Ten Problems About Twenty-Five Horses

Stephen Morris, Newbury, Berkshire, England stephenmor@gmail.com
Richard Stong, Center for Communications Research, La Jolla, California, stong@ccrwest.org
Stan Wagon, Macalester College, St. Paul, Minnesota, wagon@macalester.edu

The following horse-racing problem has achieved the status of folklore, having apparently been used as a test question for Facebook job candidates [1].

**Problem 1.** Given 25 horses and a track that can accommodate five horses at a time, show how to arrange seven races so that you can, in all cases, determine the three fastest horses, in order.

It is understood that each horse always runs the distance in the same time and the horses have distinct speeds. You have no stopwatch, but can make deductions from the finishing order in the races. When deciding which horses to race, information obtained from previous races may be used. These conditions apply for all problems we present here. As often happens, the problem leads to many interesting variations and some unsolved problems.

Our statement above reveals the answer of seven; the more common version is “Find the fewest number of races allowing you to determine the top three, in order.” Such problems generally do not expect a proof of optimality, but that can be done here.

**Problem 2.** Show that in six races it is not possible to always determine the top three horses, even not in order.

Next we ask about what is possible in only six races when we want the fastest horses in order. The next two problems appeared in [2].

**Problem 3.** Devise an algorithm that uses six races and determines with certainty the top two horses, in order, over 30% of the time.

**Problem 4.** Devise an algorithm that uses six races and determines with certainty the top three horses in order over 5% of the time.

And now we relax the condition on order.

**Problem 5.** Devise an algorithm that uses six races and determines with certainty the top two horses (not in order) over 45% of the time.

**Problem 6.** Devise an algorithm that uses six races and determines with certainty the top three horses (not in order) over 30% of the time.

We can also relax the certainty condition, meaning that we will produce a ranking of horses that might be the
desired order, but we won't be certain that the order is correct. Problem 7 presents the result that with six races one can get the top two, not in order, a remarkable 91% of the time, more than double the probability of Problem 5. But in many of the cases one would not know whether the two horses chosen were really the top two.

**Problem 7.** Devise an algorithm that uses six races and determines the top two horses (not in order) not with certainty but over 91% of the time.

**Problem 8.** Devise an algorithm that uses six races and determines the top three horses (not in order) not with certainty but over 72% of the time.

And finally we return to the order condition, but continue with uncertainty.

**Problem 9.** Devise an algorithm that uses six races and determines the top two horses in order and not with certainty but over 84% of the time.

**Problem 10.** Devise an algorithm that uses six races and determines the top three horses in order and not with certainty, but over 62% of the time.

**Solutions**

**Notation.** Let $R_i$ denote the $i$th race and let $A, B, C, D, E$ always denote the five fastest horses, in order.

**Solution 1.** Run five heats of five different horses in each. In $R_6$ run the five winners and let $X, Y, Z$ be the top three in order. Then $X = A$. In $R_7$ place $Y, Z$, the two top finishers behind $X$ in $X$'s first race, and the runnerup to $Y$ in $Y$'s first race. Then the top two in $R_7$ are $B$ and $C$ in order.

**Solution 2.** After five races at most 20 horses have lost, so there remain five that have not lost. These must appear in $R_6$. But suppose the runnerup in $R_1$ is $B$. Then $B$ is not in $R_6$ and it cannot be discovered that $B$ is in the top three. Indeed, this establishes the stronger result that one cannot determine the top two horses, not in order, using six races.

**Solution 3.** Run five randomly chosen horses in $R_1$. Run the winner, $Y$, with four new horses in $R_2$. Run the $R_2$ winner $Z$ (possibly equal to $Y$) with four new horses in $R_3$; run the $R_3$ winner $X$ (possibly equal to $Z$) in $R_4$ and beyond until $X$ loses. Let $R_i$ ($i \geq 4$) be the first race that $X$ loses, let $V$ and $W$ be that race’s top two, in order; run $W$ against four new horses in all subsequent races. If $W$ never again loses then $V$ and $W$ are proved to be top two overall. If $i$ does not exist, then we learn only that $A = X$. If $W$ loses a race after $R_i$, then we have no certainty about any position. The algorithm succeeds with probability $\frac{2746}{8925}$, or about 30.8%.

To derive the probability for the preceding strategy, let $P_n \ (n \geq 2)$ be the probability that $A$ first occurs in $R_n$ and $B$ in $R_i$, $i \leq n$. Because $n$ is fixed, four of the 25 possible locations for $A$ contribute to $P_n$, while among the 24 remaining possibilities for $B, 5 + 4 + \ldots + 0 = 4n$ are good. So $P_n = \frac{4n}{25 \cdot 24} = \frac{2}{75} \ n$. The full strategy succeeds precisely when $A$ occurs first in $R_i$ with $i \in \{4, 5, 6\}, B$ occurs first no later than $A$, and it never happens that the repeating horse loses to a horse different than $A$. For example, using numbers as ranks, $(5, 20, 21, 22, 23), (2, 3, 4, 5, 6), (2, 10, 11, 12, 13), (1, 2, 7, 8, 9)$ is a successful scenario for the first four races since 2, the fastest in the first three races, loses $R_4$ to $A$, and then wins the rest. But $(21, 22, 23, 24, 25), (15, 16, 17, 18, 21), (3, 4, 5, 6, 15), (2, 3, 7, 8, 9), (1, 3, 10, 11, 12)$ is a failure, since even though 3 will win the remaining race after the loss to 1, we have no information with which to compare 1 and 2. Therefore the overall probability of success is $\sum_{i=4}^6 P_i Q_{i-1}$ where $Q_i$ is the probability that the repeat-
ing horse \((X)\) wins \(R_4\) through \(R_i\). This probability is simply \(\frac{13}{4i+1}\), the probability that the fastest horse among the first \(4i+1\) occurs in the first 13. So the total probability is \(\frac{213}{75} \sum_{i=3}^{6} \frac{i}{4i-3} = \frac{2746}{8925}\).

