

CS 574: Randomized Algorithms

Midterm 2 — Practice Solutions

Lectures 8–16 and review of earlier material

This practice exam is intended for students. Allow 60 minutes for 20 questions. Choose exactly one answer per question; each correct answer earns 1 point. No calculator is needed. You may consult the reference sheet if useful.

Question 1 A treap has distinct keys and priorities. A leaf is deleted, and nothing else is changed. Are rotations needed to restore the BST and min-heap invariants?

- A. Yes: every deletion changes the root.
- B. Yes: the remaining priorities must be resampled.
- C. No: removing a leaf preserves all remaining key and priority comparisons.
- D. Only if the deleted key was larger than its parent.
- E. Only if the deleted priority was larger than its parent's.

Correct answer: C. Removing a leaf removes constraints but changes no remaining parent-child relation. Both invariants therefore continue to hold.

Why the alternatives fail: No new relation needs repair, on either side of the parent. Priorities need not be resampled, and the root changes only when the tree consisted of that single leaf.

Question 2 Draw two labels independently and uniformly from $\{1, \dots, N\}$, where $N \geq 2$. If they are equal, discard both and repeat with fresh randomness until they differ. What is the distribution of the returned pair?

- A. Uniform over ordered pairs of distinct labels.
- B. Two independent uniform labels, still allowing equal labels.
- C. Uniform over pairs whose first label is smaller than the second.
- D. Always a pair containing label 1.
- E. Biased toward pairs of consecutive numerical labels.

Correct answer: A. Every ordered distinct pair has the same probability in each trial, and the same acceptance condition. Conditioning on acceptance leaves a uniform distribution over these pairs.

Why the alternatives fail: The returned labels cannot be equal and are not independent. Rejection does not distinguish increasing from decreasing pairs, special labels, or numerical distances.

Question 3 An incremental algorithm uses $\Theta(n)$ ordinary work in total. You know only that its expected number of rebuilds is $O(\log n)$ and each rebuild costs at most n . Which conclusion follows from these facts alone?

- A. Its total expected work is necessarily $O(\log n)$.
- B. Its total expected work is necessarily $O(n)$.
- C. It has $O(n)$ worst-case work because rebuilding is rare.
- D. It has $O(n \log n)$ expected work; a linear bound needs more information about rebuild costs and locations.
- E. No expected-work bound follows unless the rebuild events are independent.

Correct answer: D. If K counts rebuilds, their total cost is at most nK , so its expectation is at most $n\mathbb{E}K = O(n \log n)$. For example, about $\log n$ late rebuilds costing order n each are consistent with these facts and rule out a general linear conclusion.

Why the alternatives fail: Counting rebuilds is not the same as weighting their costs. The assumptions give neither a linear expected bound nor a worst-case bound; independence is not needed for the displayed inequality.

Question 4 A coloring of a ground set U has total signed sum zero. For a subset $A \subseteq U$, what is the relation between the imbalances of A and its complement $U \setminus A$?

- A. Both signed sums must be zero.
- B. Their absolute signed sums are equal.
- C. Their signed sums are independent random variables.
- D. The larger set must have the larger absolute signed sum.
- E. Their absolute signed sums add to zero.

Correct answer: B. The two signed sums add to the total sum, zero. They are negatives of each other and have the same absolute value, even if neither is zero.

Why the alternatives fail: Overall balance does not imply balance in each part. Complementary sums are constrained, not independent, and set size alone does not decide the imbalance.

Question 5 A simple graph on $n \geq 4$ vertices is formed by choosing its edges independently. Let Z be the number of triangles, and expose the edge indicators one at a time. What is the smallest uniform bound on the change in Z caused by changing a single edge indicator, over all configurations of the other edges?

- A. 1, because an edge indicator is a bit.
- B. $\binom{n}{3}$, because any triangle may change.
- C. 0, because the edge choices are independent.
- D. No finite bound exists, because triangle indicators are dependent.
- E. $n - 2$, because an affected triangle must use that edge and one third vertex.

