: the decision whether $\tau \leq i$ uses only the history through time i . Prove directly that $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0]$. Now take the population $\{0, 1\}$ in random order. Stop after one draw if $X_1 = 1$, and otherwise take both draws. Compute $\mathbb{E}[S_\tau/\tau]$. Explain why the two conclusions do not conflict, and why the fixed- k guarantee in (D) cannot simply be asserted after substituting a data-dependent τ for k .

Hint. Write $Z_\tau - Z_0$ as $\sum_{i=1}^r (Z_i - Z_{i-1}) \mathbf{1}_{\{\tau \geq i\}}$. Which history determines the indicator?

Problem 2: The Power of Two Choices

[100 points]

Place n balls sequentially into n initially empty bins. Each ball samples two uniform bins independently, with replacement, and is put in a sampled bin of minimum current load. Break ties by a fixed rule. All probes of different balls are independent. Let $B_i(t)$ be the number of bins with load at least i after t insertions, and let H_t be the height of ball t , namely the load of its chosen bin immediately after insertion. Write $B_i = B_i(n)$ and let $\mathcal{F}_t$ record the first t pairs of probes.

- (A) [15 points] **Chernoff without independent indicators.** Let $Y_1, \dots, Y_n \in \{0, 1\}$ be determined by the respective histories $\mathcal{F}_1, \dots, \mathcal{F}_n$, and suppose $\mathbb{P}[Y_t = 1 \mid \mathcal{F}_{t-1}] \leq p$ for every possible history, where $0 \leq p \leq 1$. For $S = \sum_t Y_t$, prove

$$\mathbb{E}[e^{\lambda S}] \leq (1 - p + pe^\lambda)^n \leq \exp(np(e^\lambda - 1)) \quad (\lambda \geq 0).$$

Deduce that $\mathbb{P}[S \geq b] \leq e^{-b/6}$ whenever $np \leq b/2$. Also show, for an integer $1 \leq q \leq n$, that

$$\mathbb{P}[S \geq q] \leq \binom{n}{q} p^q.$$

Hint. For the exponential bound, condition on the past before removing the last factor. Take $\lambda = \ln 2$ for the tail bound. For the last claim, union bound over q specified successful times, bounding their joint probability by conditioning backwards.

- (B) [25 points] **Indicators controlled by the past.** Fix a level i and a deterministic real number $0 \leq b \leq n$. Define

$$Y_t = \mathbf{1}_{\{B_i(t-1) \leq b\}} \mathbf{1}_{\{H_t \geq i+1\}}, \quad S = \sum_{t=1}^n Y_t.$$

Prove that $\mathbb{P}[Y_t = 1 \mid \mathcal{F}_{t-1}] \leq (b/n)^2$. Show that, on the event $B_i \leq b$, we have $B_{i+1} \leq S$. Deduce a bound on $\mathbb{P}[\{B_i \leq b\} \cap \{B_{i+1} > b'\}]$ when $n(b/n)^2 \leq b'/2$. Explain what is wrong with simply conditioning the original probes on the future event $B_i \leq b$ and then declaring them independent uniform probes.

Hint. Bin loads never decrease. Given the actual past, a ball has height at least $i+1$ exactly when both probes hit bins already at load at least i .

- (C) [20 points] **Review: The layers shrink doubly exponentially.** Assume n is sufficiently large that $a = 24 \ln n < n/4$. Define

$$b_4 = n/4, \quad b_{i+1} = \max\{2b_i^2/n, a\},$$

and stop the recurrence at the first index L with $b_L = a$. Prove $L \leq \log_2 \ln n + O(1)$ and

$$\mathbb{P}[B_i \leq b_i \text{ for every } 4 \leq i \leq L] \geq 1 - n^{-3}.$$

Explain how an additive constant in the final load bound can cover the finitely many smaller n .

Hint. Before the cutoff, put $x_i = 2b_i/n$, so $x_{i+1} = x_i^2$ and $x_4 = 1/2$. Bound the probability of the first failed layer using (B), rather than multiplying conditional estimates.

- (D) **[25 points] Review: Finish after the logarithmic cutoff.** Show that

$$\mathbb{P}[B_{L+1} > 2] \leq n^{-3} + \frac{a^6}{6n^3}, \quad \mathbb{P}[B_{L+2} > 0] \leq n^{-3} + \frac{a^6}{6n^3} + \frac{4}{n}.$$

Conclude that maximum load is at most $\log_2 \ln n + O(1)$ with probability $1 - O(1/n)$. Make the event inclusions explicit; do not assume the indicators of high placements are independent.

Hint. At level L , use the guarded indicators with $b = a$ and the $q = 3$ bound in (A). At level $L + 1$, use $b = 2$ and $q = 1$.