The choice of \(R_4\) as the first possible race to switch strategy was to optimize the result; when the same method is carried out with \(R_3\) in place of \(R_4\) the success probability is only 29.3%; other choices are worse.

**Solution 4.** Proceed as in Problem 3, running each winner in the next race until the repeater finishes either 3rd, 4th, or 5th in \(R_i\) with \(i \geq 3\). Let \(V, W, X\) be the top three, in order, of \(R_i\). Run \(X\) in all subsequent races. If \(X\) wins them all, then we will know that \((A, B, C) = (V, W, X)\). This succeeds with probability \(\frac{9487}{177905}\), or about 5.3%.

To derive the fraction, let \(P_i\) be the probability that \(A\) and \(B\) appear first in \(R_n\) and \(C\) appears first in \(R_i\) with \(i \leq n\); then \(P_i = \frac{4 \cdot 3 \cdot (4n-1)}{25 \cdot 24 \cdot 23} = \frac{4n-1}{1150}\). We need the probability \(Q_n^i\) that the repeating horse in \(R_i\) finishes first or second (i.e., places), for each \(i = 3, \ldots, n\). Observe first that the repeating horse in \(R_i\) is the fastest among those running \(R_1, R_2, \ldots, R_{i-1}\). The probability \(Q_i\) that the repeater does not place in \(R_i\) is the chance that \(R_i\) includes the two fastest among the horses running in the first \(i\) races; this is \(\frac{4 \cdot 3}{(4i+1)(4i)} = \frac{3}{8(i+1)}\). So the probability that the repeater places is \(1 - Q_i = \frac{(i+1)(4i-3)}{i(4i+1)}\). Therefore \(Q_n^i = \prod_{i=3}^{n} \frac{(i+1)(4i-3)}{i(4i+1)} = 3 \frac{n+1}{4n+1}\). Now the total probability of success is \(\sum_{i=3}^{6} P_i Q_n^{i-1} = \frac{3}{1150} \sum_{i=3}^{6} \frac{4(4i-1)}{4i-3} = \frac{9487}{177905}\).

**Solution 5.** Run each winner in the next race until the repeater does not win \(R_i\) with \(i \geq 3\). If the top two horses in \(R_i\) are \(X, Y\) in order, run \(Y\) through all remaining races. If \(Y\) wins them all (this includes the case \(i = 6\)), then \(X\) and \(Y\) are top two in order. If \(Y\) wins the rest of the races except one, in which it is second to \(W\), then \(X\) and \(W\) are the top two. This algorithm succeeds with probability \(\frac{30956}{68425}\), about 45.2%.

The algorithm fails unless there is a “key race” \(R_i, 3 \leq i \leq 6\), which is the first race that the repeating horse loses. If there is a key race, then the algorithm selects a “key horse”, namely \(Y\). At the time of \(Y\)’s selection we know that \(Y\) is faster than all horses that have been raced except \(X\), who is faster than \(Y\). Since \(Y\) will be raced against every remaining horse in the subsequent races, at the end of six races we will know how every horse compares to \(Y\). Thus we will know the two fastest horses (and the algorithm will succeed) if \(Y\) is \(B\) or \(C\). Explicitly, if \(Y = B\), then the two fastest horses will be \(Y\) and \(X\). If \(Y = C\), then the fastest two horses will be the two horses faster than \(Y\). It is also not hard to see that these are the only cases where the algorithm succeeds. By construction, \(Y\) has lost to at least one horse \(X\) so we cannot have \(Y = A\). If \(Y\) is \(D\) or worse, then we will have at least three horses faster than \(Y\) and since \(X\) cannot be compared to any of the others we cannot know the fastest two. Thus computing the probability of success is just computing the probability that \(Y\) is \(B\) or \(C\). We split this computation into three cases.

Suppose \(B\) wins \(R_2\). Since \(A\) is not in the first two races with probability \(\frac{16}{25}\) and, given this, \(B\) is in one of those two races with probability \(\frac{9}{24}\), this occurs with probability \(\frac{6}{25}\). In this case, the repeater will be \(B\) until he encounters horse \(A\). This will be the key race and \(B\), as the second fastest horse in this race, will become the key horse, resulting in a success.

Suppose \(C\) wins \(R_2\). This occurs with probability \(\frac{16}{25} \cdot \frac{15}{24} \cdot \frac{9}{23} = \frac{18}{115}\). In this case, \(C\) will be the repeater until he encounters \(A\) or \(B\) or both. This will be the key race and either \(C\) or \(B\) will be the runnerup in this race.
Thus the key horse will be \( B \) or \( C \) and the algorithm will succeed.