Correct answer: E. Only triangles containing the changed edge can be affected. There are at most $n - 2$ choices for their third vertex, and the bound is attained when both endpoints are adjacent to all other vertices. This bounds the corresponding exposure-martingale increment under independent inputs.

Why the alternatives fail: Input values being bits does not make the output sensitivity one. Triangles not containing the changed edge are unaffected, and dependence between triangle indicators does not prevent a coordinate-sensitivity argument.

Question 6 Two stream blocks have lengths $a, b > 0$. Each block has a stored sample uniform over its own occurrence positions. To obtain a uniform position sample from the concatenation, choose one block and return its stored sample. How should the first block be chosen? Make this choice independently of the stored samples.

- A. With probability $1/2$, regardless of the lengths.
- B. With probability $b/(a + b)$.
- C. With probability $a/(a + b)$.
- D. Always choose the longer block.
- E. Choose the block whose stored value occurs earlier alphabetically.

Correct answer: C. A position in the first block is returned with probability $(a/(a + b))(1/a) = 1/(a + b)$. The same calculation holds for the second block. Weighting by block length gives uniformity over occurrences.

Why the alternatives fail: Equal block probabilities work only for equal lengths. Reversing the weights or choosing by length or stored value deterministically biases the position sample.

Question 7 A procedure computes K estimates. For every index fixed before sampling, its probability of being inaccurate is at most η . An index is then selected after examining the estimates, and that estimate is reported. Assume $K \geq 2$ and $0 < \eta < 1/K$. What failure bound follows for the reported estimate without assumptions on how the index is selected?

- A. At most $K\eta$, because an inaccurate reported estimate requires at least one inaccurate estimate.
- B. At most η , because the selected index is an index.
- C. At most η/K , because selection averages the errors.
- D. At most η^K , because all estimates must be inaccurate.
- E. Zero, because selection can always recognize errors from the estimates alone.

Correct answer: A. Whatever index is selected, failure is contained in the union of the K failure events. The union bound gives $K\eta$. Selecting after seeing results can favor errors, so a fixed-index guarantee does not by itself remain η .

Why the alternatives fail: Selection is not averaging, and one inaccurate selected estimate suffices. Nothing guarantees that errors are detectable or that error events are independent.

Question 8 A graph is a disjoint union of two components. Select a vertex if it precedes all its neighbors in an ordering. Change how the two components are interleaved, but preserve the relative vertex order within each component. What happens to the selected set?

- A. Its size must increase.
- B. Every vertex in the component placed first is selected.
- C. The selected sets in the two components are exchanged.
- D. It stays the same, because all relevant comparisons are within components.
- E. It may contain adjacent vertices after the interleaving changes.

Correct answer: D. A vertex has neighbors only in its own component. Their order relative to that vertex is preserved, so its eligibility is unchanged.

Why the alternatives fail: Interleaving vertices from another component adds no neighbor comparisons. It therefore neither changes eligibility nor breaks independence of the selected vertex set.

Question 9 Fix a configuration in which a fraction p of bins have load at least h . An insertion first samples one uniform bin. If it has load below h , place the ball there. Otherwise sample a second bin independently and uniformly, and use the less loaded of the two. With replacement and $0 < p < 1$, what are the expected number of probes and the probability of placing the ball at height at least $h + 1$?

- A. 2 probes and probability p^2 .
- B. $1 + p$ probes and probability p^2 .
- C. $1 + p$ probes and probability p .
- D. $1 + p^2$ probes and probability p^2 .
- E. 1 probe and probability $2p - p^2$.

Correct answer: B. The first probe is always made and the second is made exactly when the first bin is heavy, an event of probability p . A high placement requires both probes to hit heavy bins, giving p^2 . This preserves this particular threshold probability, not necessarily the entire ordinary two-choice load distribution.

Why the alternatives fail: Two probes are not always needed. The second-probe event concerns the first sample only, while a high placement requires two heavy samples. At least one heavy sample is insufficient.

Question 10 A randomized test accepts every prime and accepts each fixed composite with probability at most $1/100$. The test accepts an input drawn from an unspecified distribution. Does this imply that the input is prime with probability at least $99/100$?

