Problem: Verifying Matrix Multiplication.

Input: A, B, C are n x n matrices

Output: AB ? C

We will show that a (randomized) verification can be done faster than the currently best known deterministic algorithms.

Algorithm

1. Choose a - O/1 vector \( \vec{r} = (r_1, r_2, \ldots, r_n) \in \{0, 1\}^n \)

2. Compute \((A(B\vec{r}))\)

3. Compute \(C\vec{r}\)

4. If \((A(B\vec{r})) \neq C\vec{r}\) then report \(AB \neq C\)
   else report \(AB = C\).

Note that algorithm runs in \(O(n^2)\) time.

(a) Step 1 takes \(O(n)\) time
   by choosing each \(r_i\) uniformly from \(\{0, 1\}\).

(b) Step 2 takes \(O(n^2)\) time.

\[
B\vec{r} = \begin{bmatrix} \vec{r} \end{bmatrix}_{n \times 1}
= \begin{bmatrix} 1 \end{bmatrix}_{n \times 1}
= \begin{bmatrix} \vec{r} \end{bmatrix}_{n \times 1}
\]

\[
(A(B\vec{r})) = \begin{bmatrix} A \end{bmatrix}_{n \times n}
\begin{bmatrix} \vec{r} \end{bmatrix}_{n \times 1}
= \begin{bmatrix} \vec{r} \end{bmatrix}_{n \times 1}
\]

3. Step 3 takes \(O(n^2)\) time

4. Step 4 takes \(O(n)\) time.
We will show that

Claim 1: If \( AB \neq C \) and if \( \bar{r} \) is chosen
uniformly at random from \( \{0,1\}^n \), then
\[
\Pr ( AB \bar{r} = C \bar{r} ) \leq \frac{1}{2}.
\]

Before proving this let us establish some simple
claims. Let us assume that matrices \( A, B, C \) are
all integer matrices (i.e. their entries are \( 0/1 \))
and computations are done mod 2.

Let \( D = AB - C \neq 0 \)

\[ \Rightarrow \text{If } AB \bar{r} = C \bar{r} \Rightarrow D \bar{r} = 0 \]

Since \( D \neq 0 \), it contains at least one non-zero
entry. WLOG let us assume that \( d_{11} \neq 0 \).

If \( D \bar{r} = 0 \Rightarrow \sum_{j=1}^n d_{1j} r_j = 0 \)

\[
\begin{pmatrix} \mathbf{D} & \mathbf{r} \end{pmatrix} = \begin{pmatrix} d_{11} \mathbf{r}_1 + d_{12} \mathbf{r}_2 + d_{13} \mathbf{r}_3 + \cdots + d_{1n} \mathbf{r}_n = 0 \\
\end{pmatrix}
\]

\[ \Rightarrow \mathbf{r}_1 = - \frac{\sum_{j=2}^n d_{1j} \mathbf{r}_j}{d_{11}} \quad (\mathbf{1}) \]

Suppose in Step 1, we have chosen each of
\( \mathbf{r}_i \)'s, \( i = n \) down to 2, except \( \mathbf{r}_1 \).

Then the RHS of Equation 1 is fixed,
and hence there is only one value of \( \mathbf{r}_1 \) (out \( \frac{1}{2} \))
that can satisfy (1).
Since \( r_1 \) is chosen independently, there is a 50\% chance of satisfying Equation 1. Here we have set all the variables \( e_i \) and \( r_1 \) and deferred the decision to set \( r_1 \).

To formalize this we have the following claim.

**Theorem [Law of Total Probability]**

Let \( E_1, E_2, \ldots, E_n \) be mutually disjoint events in the sample space \( \Omega \) and let 
\[
\bigcup_{i=1}^{n} E_i = \Omega.
\]

Then
\[
\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B | E_i) \Pr(E_i),
\]

**Pf:** Since events \( E_i \)'s are disjoint and cover the whole space then 
\[
B = \bigcup_{i=1}^{n} (B \cap E_i).
\]

Hence
\[
\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B | E_i) \Pr(E_i).
\]

By definition of conditional probability. \( \square \)
Now back to proving Claim 1. (page 22).

Consider all possible values for \((x_2, x_3, \ldots, x_n) \in \{0, 1\}^{n-1}\).

Now \(\Pr(AB \overline{r} = C \overline{r})\)

\[
= \sum_{(x_2, x_3, \ldots, x_n) \in \{0, 1\}^{n-1}} \Pr((AB \overline{r} = C \overline{r} \cap (x_2, x_3, \ldots, x_n) = (x_2, x_3, \ldots, x_n)))
\]

\[
= \sum \Pr\left(\left(\eta_i = \sum_{j=2}^{n} \frac{d_{ij} \cdot r_j}{d_{i1}}\right) \cap (x_2, x_3, \ldots, x_n) = (x_2, x_3, \ldots, x_n)\right)
\]

\[
= \sum \Pr(\eta_1 = \sum \ldots) \cdot \Pr((x_2, x_3, \ldots, x_n) = (x_2, x_3, \ldots, x_n))
\]

\[
\leq \sum \frac{1}{2} \Pr((x_2, x_3, \ldots, x_n) = (x_2, x_3, \ldots, x_n))
\]

\[
= \frac{1}{2}.
\]

Hence \(\Pr(AB \overline{r} = C \overline{r}) \leq \frac{1}{2}\) and thus the claim is proved. \(\Box\)

Success probability of only 50% seems pretty low. We can increase the success probability by repeatedly running the algo for different choices of \(\overline{r}\). If we ever find that \(AB \overline{r} \neq C \overline{r}\) then we are done. Otherwise, after \(k\) trials,
Here is an exercise from [MU05] book to emphasize this idea.