Suppose none of \( A, B, \) or \( C \) occurs in the first two races. Let the first race where one of them occurs be \( R_n, 3 \leq n \leq 6 \). For the method to succeed, the fastest horse in \( R_2 \) (out of 9 horses) must be the fastest in the first \( n - 1 \) races (out of \( 4n - 3 \) horses). This occurs with probability \( \frac{9}{4n-3} \). Also we must have two of \( A, B, C \) occur in \( R_n \) so that the runnerup will be \( B \) or \( C \) and the third of them must occur in \( R_n \) or later. This occurs with probability \( \frac{3 \cdot 4 \cdot (24 - 4n) + 4 \cdot 3 \cdot 2}{25 \cdot 24 \cdot 23} \). The first term in the numerator corresponds to two of the three horses racing in \( R_n \). There are three choices for which two, four choices of which spot in \( R_n \) goes to the first of these two, three choices for which goes to the second and \( 24 - 4n \) choices of a spot in the remaining races for the third. The second term corresponds to all three racing in \( R_n \). Thus in this case the algorithm succeeds with probability

\[
\sum_{n=3}^{6} \frac{9}{4n-3} \frac{3 \cdot 4 \cdot (24 - 4n) + 4 \cdot 3 \cdot 2}{25 \cdot 24 \cdot 23} = \frac{3824}{68425}.
\]

Combining the three cases we see that the algorithm succeeds with probability

\[
\frac{6}{25} + \frac{18}{115} + \frac{3824}{68425} = \frac{30956}{68425}.
\]

**Solution 6.** Let \( V \) win \( R_1 \) and run \( V \) into future races until \( V \) finishes second or worse in \( R_i (2 \leq i \leq 6) \), with \( X, Y, Z \) the top three, in order, in \( R_i \). Race \( R_i \) is the key race. We choose one of the horses in this race to be the key horse depending on \( i \) and the results. If either \( i \) is 2 or 3 or \( Y = V \), then we choose the key horse to be \( Y \). If \( i \) is 4, 5, or 6 and \( Y \neq V \), then we choose the key horse to be \( Z \). We race the key horse through all the remaining races against all remaining horses. If there is no key race, then the algorithm simply fails.

At the time of the selection of the key horse, we know how every previously raced horse compares to him. Since the key horse will be raced against every remaining horse in subsequent races, at the end of six races we will know how every horse compares to the key horse. If the key horse turns out to be \( C \) or \( D \), then we will know the three fastest horses. Conversely, it is not hard to see that if the key horse is any other horse then the algorithm will fail. For example, if the key horse turns out to be \( B \), then \( C \) is one of the horses who lost only to the key horse (and possibly to \( A \)) but we cannot know which one. (There are at least four such horses since in this case either the key horse is \( V \) or \( i \leq 3 \).)

Note that a key race will occur provided \( V \neq A \) and the key horse will always be at least as fast as \( V \).

Thus computing the probability of success requires just computing the probability that \( Y \) is \( C \) or \( D \). We split this computation into three cases.

Suppose \( C \) wins \( R_1 \). This occurs with probability \( \frac{20}{25} \cdot \frac{19}{24} \cdot \frac{5}{23} = \frac{19}{138} \). Then \( V = C \) and a key race will occur. The algorithm will succeed unless the key race is \( R_2 \) or \( R_3 \) and involves both \( A \) and \( B \). Thus the algorithm succeeds following this case with probability \( \frac{19}{138} \left( 1 - \frac{4 \cdot 3 \cdot 4 \cdot 3}{20 \cdot 19} \right) = \frac{89}{690} \), since there are \( 20 \cdot 19 \) remaining choices for spots to place \( A \) and \( B \) in and each of \( R_2 \) and \( R_3 \) gives 12 choices that lead to a failure of the algorithm.

Suppose \( D \) wins \( R_1 \). This occurs with probability \( \frac{20}{25} \cdot \frac{19}{24} \cdot \frac{18}{23} \cdot \frac{5}{22} = \frac{57}{506} \). Then \( V = D \) and a key race will occur as soon as one of \( A, B, \) or \( C \) races. The algorithm will succeed unless the key race is \( R_2 \) or \( R_3 \) and involves both \( A \) and \( B \). The probability that \( R_2 \) involves both \( A \) and \( B \) (and hence is key) given that \( D \) won \( R_1 \) is \( \frac{4}{20} \cdot \frac{3}{19} = \frac{3}{95} \).
Similarly the chance that \( R_3 \) is key and involves both \( A \) and \( B \) given that \( D \) won \( R_1 \) is \( \frac{4}{20} \cdot \frac{3}{19} \cdot \frac{14}{18} = \frac{7}{285} \). (The last term in the numerator comes since \( C \) must not be in \( R_2 \), otherwise \( R_2 \) would be key.) Thus the final probability for following this case is \( \frac{57}{506} \left( 1 - \frac{3}{95} - \frac{7}{285} \right) = \frac{269}{2530} \).