- A. Yes: conditional error probabilities can always be reversed.
- B. Yes: acceptance is a proof of primality.
- C. No: a randomized test cannot ever give useful information.
- D. Yes: the distribution of inputs cannot matter.
- E. No: the probability of primality after acceptance also depends on the input distribution.

Correct answer: E. The supplied guarantee concerns acceptance conditional on a composite input. A posterior probability also depends on how often primes and composites are supplied. For example, a distribution containing only composites can still produce an acceptance if the test has a positive false-acceptance probability.

Why the alternatives fail: A likelihood is not a posterior probability, and an acceptance need not certify primality. This does not make the test useless; it identifies the missing prior information.

Question 11 You want to split a treap into keys below q and keys above q ; assume q is not a stored key. At the current root $r < q$, which part may contain keys on both sides of q and therefore needs further examination?

- A. Only the left subtree; the entire right subtree is above q .
- B. Both subtrees, because priorities determine the key order.
- C. The right subtree; the root and its entire left subtree belong below q .
- D. Neither subtree; knowing $r < q$ puts every key below q .
- E. Only the subtree whose root has smaller priority.

Correct answer: C. Every key in the left subtree is below r , hence below q . Right-subtree keys exceed r but can lie on either side of q . The split follows key comparisons; heap priorities do not change these inequalities.

Why the alternatives fail: The right subtree need not be entirely above q , nor need the whole tree be below it. Priority comparisons organize the shape, not the partition by a search key.

Question 12 M_i and N_i are martingales with respect to the same revealed history. They may be dependent. What can be said about $M_i + N_i$?

- A. It is a martingale by linearity of conditional expectation.
- B. It is a martingale only if the two processes are independent.
- C. It must be nondecreasing.
- D. It must have zero variance at every time.
- E. It is a martingale only if both processes start at zero.

Correct answer: A. Conditional linearity gives $\mathbb{E}[M_i + N_i \mid \mathcal{F}_{i-1}] = M_{i-1} + N_{i-1}$. Integrability and determination by the current history are also preserved under addition. Independence and zero initial values are unnecessary.

Why the alternatives fail: The martingale condition is linear. It imposes neither monotonicity nor zero variance, and dependence between the two processes does not invalidate linearity.

Question 13 Assign independent fair signs $\sigma_1, \dots, \sigma_n$. For m constraints let $Y_j = \sum_i a_{ji}\sigma_i$, with fixed real coefficients. Suppose $\sum_i a_{ji}^2 \leq V$ for every j , where $V > 0$. Which bound is valid for every $t > 0$? You may consult the reference sheet at the end.

- A. $\Pr(\max_j |Y_j| \geq t) \leq \min\{1, 2 \exp(-mt^2/(2V))\}$.
- B. $\Pr(\max_j |Y_j| \geq t) \leq \min\{1, 2m \exp(-Vt^2/2)\}$.
- C. $\Pr(\max_j |Y_j| \geq t) \leq \min\{1, 2m \exp(-t^2/2)\}$, independently of the coefficients.
- D. $\Pr(\max_j |Y_j| \geq t) \leq \min\{1, 2m \exp(-t^2/(2V))\}$.
- E. No union bound is available because the constraints share signs.

Correct answer: D. The range length of $a_{ji}\sigma_i$ is $2|a_{ji}|$. The sum of squared range lengths is at most $4V$, so one constraint has two-sided tail at most $2 \exp(-t^2/(2V))$. A union bound handles the m constraints without independence between their sums.

Why the alternatives fail: More constraints cannot improve the union guarantee as in the first expression. Larger coefficient squares weaken, rather than strengthen, the scale of concentration. Ignoring coefficients is not valid, and shared signs do not prevent a union bound.

Question 14 A routing implementation chooses one uniform random vertex H , then sends every packet through H . Each packet individually has a uniform intermediate vertex. Why does this fail to preserve the usual argument based on independent random intermediates?

- A. A uniform vertex cannot be represented by a binary string.
- B. The common choice correlates all routes and can concentrate traffic at one hub.
- C. Every route now has zero length.
- D. Uniform marginal choices automatically imply independence, so the argument is unchanged.
- E. Packets no longer have final destinations.