- (E) **[15 points] Review and extension: More choices, and fake choices.** For fixed $d \geq 2$, adapt the proof to d independent uniform probes: show maximum load at most $\log_d \ln n + O_d(1)$ with probability $1 - O_d(1/n)$. Give the new recurrence and both final error bounds; a sketch suffices. If instead each ball uses two identical uniform probes, independently across balls, what process results and which probability bound fails?

Hint. Use $p = (b/n)^d$ and $b_{i+1} = \max\{2n(b_i/n)^d, a\}$. To solve the uncapped recurrence, scale b_i/n by $2^{1/(d-1)}$.

Problem 3: Learning from a Stream

[100 points]

Part (A) concerns candidates in uniformly random order. In (B)–(E), the stream $x_1, \dots, x_m$ is an arbitrary fixed nonempty sequence of items from a finite universe U , and its length is not known in advance. Let f_v be the number of occurrences of item v . You may use the following consequence of Hoeffding's inequality: for independent $Z_1, \dots, Z_s \in \{0, 1\}$ with $\bar{p} = s^{-1} \sum_j \mathbb{E}[Z_j]$,

$$\mathbb{P} \left[\left| s^{-1} \sum_j Z_j - \bar{p} \right| \geq \eta \right] \leq 2e^{-2s\eta^2}.$$

Its one-sided version has no leading factor of two.

(A) [15 points] **Review: Choose once: the secretary problem.**

There are $N \geq 2$ candidates of distinct quality, arriving in uniformly random order; N is known. Reject the first r candidates, where $1 \leq r < N$, and then accept the first candidate better than all previously seen candidates. If none appears, accept nobody. Derive the exact probability of accepting the best candidate. For $r/N \rightarrow \alpha \in (0, 1)$ as $N \rightarrow \infty$, find its limit and the choice of α maximizing that limit. Explain why the same guarantee need not hold for an adversarial arrival order.

Hint. Condition on the position j of the best candidate. Where must the best among the preceding $j - 1$ candidates have appeared?

(B) [20 points] **Review: Keep a sample without knowing the length.** Design a one-pass algorithm maintaining s independent uniform samples from the positions seen so far, with replacement, using $O(s)$ words. Prove their joint distribution after every fixed nonempty prefix. State the update time. Why is proving that each sample is individually uniform not enough for parts (C) and (D)?

Hint. Use s copies of a one-item reservoir, each with its own fresh random choices. At position t , replace a copy's item with probability $1/t$. Samples may store the same position.

(C) [20 points] **Review: Approximate a median.** Suppose U is totally ordered. Fix $0 < \varepsilon, \delta < 1/2$, and take an odd number s of samples from (B). Return their median z . Prove that, with probability at least $1 - \delta$,

$$|\{t : x_t < z\}| \leq (1/2 + \varepsilon)m, \quad |\{t : x_t > z\}| \leq (1/2 + \varepsilon)m$$

when $s \geq (2\varepsilon^2)^{-1} \ln(2/\delta)$. Your proof must handle repeated values. Give space, per-item update time, and final processing time for a simple implementation.

Hint. Break equal-value ties by stream position for analysis only. If the median is too high, at least half the samples lie in a fixed upper tail containing fewer than $(1/2 - \varepsilon)m$ positions. Use a symmetric argument for a median that is too low.

- (D) **[20 points] Review: Find every heavy hitter.** Fix $0 < \theta, \delta < 1/2$. An item is a heavy hitter if $f_v \geq \theta m$. Using (B), output a candidate set of size at most s that contains every heavy hitter with probability at least $1 - \delta$, for

$$s = \left\lceil \frac{1}{\theta} \ln \frac{1}{\theta\delta} \right\rceil.$$

Bound the number of heavy hitters, and account for all of them in your failure probability. If a second pass is allowed, explain how to remove every false positive without introducing additional random error. State a simple time and space bound for this second pass.

- (E) **[25 points] Review: Remove the randomness.** Fix an integer $k \geq 2$. Maintain at most $k - 1$ item-counter pairs, initially none. On an arriving item v : if it has a counter, increment it; otherwise, if there is space, insert $(v, 1)$; otherwise decrement all $k - 1$ counters, remove zeros, and do not insert v . Let D be the number of decrement-all steps and c_v the final counter for v (zero if absent). Prove

$$D \leq m/k, \quad 0 \leq f_v - c_v \leq D \quad \text{for every } v.$$

Conclude that all items with $f_v > m/k$ survive. Choose k so that all items with $f_v \geq \theta m$ survive, including the equality case. Compare the space and error guarantees with (D), and describe how a second pass gives the exact heavy-hitter set.

Hint. Interpret a decrement-all step as canceling k occurrences with distinct values, including the arriving item. Distinguish $f_v > m/k$ from $f_v \geq m/k$.