Suppose none of the horses \( A, B, C, D \) race in \( R_1 \). Suppose the first race that involves one of these four is \( R_n \), \( (2 \leq n \leq 6) \). For the algorithm to succeed we need the winner \( V \) of \( R_1 \) to be the fastest horse in the first \( n - 1 \) races, an event that occurs with probability \( \frac{5}{4n-3} \). Given this we still need \( R_n \) to involve several of the four fastest horses. If \( n \leq 3 \), then we need \( R_n \) to involve at least two of them, but not both \( A \) and \( B \). If \( n \geq 4 \), then we need \( R_n \) to involve at least three of the four. In either case, all remaining members of \( A, B, C, D \) must occur in later races. For \( n \leq 3 \), there are five pairs \( ((A, C), (A, D), (B, C), (B, D), \text{and} (C, D)) \) that could occur in \( R_n \) and two triples \( ((A, C, D) \text{and} (B, C, D)) \). Thus the probability that two or more occur in \( R_n \) and the rest occur later is

\[
Q_n = \frac{5 \cdot 4 \cdot 3 \cdot (24 - 4) \cdot n - 2 \cdot 4 \cdot 3 \cdot 2 \cdot (24 - 4) \cdot n}{25 \cdot 24 \cdot 23 \cdot 22}.
\]

For \( n \geq 4 \), there are four triples that could occur and one quadruple so the probability is

\[
Q_n = \frac{4 \cdot 4 \cdot 3 \cdot 2 \cdot (24 - 4) \cdot n + 4 \cdot 3 \cdot 2 \cdot 1}{25 \cdot 24 \cdot 23 \cdot 22}.
\]

Combining these, the probability of following this case is \( \sum_{n=2}^{6} \frac{5}{4n-3} Q_n = \frac{21011}{313950} \).

Combining the three cases gives a probability of \( \frac{89}{690} + \frac{269}{2530} + \frac{21011}{313950} = \frac{347917}{1151150} \), or about 30.2\%, for the algorithm to succeed.

Using a backtracking strategy with some computer assistance, it is possible to prove that the results of Problems 3, 4, 5, and 6 are best possible. We do not know whether the results if Problem 7–10 are optimal. Computer assistance was instrumental in finding the algorithms in the solutions of Problems 5 through 10.

**Solution 7.** Run 20 horses in four disjoint heats. Race the four winners in \( R_5 \) against a new horse \( T \) and let \( X, Y \) be the top two in order.

**Case 1.** \( X \neq T \). Let \( W \) be the runnerup to \( X \) in \( X \)'s heat. Race \( Y \) and \( W \) against three new horses in \( R_6 \) and choose \( X \) and the winner of \( R_6 \). This finds the top two provided both \( A \) and \( B \) are among the 24 horses seen and it is not the case they each appear for the first time in \( R_6 \). The chance of this is \( \frac{21}{25} \frac{23}{24} + \frac{3}{25} \frac{21}{24} = \frac{91}{100} \).

**Case 2.** \( X = T \). The top two in \( R_5 \) are \( T \) and \( Y \). Race \( Y \) against the remaining four unseen horses in \( R_6 \). If \( Y \) is not last, we choose \( T \) and the winner of \( R_6 \). If \( Y \) is last then choose the top two in \( R_6 \). This case succeeds if \( A \) and \( B \) are both in the first 21 positions; or one of \( A \) and \( B \) is in the first 21, the other is not, and \( Y \) is one of \( C, D, \text{or} E; \) or the last race has, for its four new horses, \( A, B, \text{and either} C, D, \text{or} C, E, \text{or} D, E. \) The probability of one of these three things happening is the sum \( \frac{21}{25} \frac{20}{24} + 2 \frac{21}{25} \frac{4}{24} \left( 1 - \frac{1}{23} \frac{1}{22} \frac{1}{21} \right) + \frac{4}{25} \frac{3}{24} \left( \frac{3}{23} \frac{2}{22} \frac{1}{21} \right) = \frac{6199}{6325} \).

Because Case 2 occurs with probability \( \frac{1}{21} \), the overall success probability is \( \frac{20}{21} \cdot \frac{91}{100} + \frac{1}{21} \cdot \frac{6199}{6325} = \frac{40438}{44275} = 0.913337 \ldots \).

**Solution 8.** Start by racing four heats of five horses each. Then race the four winners in \( R_5 \) against a new horse \( T \). We consider three cases for possible outcomes to \( R_5 \).

**Case 1.** \( T \) does not finish in the top two in \( R_5 \). This case occurs with probability \( \frac{25}{28} \). Suppose the top two
horses in $R_5$ are $X$, $Y$, $Z$ in order (where $Z$ may be $T$).

(Digression on $\frac{25}{28}$. Horse $T$ wins $R_5$ with probability $\frac{1}{21}$ since this occurs if and only if $T$ is the fastest horse among the 21 raced up through $R_5$. Otherwise (probability $\frac{20}{21}$), the fastest horse is the winner of one of the first four races. Given this, $T$ finishes second in $R_5$ if and only if $T$ is the fastest of the 16 horses raced, the five who raced with the fastest horse being excluded. Thus overall $T$ finishes second in $R_5$ with probability $\frac{20}{21} \cdot \frac{5}{84}$. Hence $T$ finishes third or worse with probability $1 - \frac{5}{84} - \frac{1}{21} = \frac{25}{28}$. [Alternately, continuing this argument $T$ finishes third with probability $\frac{20 \times 15}{21 \times 16 \times 11} = \frac{25}{308}$, and $T$ finishes last with probability $\frac{20 \times 15 \times 10}{21 \times 16 \times 11 \times 6} = \frac{625}{924}$. Hence $T$ finishes third or lower with probability $\frac{25}{308} + \frac{125}{924} + \frac{625}{924} = \frac{25}{28}$.)]

