Problem 1 (25 points) The class $\text{BPP}$ contains all languages $L$ for which there is a precise polynomial-time NTM $N$ satisfying $\text{prob}[N(x) = L(x)] \geq 3/4$ for all $x$. (That is, if $x \in L$, then at least $3/4$ of the computation paths of $N$ on $x$ lead to “yes.” Otherwise, at least $3/4$ of the computation paths of $N$ on $x$ lead to “no.”) Let $p\text{-BPP}$ denote the class of languages $L$ for which there is a precise polynomial-time NTM $N$ satisfying $\text{prob}[N(x) = L(x)] \geq 1/2 + 1/p(n)$ for all $x$, where $p(n)$ is a polynomial function and $n = |x|$. Prove that $\text{BPP} = p\text{-BPP}$. (Hint: Recall that one version of the Chernoff bound is that $\text{prob}[\sum_{i=1}^{k} x_i \leq k/2] \leq e^{-\epsilon^2 k/2}$ where $p = 1/2 + \epsilon$ for any $0 \leq \epsilon \leq 1/2$.)

Proof: It is obvious that $\text{BPP} \subseteq p\text{-BPP}$. We will show that $p\text{-BPP} \subseteq \text{BPP}$. For all languages $L$ in $p\text{-BPP}$, for every input $x \in L$, run machine $N$ for $k$ times. The Chernoff bound $\text{Pr}[\sum_{i=1}^{k} x_i \leq k/2] \leq e^{-\epsilon^2 k/2}$ implies that the probability of a false answer is at most $e^{-\epsilon^2 k/2}$. By taking $k = \lceil 2p(n)^2 \ln 4 \rceil$, the error probability is at most $1/4$. ■

Problem 2 (25 points) Consider the interactive proof system for GRAPH NONISOMORPHISM in the lecture. Suppose Bob’s behavior is the same except that he flips his answer in the end: he accepts if and only if the original protocol rejects. Is the resulting system a system for GRAPH ISOMORPHISM? Why?

Proof: To be a system for GRAPH ISOMORPHISM, it must satisfy (1) if $G, G'$ are isomorphic then accepted with extremely high probability, and (2) if they are not isomorphic, then reject with extremely high probability regardless of the prover. For (1), suppose $G, G'$ are isomorphic. Then Alice sees all graphs as the same, and she answers one always. Hence all the answers will be 1s. For (2), suppose $G, G'$ are not isomorphic. Then does Bob’s strategy prevent cheating? No, because Alice can cheat Bob by always answering 1s only! Hence, it is not a system for GRAPH ISOMORPHISM. ■

Problem 3 (25 points) Given three disjoint sets $A$, $B$, and $C$, each containing $n$ elements, and a ternary relation $T \subseteq A \times B \times C$, a tripartite matching is a set of triples in $T$, none of which has an element in common. The problem MAXIMUM TRIPARTITE MATCHING seeks the largest tripartite matching. There is an approximation algorithm for MAXIMUM TRIPARTITE MATCHING:

1: $M := \emptyset$
2: while there is a triple $(a, b, c)$ in $T$ such that $a \in A, b \in B, c \in C$ do
3: Add $(a, b, c)$ to $M$;
4: Delete $a, b, c$ from $A, B, C$, respectively;
5: \textbf{end while}
6: \textbf{return} $M$;

Prove that the above algorithm is a $2/3$-approximation algorithm. (That is, prove that
the approximation ratio of the above algorithm is $c(M(x))/\text{OPT}(x) \geq 1/3$.)

**Proof:** Let $M^*$ be a maximum tripartite matching. Since we can not add any triple
from $M$ to $M^*$ when the algorithm ends, each triple in $M^*$ must have at least one element
in common with some element in $M$. Because all triples in $M^*$ contain disjoint elements,
the size of $M^*$ is upper-bounded by the number of elements of all the triples in $M$. That
is, $|M^*| \leq 3|M|$. The approximation ratio is hence $|M|/|M^*| \geq 1/3$. ■

**Problem 4 (25 points)** Prove that a $k$-sat expression where literals in a clause (which
contains $k$ literals by definition) are distinct must be satisfiable if it has fewer than $2^k$
clauses. (Hint: Consider a random truth assignment that independently assigns \textbf{true}
to every variable with probability $1/2$. Calculate its expected number of unsatisfiable
clauses.)

**Proof:** Consider a random truth assignment that independently assigns \textbf{true}
to every variable with probability $1/2$. Let $X$ count the number of unsatisfiable clauses. It suffices
to prove $E[X] < 1$: prob[$X = 0$] = 0 implies $E[X] \geq 1$ because $X$ is integer-valued. Let

\[
\phi = C_1 \land C_2 \land \cdots \land C_m
\]

be such a $k$-sat expression with $m < 2^k$ clauses. Define

\[
X_i = \begin{cases} 
1, & \text{if } C_i \text{ is unsatisfied}, \\
0, & \text{otherwise}.
\end{cases}
\]

Clearly

\[
X = \sum_{i=1}^{m} X_i.
\]

Clearly,

\[
E[X_i] \leq 2^{-k}
\]
as $X_i$ is a disjunction of distinct literals.$^a$ Hence

\[
E[X] = \sum_{i=1}^{m} E[X_i] \leq m2^{-k} < 1,
\]
as desired. ■

$^a$If $x$ and $\bar{x}$ both appear in a literal, then the probability that $C_i$ is unsatisfiable is zero, hence the inequality.