Problem 4: Estimating Frequency Moments

[100 points]

Let $x_1, \dots, x_m \in [U]$ be a fixed nonempty insertion-only stream, where $U \geq 2$. Define $f_v = |\{t : x_t = v\}|$ and $F_k = \sum_{v=1}^U f_v^k$ for positive integers k . The length m is not known in advance. In (C)–(E), you may assume a family of functions $\sigma : [U] \rightarrow \{-1, +1\}$ whose values on any set of at most four distinct keys are independent unbiased signs. A sampled function can be represented in $O(\log U)$ bits and evaluated in $O(1)$ word operations; you do not need to construct this family. Different copies use independent functions. Assume $0 < \varepsilon, \delta < 1/2$.

- (A) [15 points] **Review: A suffix estimator.** Choose a stream position J uniformly from $[m]$, and let $R = |\{t \geq J : x_t = x_J\}|$ count occurrences of its value in the suffix, including position J . Prove that

$$X_k = m(R^k - (R-1)^k) \quad \text{satisfies} \quad \mathbb{E}[X_k] = F_k.$$

Give a one-pass implementation using a one-item reservoir and a counter. Specify the update order and what happens if a replacement selects the same value as the old representative.

Hint. For an item of frequency f , the possible suffix counts at its occurrences are exactly $f, f-1, \dots, 1$. A replacement selects a new *position*, even when its value is unchanged.

- (B) [20 points] **Unbiased need not mean low variance.** For $k = 2$, derive the exact second moment

$$\mathbb{E}[X_2^2] = \frac{m}{3}(4F_3 - F_1).$$

Now take an integer $q \geq 2$ and a stream of length $m = q^2$ with one value of frequency q and $q^2 - q$ other values of frequency one (assume the universe is large enough). Show that

$$\frac{\text{Var}(X_2)}{F_2^2} = \Theta(q) = \Theta(\sqrt{m}).$$

Explain what this implies for the number of independent copies needed by the usual variance-plus-Chebyshev analysis of an averaged estimator. Do not claim a lower bound for all streaming algorithms.

Hint. Use $\sum_{r=1}^f (2r-1)^2 = (4f^3 - f)/3$. Here $F_1 = q^2$, $F_2 = 2q^2 - q$, and $F_3 = q^3 + q^2 - q$.

- (C) [25 points] **Review: Random signs remove the length dependence.** Maintain $Z = \sum_{t=1}^m \sigma(x_t) = \sum_v f_v \sigma(v)$ and return $Y = Z^2$. Prove

$$\mathbb{E}[Y] = F_2, \quad \mathbb{E}[Y^2] = 3F_2^2 - 2F_4, \quad \text{Var}(Y) \leq 2F_2^2.$$

State precisely where pairwise independence suffices and where four-wise independence is used. Give the state size in bits and the update time for one copy.

Hint. Expand Z^4 and group ordered quadruples of keys by their multiplicities. Which products have a key appearing an odd number of times? How many orderings have multiplicities two and two?

- (D) [25 points] **Review and extension: Average, then take a median.** In each of b independent groups, average $a = \lceil 8/\varepsilon^2 \rceil$ independent copies of Y . Return the median of the b group averages, where b is odd. Give an explicit sufficient $b = O(\ln(1/\delta))$ and prove

$$\mathbb{P}[|\widehat{F}_2 - F_2| > \varepsilon F_2] \leq \delta.$$

Give total space in bits and update time. Explain why independent groups are required, and why pairwise independence of signs inside one copy is insufficient for the preceding variance guarantee. For the latter point, analyze the following example: let $U = 2^d$, index keys by $v \in \{0, 1\}^d$, choose independent uniform $A \in \{0, 1\}^d$ and $B \in \{0, 1\}$, and put $\sigma(v) = (-1)^{A \cdot v + B}$, with the dot product modulo two. Let the stream contain every key once.

Hint. Chebyshev makes each group bad with probability at most $1/4$. Apply Hoeffding to the bad-group indicators. In the example, pair terms in $\sum_v (-1)^{A \cdot v}$ when $A \neq 0$.

- (E) [15 points] **Merge streams before squaring.** Two machines process streams with frequency vectors f and g . Describe how to combine their signed counters to estimate $\sum_v (f_v + g_v)^2$ with the same guarantee as in (D), without replaying either stream. Specify exactly which random seeds must agree and which must remain independent. Why is adding the two final estimates of F_2 generally wrong? What goes wrong if the machines use independent sign functions and add their counters?

Hint. The counter is linear in the frequency vector; its square is not. Test a stream where both machines see the same single key.

References

- [MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
- [MU17] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. 2nd ed. Cambridge University Press, 2017.