In this case we race the following horses in $R_6$: (i) $Z$, (ii) the runnerups to $X$ and $Y$ in their heats, and two new horses. As our selected top three we take $X$, $Y$, and the winner in $R_6$ unless: The runnerup to $X$ finishes second or third in $R_6$ and $Z$ and the runner up to $Y$ are the bottom two finishers in $R_6$. In this exceptional case we choose $X$ and the top two horses in $R_6$. The probability of success in this case is $\frac{1386}{1955}$.

- **Proof of $\frac{1386}{1955}$**

(Proof. We have raced only 23 of the 25 horses. Hence the probability of success is $\frac{23}{25} \cdot \frac{22}{24} \cdot \frac{21}{23} \cdot p$ where $p$ is the probability of finding the three fastest conditioned on the fact that the top three horses are among the 23 actually raced. That is, to compute $p$ we assume there are only 23 horses. There are four ways we could get the wrong top three:

(i) The second runner up to $X$ in his first race (who did not even race in $R_6$) was actually $C$. Given that we are in Case 1 this occurs with probability $p_1 = \frac{21}{23} \cdot \frac{4}{22} \cdot \frac{3}{21} = \frac{6}{253}$, since there is a $\frac{21}{23}$ chance that $X = A$ is the fastest and then there are 4 and 3 ways to successively place $B$ and $C$ in $X$'s first race. Note that if this error does not occur, then the fastest three horses are among $X$, $Y$, and the entrants in $R_6$.

(ii) $X$, $Y$, and one of the new horses are the fastest three, but the exceptional case above occurs. The chance that $X$, $Y$, and a new horse are the fastest three is $\frac{3 \cdot 2}{23} \cdot \frac{21}{22} \cdot \frac{16}{21}$ since there are three choices for which of the top three places goes to the new horse and two horses that could be assigned this place, given this there is a $\frac{21}{22}$ chance that $X$ will be the fastest unassigned horse and further given that a $\frac{16}{21}$ chance that $Y$ will be the next fastest unassigned horse. The exception will only be invoked in this case if the other new horse and the runner-up to $X$ are faster than $Z$ and the runnerup to $Y$. This occurs with probability $\frac{4}{20 \times 19} + \frac{4}{20 \times 19} = \frac{7}{304}$, where the first term counts cases where the runner-up to $X$ is faster and the second term where the other new horse is faster. Thus this error occurs with probability $\frac{3 \cdot 2}{23} \cdot \frac{21}{22} \cdot \frac{16}{21} \cdot \frac{7}{304} = \frac{21}{4807}$.

(iii) $X$, the runnerup to $X$, and one of the new horses are the three fastest and the exception rule is not invoked. The chance that these three horses are the three fastest is $\frac{3 \cdot 2}{23} \cdot \frac{21}{22} \cdot \frac{4}{21}$ by essentially the same calculation as above. The exception rule will be invoked if the fast new horse is faster than the runner-up to $X$ (a $\frac{2}{3}$

\[\text{...}\]
chance given the above) and the other new horse beats both Z and the runnerup to Y. This last occurs exactly when this new horse one of the top 2 among the 17 horses obtained by excluding the five horses in X's first race and the fast new horse — thus with probability \(\frac{2}{17}\). Hence this error occurs with probability
\[
\frac{3\cdot2\cdot21\cdot22\cdot4\cdot21}{23\cdot22\cdot21\cdot22\cdot4\cdot21} \left(1 - \frac{4}{51}\right) = \frac{188}{4301}.
\]

(iv) X and the two new horses are the three fastest and the exception rule is not invoked. The chance that these are the fastest three horses is \(\frac{6}{23\cdot22}\) since there are six possible orders for these three horses. For each order there is a \(\frac{1}{23\cdot22}\) chance that the new horses will be have the assigned rankings and X will automatically be the fastest remaining horse. In this case the exception rule will occur if and only if the runnerup to X in his first race is faster than Z and the runnerup to Y. This occurs exactly when the runnerup to X is D or E. Hence with probability
\[
\frac{4}{20} + \frac{16\cdot4}{20\cdot19} = \frac{7}{19},
\]
where the first term corresponds to the runnerup to X being D and the second to Y = D and the runnerup being E. Thus this error occurs with probability
\[
\frac{6\cdot23\cdot22}{23\cdot22} \left(1 - \frac{7}{19}\right) = \frac{36}{4807}.
\]

These four cases are disjoint so the total probability we are wrong is
\[
\frac{6}{253} + \frac{21}{4807} + \frac{188}{4301} + \frac{36}{4807} = \frac{31}{391},
\]
and hence \(p = \frac{360}{391}\) and the overall probability is
\[
\frac{23\cdot22\cdot21}{25\cdot24\cdot23} p = \frac{1386}{1955}.
\]

**Case 2.** T finishes second in \(R_5\), which occurs with probability \(\frac{5}{84}\). Suppose the top three horses in \(R_5\) are X, T, Y in order. Then we race the following horses in \(R_6\): (i) Y, (ii) The runnerup to X in X's heat, and (iii) three new horses. As our selected three we take X, T, and the winner of \(R_6\) unless: Y finishes last in \(R_6\) or Y finishes next to last and one of the new horses wins. In this exceptional case we choose X and the top two horses in \(R_6\). The probability of success in this case is \(\frac{92519}{117300}\).