Input: Two fair coins and one biased coin where
\( \Pr(H) = \frac{2}{3}, \ Pr(T) = \frac{1}{3}. \)
We do not know a priori which is a biased coin.

Problem: Permute coins, and toss them one by one.
First two coins show up heads and last one is tails.
Q: \( \Pr[ \text{first coin is a biased coin}] = \) ?

Answer: Let \( E_i \) be the event that the \( i^{th} \) coin is biased one (\( i = 1, 2, 3 \)).

Let \( B \) be the event that the three tosses result in HHT.

Note that \( \Pr(E_i) = \frac{1}{3} \) for \( i = 1, 2, 3 \) as we have no a priori knowledge before permanently tossed.

\[
\Pr(B|E_1) = \frac{2/3 \cdot 1/2 \cdot 1/2}{HHT} = \frac{1}{6}
\]

\[
\Pr(B|E_2) = \frac{1/2 \cdot 2/3 \cdot 1/2}{HHT} = \frac{1}{6}
\]

\[
\Pr(B|E_3) = \frac{1/2 \cdot 1/2 \cdot 1/3}{HHT} = \frac{1}{12}
\]

Applying Bayes Law (Note that \( E_1, E_2 \) and \( E_3 \) are mutually disjoint)

\[
\Pr(E_1|B) = \frac{\Pr(E_1 \cap B)}{\Pr(E)} = \frac{\Pr(B|E_1) \Pr(E_1)}{\sum_{j=1}^{3} \Pr(B|E_j) \Pr(E_j)}
\]

\[
= \frac{1/6 \cdot 1/3}{1/6 \cdot 1/3 + 1/6 \cdot 1/3 + 1/2 \cdot 1/3}
\]

\[
= \frac{1/6 \cdot 1/3}{2/18 + 1/36} = \frac{2/36}{4/36 + 1/36} = \frac{2}{5}
\]
where one trial is independent of other trials, probability that we report that \( AB\overline{r} = C\overline{r} \) for each of runs is at most \( \left( \frac{1}{2} \right)^k \).

[This is same as tossing a fair coin \( k \)-times, and in each toss we obtain a Tail.]

A Related Question

* Suppose we test \( AB\overline{r} = C\overline{r} \) and obtain that \( AB\overline{r} = C\overline{r} \).

* How confident are we that when we run the algorithm again, that we obtain that \( AB\overline{r} = C\overline{r} \).

For this first look at Bayes theorem

\[ \sum_{i=1}^{n} \Pr(E_i | B) = \frac{\Pr(E_j \cap B)}{\Pr(B)} = \frac{\Pr(B | E_j) \Pr(E_j)}{\sum_{j=1}^{n} \Pr(B | E_j) \Pr(E_j)} \]

Assume that \( E_1, \ldots, E_n \) are mutually disjoint sets such that \( \bigcup_{i=1}^{n} E_i = E \).
Note that without the knowledge of tosses, \( \Pr \) that the first coin is biased is only \( \frac{1}{3} \). But when we know the outcomes of coin flips (HHT), then the probability that the first coin is biased becomes \( \frac{2}{5} \).

Back to Matrix Multiplication.

Let \( E \) be the event that \( AB \subseteq C \) and let \( B \) be the event that the test returns that the identity is correct.

Note: Since we do not know whether \( AB = C \) or \( AB \neq C \), we assume that \( \Pr(AB = C) = \Pr(AB \neq C) = \frac{1}{2} \) to begin with.

Hence \( \Pr(E) = \Pr(\overline{E}) = \frac{1}{2} \).

What is \( \Pr(B|E) \)?

\[
\Pr(B|E) = \frac{1}{2}. 
\]

By Bayes theorem

\[
\Pr(E|B) = \frac{\Pr(B|E)\Pr(E)}{\Pr(B|E)\Pr(E) + \Pr(B|\overline{E})\Pr(\overline{E})} = \frac{1/2}{1/2 + 1/4} = \frac{2/3}{3/2} = \frac{2}{3}. 
\]
Let us run the algorithm again and suppose the outcome is that the identity is correct.

After 1st run we have \( \Pr(E) \geq \frac{2}{3} \), \( \Pr(E) \leq \frac{1}{3} \)

Now let \( B = \) Event that the 2nd test returns that the identity is correct.

Since tests are independent, we have

\[
\Pr(B|E) = 1, \quad \Pr(B|\overline{E}) = \frac{1}{2}
\]

Applying Bayes' we have

\[
\Pr(E|B) = \frac{\Pr(B|E) \cdot \Pr(E)}{\Pr(B|E) \cdot \Pr(E) + \Pr(B|\overline{E}) \cdot \Pr(\overline{E})}
\]

\[
= \frac{1 \cdot \frac{2}{3}}{\frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{\frac{2}{3}}{\frac{7}{6}} = \frac{4}{5}
\]

In general if before running the test we have that

\[
\Pr(E) = \frac{2^i}{2^i + 1}
\]

and the test returns that the identity is correct (i.e. the event \( B \)) then

\[
\Pr(E|B) \geq \frac{\frac{2^i}{2^i + 1}}{\frac{2^i}{2^i + 1} + \frac{1}{2} \cdot \frac{1}{2^i + 1}} = 1 - \frac{1}{2^i + 1}
\]

E.g. after 100 calls, if the test returns that in each of the calls identity is correct then we are certain, \( 1 - \frac{1}{2^{100} + 1} \).