Correct answer: B. All packets share the same random intermediate, so their intermediate choices are completely dependent. Uniform marginals alone do not spread the simultaneous traffic as independent choices do.

Why the alternatives fail: The network representation and final destinations are unchanged, and paths need not have zero length. The missing ingredient is joint independence, not marginal uniformity.

Question 15 Three balls independently choose uniform bins among $N \geq 2$ bins. Let E_{12}, E_{13}, E_{23} be the events that the indicated pair shares a bin. Which statement describes these three events?

- A. They are mutually independent.
- B. They are pairwise disjoint.
- C. If any one occurs, all three must occur.
- D. Every pair is dependent because it shares a ball.
- E. They are pairwise independent, but not mutually independent.

Correct answer: E. Each event has probability $1/N$. Any two together mean all three balls share a bin, with probability $1/N^2$, equal to the product of the two marginals. All three together still have probability $1/N^2$, not $1/N^3$. Thus pairwise independence holds but mutual independence fails.

Why the alternatives fail: The events can all occur together, but a single collision need not involve the third ball. Sharing a random input does not by itself prove dependence; here the pairwise products agree, while the triple product does not.

Question 16 Fix the two bins sampled by an insertion, and use a fixed rule for ties. Before choosing the less loaded bin, add the same positive integer c to the load of every bin. What changes for this insertion?

- A. The chosen bin is unchanged, and the new ball's height increases by c .
- B. The chosen bin is always exchanged with the other sampled bin.
- C. The chosen bin becomes the first bin in numerical order.
- D. The new height is unchanged.
- E. The two sampled bins must be sampled again.

Correct answer: A. Adding a common offset preserves all load comparisons and all ties. The same bin is chosen, but its load before and after insertion is larger by c .

Why the alternatives fail: No relative ordering changes, so there is no reason to exchange candidates or resample. The absolute height, unlike the choice, does change.

Question 17 Two nonempty streams are concatenated. Their frequencies for value x are a_x and b_x . Someone estimates the concatenation's second moment by adding the two separate second moments. Which statement about this sum is correct?

- A. It is always exactly the concatenation's second moment.
- B. It is always an upper bound, with equality only for identical streams.
- C. It is a lower bound, with equality exactly when no value occurs in both streams.
- D. It can be either larger or smaller, depending on arrival order.
- E. It is correct whenever the streams have the same length.

Correct answer: C. The true moment is $\sum_x (a_x + b_x)^2 = \sum_x a_x^2 + \sum_x b_x^2 + 2 \sum_x a_x b_x$. The missing cross term is nonnegative and is zero exactly when the supports are disjoint.

Why the alternatives fail: Second moments are not generally additive across concatenation. Arrival order and equal lengths do not remove cross terms from values present in both blocks.

Question 18 In an exponential-moment argument for a lower-tail event $S \leq a$, take $\lambda > 0$. Which event is equivalent to $S \leq a$ and is in the upper-tail form used by Markov's inequality?

- A. $e^{-\lambda S} \leq e^{-\lambda a}$.
- B. $e^{\lambda S} \geq e^{\lambda a}$.
- C. $S - a \geq 0$.
- D. $e^{-\lambda S} \geq e^{-\lambda a}$.
- E. $e^{-\lambda S} < 0$.

Correct answer: D. The map $s \mapsto e^{-\lambda s}$ is decreasing and positive. It reverses the inequality and turns the lower tail into an upper tail of a nonnegative variable.

Why the alternatives fail: Using the wrong direction or a positive exponent describes the opposite tail. The exponential cannot be negative.

Question 19 A random-order stream is processed in one pass. Store exact counts of all values in an initial prefix, then freeze these values as the candidate list and continue incrementing only their counters through the suffix. Suppose the prefix contains all final heavy hitters with probability at least $1 - \delta$. What can be reported at the end?