**Proof of** \(\frac{92519}{117300}\)

We have raced only 24 of the 25 horses. Hence the probability of success is \(\frac{22}{25}\ p\), where \(p\) is the probability of finding the three fastest conditioned on the fact that the top three horses are among the 24 actually raced. That is, to compute \(p\) we assume there are only 24 horses. There are five disjoint ways in which we could get the correct three fastest horses: (i) The three fastest could be X, T, Y. This occurs with probability
\[
\frac{21\cdot16\cdot15}{24\cdot23\cdot22} = \frac{105}{253}.
\]

(ii) The three fastest could be X, T, and a new horse with the exception not invoked. The chance that these are the three fastest is \(\frac{3\cdot3\cdot21\cdot16}{24\cdot23\cdot22} = \frac{63}{253}\) by the usual counts. The exception will be invoked in this case if and only if Y finishes fourth or fifth in \(R_6\). Horse Y will finish second with probability \(\frac{15}{21} = \frac{5}{7}\) and third with probability
\[
\frac{4\cdot15}{21\cdot17} + \frac{2\cdot15}{21\cdot20} = \frac{57}{238}\ (\text{where the two terms correspond to the runnerup to X being second and a new horse being second, respectively}).
\]
Thus the algorithm succeeds following this case with probability
\[
\frac{63}{253} \left(\frac{5}{7} + \frac{57}{238}\right) = \frac{2043}{8602}.
\]

(iii) The three fastest horses could be X, T, and the runnerup to X with the exception not invoked. The chance that these are the three fastest is \(\frac{2\cdot21\cdot16\cdot4}{24\cdot23\cdot22} = \frac{56}{253}\). In this case the exception rule will only be invoked if Y finishes last in \(R_6\), that is, if and only if all three new horses are faster than Y. This occurs with probability
\[
\frac{6\cdot23\cdot22}{23\cdot22} \left(1 - \frac{7}{19}\right) = \frac{36}{4807}.
\]
\[
\frac{3 \times 2}{18 \times 17 \times 16} = \frac{1}{816}.
\] Thus the algorithm succeeds following this case with probability \(\frac{56}{253} \left(1 - \frac{1}{816}\right) = \frac{5705}{25806}\).

(iv) The three fastest horses could be \(X\), the runnerup to \(X\) and a new horse with the exception rule invoked. The chance that these are the three fastest is \(\frac{3 \times 21 \times 4 \times 3}{24 \times 23 \times 22} = \frac{63}{1012}\). In this case the exception rule will only be invoked if \(Y\) finishes last in \(R_6\) or the new horse is faster than the runnerup to \(X\) and \(Y\) finishes fourth in \(R_6\). Since \(Y\) finishes third if and only if \(T\) and \(Y\) are the next two fastest horses, we have

\[
\text{Prob}(Y \text{ finishes 3rd in } R_6 \mid X, \text{runnerup to } X, \text{new are 3 fastest}) = \frac{16 \times 15}{18 \times 17} = \frac{40}{51},
\]

\[
\text{Prob}(Y \text{ finishes 4th in } R_6 \mid X, \text{runnerup to } X, \text{new are 3 fastest}) = \frac{2 \times 2 \times 16 \times 15}{18 \times 17 \times 16} = \frac{10}{51},
\]

\[
\text{Prob}(Y \text{ finishes 5th in } R_6 \mid X, \text{runnerup to } X, \text{new are 3 fastest}) = \frac{3 \times 2 \times 16 \times 15}{18 \times 17 \times 16 \times 15} = \frac{1}{51}.
\]

Since \(\frac{2}{3}\) of the orders on the three fastest have the new horse faster than the runner-up to \(X\), the conditional probability of invoking the exception rule in this case is \(\frac{2 \times 10}{3 \times 51} + \frac{1}{51} = \frac{23}{153}\) and the final probability of success following this case is \(\frac{63 \times 23}{1012 \times 153} = \frac{7}{748}\).

(v) The three fastest horses could be \(X\) and two new horses with the exception rule invoked. The chance that these are the three fastest is \(\frac{3 \times 21 \times 3 \times 2}{24 \times 23 \times 22} = \frac{63}{2024}\). The exception rule will be invoked if \(Y\) finishes last or fourth in \(R_6\). Since \(Y\) finishes third in \(R_6\) if and only if \(T\) and \(Y\) are the next two fastest horses, this probability is \(1 - \frac{16 \times 15}{21 \times 20} = \frac{3}{7}\) and the final probability of success following this case is \(\frac{63 \times 3}{2024 \times 7} = \frac{27}{2024}\).

Summing the five probabilities above the total probability of success in this case (with 24 horses) is

\[
p = \frac{105}{253} + \frac{2043}{8602} + \frac{5705}{25806} + \frac{7}{748} + \frac{27}{2024} = \frac{92519}{103224}.
\]

Thus the final probability is \(\frac{22}{25} p = \frac{92519}{117300}\).

