00314 In a television game show a contestant is successively asked questions $Q_1, \dots, Q_9$. After correctly answering Q_i and hearing Q_{i+1} she has the option of either going home with 2^i pounds or attempting to answer Q_{i+1}. If she answers Q_{i+1} incorrectly then she goes home with nothing. If she answers Q_9 correctly then the game ends and she takes home 2^9 pounds.

Upon hearing Q_i she is able to classify it as either easy or hard; these types occur with probabilities 0.95 and 0.05 respectively, independently for each question. If Q_i is easy her probability of answering it correctly is $9/(9+i)$, but if it is hard the probability is only $6/(6+i)$. At most once in the game she may choose to 'phone a friend'; the effect is to increase her chance of correctly answering Q_i to $10/(10+i)$, if it is easy, and to $7/(7+i)$, if it is hard.

Let W_i (and V_i) denote her expected winnings if she plays optimally from the point that she has correctly answered $i-1$ questions and has (or has not yet) phoned a friend. Write down dynamic programming equations from which you could compute V_1.

Show that $W_9 = 2^8$ and $V_9 = (21/20)2^8$.

Suppose she has answered 7 questions correctly and has not yet phoned a friend. What should she do if Q_8 is an easy question?

The producers of the show are considering a new game in which everything is the same except that the potential number of questions is unlimited. The game ends only when the contestant answers incorrectly or chooses to retire. Quoting any theorem necessary to justify your answer, show that for a contestant who plays optimally the new game is the same as the old.

Solution follows on next page. So don't
read on if you want to try this yourself first.

00314

The dynamic programming equations are

$$\begin{aligned}
W_i &= 0.95 \max\left\{2^{i-1}, \frac{9}{9+i} W_{i+1}\right\} + 0.05 \max\left\{2^{i-1}, \frac{6}{6+i} W_{i+1}\right\} \\
V_i &= 0.95 \max\left\{2^{i-1}, \frac{9}{9+i} V_{i+1}, \frac{10}{10+i} W_{i+1}\right\} \\
&\quad + 0.05 \max\left\{2^{i-1}, \frac{6}{6+i} V_{i+1}, \frac{7}{7+i} W_{i+1}\right\}
\end{aligned}$$

$i = 1, \dots, 9$, where as boundary conditions we take $V_{10} = W_{10} = 2^9$.

From these we find $W_9 = 2^8$ and $V_9 = (19/20)(10/19)2^9 + (1/20)2^8 = (21/20)2^8$.

If Q_8 is easy the contestant can either retire (reward 2^7), attempt to answer (expected reward $(9/17)(21/20)2^8$), or phone a friend and then answer (expected reward $(10/18)2^8$). Now a short calculation verifies that $(9/17)(21/20) > (10/18)$, so the best option is to answer without phoning a friend.

If the number of potential question is to be unlimited this is a case of an positive programming over the infinite horizon (i.e., maximizing positive rewards). It is a theorem for positive programming that *if a policy has a value function that satisfies the dynamic programming equation, then that policy is optimal.*

So consider a policy in which the contestant retires whenever she has answered 9 or more questions. This policy has $V_i = W_i = 2^{i-1}$ for all $i > 9$. Easily observe that these values satisfy the dynamic programming equation for all $i > 9$. Therefore, by the theorem quoted above, it is optimal to retire once 9 questions have been correctly answered. So far as optimal play is concerned, the contestant will never wish to attempt $Q_{10}, Q_{11}, \dots$.