- A. Exact frequencies of every distinct value, including values first seen in the suffix.
- B. Exact frequencies for all candidates; with probability at least $1 - \delta$, all heavy hitters are among them.
- C. Only approximate frequencies for candidates, because the stream order is random.
- D. The heavy-hitter guarantee holds even if the prefix counters are discarded and the suffix counts are treated as full counts.
- E. Every candidate is necessarily a heavy hitter, so thresholding is unnecessary.

Correct answer: B. Each candidate counter includes all its prefix and suffix occurrences, so it is exact. The supplied random-order discovery guarantee controls whether any heavy hitter is missing. Exact thresholding can remove false-positive candidates, provided the final stream length is tracked.

Why the alternatives fail: Untracked values first appearing in the suffix need not have exact counts. Discarding the prefix counts undercounts candidates, while candidate discovery alone does not exclude false positives.

Question 20 A randomized search repeatedly proposes a candidate and verifies it. Each trial costs at most Cn time and either returns a correct answer or requests a restart. Conditional on any history of failures, the next trial succeeds with probability at least p , where $0 < p < 1$. Which guarantee follows?

- A. The expected time is $O(np)$ and error probability is p .
- B. The time is at most Cn/p in every execution.
- C. The probability of returning an incorrect answer is $1 - p$.
- D. No expected-time bound is possible without identically distributed trials.
- E. It returns a correct answer almost surely and has expected time $O(n/p)$.

Correct answer: E. The probability of failing the first k trials is at most $(1 - p)^k$, by conditioning successively on the history. The expected number of trials is therefore at most $1/p$, giving time at most Cn/p in expectation. The failure tail tends to zero, and verification prevents incorrect answers.

Why the alternatives fail: Small success probability increases the waiting time. An expectation is not a deterministic cutoff, and failed trials cause retries rather than incorrect output. The conditional success lower bound suffices without identical trial distributions.

CS 574: Randomized Algorithms

Midterm 2 reference sheet

Lectures 8–16

Probability and concentration

Expectation and variance

For an event A , $\mathbf{1}_A$ is 1 when A occurs and 0 otherwise:

$$\mathbb{E}[\mathbf{1}_A] = \Pr(A), \quad \mathbb{E}\left[\sum_i a_i X_i\right] = \sum_i a_i \mathbb{E}[X_i].$$

Linearity of expectation does not require independence. With finite second moments,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2, \quad \text{Var}(aX + b) = a^2 \text{Var}(X).$$

For pairwise independent variables, $\text{Var}(\sum_i X_i) = \sum_i \text{Var}(X_i)$. If r independent copies have mean μ and variance σ^2 , then

$$\bar{X} = \frac{1}{r} \sum_{j=1}^r X_j, \quad \mathbb{E}[\bar{X}] = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{r}.$$

Events and elementary bounds

For any events $A_1, \dots, A_m$,

$$\Pr\left(\bigcup_{j=1}^m A_j\right) \leq \sum_{j=1}^m \Pr(A_j), \quad \Pr\left(\bigcap_{j=1}^m A_j^c\right) \geq 1 - \sum_{j=1}^m \Pr(A_j).$$

No independence is required for these union bounds.

For independent events, $\Pr(A \cap B) = \Pr(A) \Pr(B)$. For $\Pr(B) > 0$, $\Pr(A \mid B) = \Pr(A \cap B) / \Pr(B)$.

$$1 - x \leq e^{-x} \quad (x \in \mathbb{R}), \quad e^{\ln x} = x \quad (x > 0).$$

Here $\ln$ is the natural logarithm. A probability bound larger than 1 may always be replaced by 1.

Markov and Chebyshev

For $X \geq 0$ with finite expectation and $t > 0$,

$$\Pr(X \geq t) \leq \frac{\mathbb{E}X}{t}.$$

For any X with finite variance and $t > 0$,

$$\Pr(|X - \mathbb{E}X| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$

These statements do not assume X is a sum of independent variables.