Case 3. \(T\) wins \(R_5\), which occurs with probability \(\frac{1}{21}\). Suppose the top three in \(R_5\) are \(T, X, Y\) in order. Then we race the following horses in \(R_6\): (i) \(Y\), (ii) the runnerup to \(X\) in \(X\)'s heat, and (iii) three new horses. For the top three we choose \(T, X\), and the winner of \(R_6\) unless: The three new horses are the top three in \(R_6\). In this exceptional case we take \(T\) and the top two horses in \(R_6\). The probability of success in this case is \(\frac{1963}{2300}\).

\textbf{Proof of } \frac{1963}{2300}

We have raced only 24 of the 25 horses. Hence the probability of success is \(\frac{22}{25} p\), where \(p\) is the probability of finding the three fastest conditioned on the fact that the top three horses are among the 24 actually raced. That is, to compute \(p\) we assume there are only 24 horses. The algorithm can succeed in three ways:

(i) None of the new horses is among the three fastest. The probability that this occurs is \(\frac{21 \times 20 \times 19}{24 \times 23 \times 22} = \frac{665}{1012}\). In this case the exception cannot arise since one of \(Y\) and the runnerup to \(X\) wins \(R_6\) and hence \(T\) and \(X\) are the fastest two horses and the winner of \(R_6\) is the third-fastest.
(ii) The three fastest horses are \( T, X \), and one of the new horses and the exception is not invoked. The probability that the three fastest horses are \( T, X \) and a new horse is \( \frac{3 \times 3 \times 21 \times 20}{24 \times 23 \times 22} = \frac{315}{1012} \). The exception is invoked if and only if the next two fastest horses are both new. Given that we are in this case this occurs with probability \( 1 - \frac{2}{21 \times 20} = \frac{209}{210} \). Hence the probability of succeeding following this case is \( \frac{315 \times 209}{1012 \times 210} = \frac{627}{2024} \).

(iii) The three fastest horses are \( T \) and two new horses and the exception is invoked. The probability that these are the three fastest horses is \( \frac{3 \times 3 \times 2 \times 21}{24 \times 23 \times 22} = \frac{63}{2024} \). In this case the exception is invoked if and only if the next two fastest horses are \( X \) and the remaining new horse, which occurs with probability \( \frac{2 \times 20}{21 \times 20} = \frac{2}{21} \). Hence the probability of succeeding following this case is \( \frac{2}{21} \frac{63}{2024} = \frac{3}{1012} \).

Adding these three cases gives \( p = \frac{665}{1012} + \frac{627}{2024} + \frac{3}{1012} = \frac{1963}{2024} \) so the overall probability of success in this case is \( \frac{22}{25} p = \frac{1963}{2300} \).

Thus the overall probability of success is: \( \frac{25}{28} \frac{1386}{1955} + \frac{5}{24} \frac{92519}{117300} + \frac{1}{21} \frac{1963}{2300} = \frac{710047}{9853200} = 0.72058 \ldots \)

**Solution 9.** Run 20 horses in four heats. Race the winners in \( R_5 \) against a new horse \( T \) and let \( X, Y \) be top two in order. If \( X \neq T \), then the only contenders for the top two are \( X, Y, \) and \( Z \), the runnerup to \( X \) in \( X \)'s heat. Race \( X, Y, Z \) against two new horses in \( R_6 \). If instead \( X = T \), then the only contenders are \( T \) and \( Y \) and in \( R_6 \) we can race \( T \) and \( Y \) against three new horses. In either case the final choice is the top two in \( R_6 \). This method determines the top two in order and with certainty from the horses seen so far. So we fail only if one of \( A \) or \( B \) is among the unseen horses. The probability of this depends on the two cases. The probability in the first case is \( \frac{23}{25} \frac{22}{24} \) and in the second is \( \frac{24}{25} \frac{23}{24} \). So since the second case occurs with chance \( \frac{1}{21} \), we have \( \frac{20}{21} \frac{23}{25} \frac{22}{24} + \frac{1}{21} \frac{24}{25} \frac{23}{24} = \frac{1334}{1575} = 84.698 \% \).

**Solution 10.** A first try here would be to run five heats that are disjoint and then run the winners, choosing the top three from \( R_6 \). This succeeds when \( A, B, \) and \( C \) are in different races among the first five; the probability is \( \frac{20}{24} \frac{15}{23} = \frac{75}{138} \), or about 54.3\%. But there is a much better strategy.

Run 20 horses in four heats and race the winners in \( R_5 \) against a new horse \( T \). There are three possible outcomes to \( R_5 \):

**Case 1.** \( T \) does not finish in the top two in \( R_5 \). This case occurs with probability \( \frac{25}{28} \). Suppose the top three horses in \( R_5 \) are \( X, Y, Z \) in order (where \( Z \) may be \( T \)). Then we race the following horses in \( R_6 \): (i) \( Y, Z \), (ii) the runnerups to \( X \) and \( Y \) in their heats, and (iii) one new horse. As our alleged top three we take \( X \) and the top two horses in \( R_6 \) with \( X \) declared fastest.

The probability of success in this case is \( \frac{357}{575} \). (Proof. We have raced only 22 of the 25 horses. Hence the probability of success is \( \frac{22}{25} \frac{21}{24} \frac{20}{23} p \), where \( p \) is the probability of finding the three fastest conditioned on the fact that the top three horses are among the 22 actually raced. To compute this probability we may assume there are only 22 horses. The horses in \( R_5 \) and \( X \) are all the contenders for the three fastest except the second runnerup to \( X \) in \( X \)'s first race. Thus we will get the wrong top three horses if this horse is actually horse \( C \).
Also we have assumed \( X \) is the fastest even though he was never actually raced against the new horse from \( R_6 \). Thus if this new horse was \( A \), we will have the correct three horses but the order wrong. With probability \( \frac{1}{22} \) the new horse will be the fastest of the 22. Of the \( \frac{21}{22} \) of the time where \( X \) is the fastest of the 22, the chance that his runnerups in his first race were the next two fastest is \( \frac{4}{21} \cdot \frac{3}{20} \). Hence we get 

\[
p = 1 - \frac{1}{22} - \frac{21}{22} \cdot \frac{4}{21} \cdot \frac{3}{20} = \frac{51}{55} \quad \text{and the overall probability is } \frac{22}{25} \cdot \frac{21}{24} \cdot \frac{20}{23} \cdot \frac{51}{55} = \frac{357}{575} \left( \frac{1}{21} \right).
\]

**Case 2.** Horse \( T \) finishes second in \( R_5 \). This case occurs with probability \( \frac{1}{21} \). Suppose the top three in \( R_5 \) are \( X, T, Y \) in order. In this case we race the following horses in \( R_6 \): (i) \( T, Y \), (ii) The runnerup to \( X \) in \( X \)'s heat, and (iii) two new horses. As our alleged top three we take \( X \) and the top two horses in \( R_6 \) with \( X \) declared fastest. The probability of success in this case is \( \frac{63}{92} \). (Proof. We have raced only 23 of the 25 horses. Hence the probability of success is \( \frac{23}{25} \cdot \frac{22}{24} \cdot \frac{21}{23} \cdot p \), where \( p \) is the probability of finding the three fastest conditioned on the fact that the top three horses are among the 23 actually raced. That is, to compute \( p \) we assume there are only 23 horses. The horses raced in \( R_6 \) and \( X \) are all the contenders for the three fastest except the second runnerup to \( X \) in \( X \)'s heat. Thus as in the previous case we will get the wrong order unless \( X \) is the fastest (a \( \frac{21}{23} \) probability hit) and further if \( X \) is fastest we will get the wrong horses if the next two fastest where both in \( X \)'s first race (a \( 1 - \frac{4}{22} \cdot \frac{3}{21} \) probability hit). Hence 

\[
p = \frac{21}{23} \left( 1 - \frac{4}{22} \cdot \frac{3}{21} \right) = \frac{225}{253} \quad \text{and the overall probability is } \frac{23}{25} \cdot \frac{22}{24} \cdot \frac{21}{23} \cdot p = \frac{63}{92} \left( \frac{1}{21} \right).
\]

**Case 3.** Horse \( T \) wins \( R_5 \). This case occurs with probability \( \frac{1}{21} \). Suppose the top three horses in \( R_5 \) are \( T, X, Y \) in order. In this case we race the following horses in \( R_6 \): (i) \( X, Y \), (ii) the runnerup to \( X \) in \( X \)'s heat, and (iii) two new horses. As our alleged top 3 we take \( X \) and the top two horses in \( R_6 \) with \( X \) declared fastest.

The probability of success in this case is \( \frac{1617}{2300} \). (Proof. We have raced only 23 of the 25 horses. Hence the probability of success is \( \frac{23}{25} \cdot \frac{22}{24} \cdot \frac{21}{23} \cdot p \), where \( p \) is the probability of finding the three fastest conditioned on the fact that the top three horses are among the 23 actually raced. That is, to compute \( p \) we assume there are only 23 horses. The horses raced in \( R_6 \) and \( T \) are the only contenders for the three fastest horses. Thus we will be correct provided \( T \) is faster than the two new horses. Hence 

\[
p = \frac{21}{23} \quad \text{and the overall probability is } \frac{23}{25} \cdot \frac{22}{24} \cdot \frac{21}{23} \cdot p = \frac{1617}{2300} \left( \frac{1}{21} \right).
\]

Thus the overall probability of success is: 

\[
\frac{25}{28} + \frac{357}{575} + \frac{5}{84} + \frac{63}{92} + \frac{1}{21} \cdot \frac{1617}{2300} = \frac{5783}{9200} = 0.628587 \ldots
\]

The strategy in Solution 9 solves the “find top two, with certainty, in order” problem using six races when there are only 23 horses. But it is not possible do accomplish this with 24 horses. In the first four races there are 16 losing positions so we have at least eight unbeaten horses going into the fifth race. In any successful strategy five of these are run in race five and then the two winners are run against the remaining three in race six. But it could happen that \( B \) is been defeated in race 1 and the strategy fails.

Similarly it is possible to find the top three in order in six races when there are 21 horses, but not when there are 22. For the first, run 20 horses in four heats. Run the four winners in \( R_5 \) against the unseen horse. All horses have been run and the winner of \( R_5 \) is certain to be \( A \). There are at most five contenders to be \( B \) or \( C \). Run them in \( R_6 \); the top two will be \( B \) and \( C \). We have a rather complicated proof by cases that this is not possible with 22 horses.

**More Problems.** In five races, there is no choice of method if one wants to see all 25 horses run. But one
unsolved problems. prove optimality of any of the algorithms in problems 6–10. or find improvements to any of them!

reference


2. s. morris and s. wagon, problem 11613, american mathematical monthly 118 (2011) 937.