Multiplicative Chernoff bounds

Let $S = \sum_{i=1}^n X_i$ for independent Bernoulli variables $X_i \in \{0, 1\}$, and let $\mu = \mathbb{E}S > 0$. Their success probabilities may differ. For $0 < \varepsilon < 1$:

$$\Pr(S \geq (1 + \varepsilon)\mu) \leq e^{-\varepsilon^2 \mu / 3}, \quad \text{upper tail,}$$

$$\Pr(S \leq (1 - \varepsilon)\mu) \leq e^{-\varepsilon^2 \mu / 2}, \quad \text{lower tail,}$$

$$\Pr(|S - \mu| \geq \varepsilon \mu) \leq 2e^{-\varepsilon^2 \mu / 3}, \quad \text{two tails.}$$

The displayed independent-sum formulas are not justified by marginal distributions alone. If $\mu = 0$, then $S = 0$ almost surely.

CS 574: Randomized Algorithms

Midterm 2 reference sheet

Lectures 8–16

Exposure, sampling, and notation

Hoeffding for bounded independent variables

If independent $X_i \in [a_i, b_i]$ almost surely, $S = \sum_i X_i$, and $D = \sum_i (b_i - a_i)^2 > 0$, then for $t > 0$,

$$\Pr(|S - \mathbb{E}S| \geq t) \leq 2 \exp\left(-\frac{t^2}{D}\right).$$

If $D = 0$, the sum is constant. A fair sign takes values $-1, +1$ with equal probabilities and has mean 0.

Conditional expectation and martingales

Let $\mathcal{F}_i$ denote the information revealed through step i , with $\mathcal{F}_{i-1} \subseteq \mathcal{F}_i$. The tower property gives

$$\mathbb{E}[\mathbb{E}[Z \mid \mathcal{F}_i] \mid \mathcal{F}_{i-1}] = \mathbb{E}[Z \mid \mathcal{F}_{i-1}], \quad \mathbb{E}[\mathbb{E}[Z \mid \mathcal{F}_i]] = \mathbb{E}Z.$$

A martingale M_i is determined by $\mathcal{F}_i$, has finite absolute expectation, and satisfies $\mathbb{E}[M_i \mid \mathcal{F}_{i-1}] = M_{i-1}$. An exposure martingale is $M_i = \mathbb{E}[Z \mid \mathcal{F}_i]$; initially $M_0 = \mathbb{E}Z$ when no information is revealed, and $M_n = Z$ once Z is known.

Azuma for bounded martingale increments

If $|M_i - M_{i-1}| \leq c_i$ almost surely for deterministic bounds c_i , and $C = \sum_{i=1}^n c_i^2 > 0$, then for $t > 0$,

$$\Pr(|M_n - M_0| \geq t) \leq 2 \exp\left(-\frac{t^2}{2C}\right).$$

If $C = 0$, the martingale does not change. Its increments need not be independent. For independent inputs $X_1, \dots, X_n$, if changing only input i changes $f(X_1, \dots, X_n)$ by at most c_i , the exposure martingale for f has increments bounded by c_i .

Permutations, sampling, and stream notation

In a uniform permutation, each of n objects is equally likely to occupy any specified position. Sampling with replacement permits repeats; sampling without replacement produces distinct positions. Repeated values can still occur at distinct positions.

For a stream $(s_1, \dots, s_m)$ over $[n] = \{1, \dots, n\}$, write $f_i = |\{j : s_j = i\}|$. Its frequency moments are

$$F_0 = |\{i : f_i > 0\}|, \quad F_k = \sum_{i=1}^n f_i^k \quad (k \geq 1), \quad F_1 = m.$$

An unbiased estimator X of a quantity θ satisfies $\mathbb{E}X = \theta$; this alone does not bound its error probability.

Algorithmic conventions

A treap is a binary search tree on keys and a heap on priorities; this exam uses a min-heap. Depth counts edges from the root, whose depth is 0. An independent vertex set contains no edge between its vertices. A d -dimensional hypercube has binary strings of length d as vertices; an edge changes one bit.

For a coloring $\chi : U \rightarrow \{-1, +1\}$ and set family $\mathcal{S}$, its discrepancy is $\max_{A \in \mathcal{S}} |\sum_{x \in A} \chi(x)|$. Bin load is the number of balls in a bin; a ball's height is its chosen bin's load immediately after insertion.