Q1

(a) We can only interview at most k candidates, so

$$\sum_{i=1}^{n} x_i \le k.$$

We can only hire one candidate, so

$$\sum_{i=1}^{n} \sum_{v \sim D_i} \mathbf{Pr}[q_i = v] y_{iv} \le 1.$$

The probability that candidate i is hired conditioning on $q_i = v$ should be no more than the probability that the candidate is interviewed conditioning on $q_i = v$. Since the latter is x_i, regardless of what q_i is, we need

$$y_{iv} \le x_i.$$

To sum up, the linear program is

$$\begin{aligned}
\text{maximize} \quad & \sum_{i=1}^{n} \sum_{v \sim D_i} \mathbf{Pr}[q_i = v] y_{iv} v \\
\text{subject to} \quad & \sum_{i=1}^{n} x_i \le k \\
& \sum_{i=1}^{n} \sum_{v \sim D_i} \mathbf{Pr}[q_i = v] y_{iv} \le 1 \\
& y_{iv} \le x_i, \ \forall i \in [n] \ \forall v \\
& x_i \in [0, 1], \ \forall i \in [n] \\
& y_{iv} \in [0, 1], \ \forall i \in [n] \ \forall v
\end{aligned}$$

(b) Let (x^*, y^*) be an optimal solution for the LP in part (a). Consider the following algorithm where c and d are two constants.

(I) Start with $i = 1$.

(II) For candidate i, interview him/her with probability cx_i^*.

(III) Suppose candidate i is interviewed, hire him/her with probability $d\frac{y_{iv}^*}{x_i^*}$

(IV) If candidate i is hired, terminate.

(V) If k candidates have been interviewed or $i = n$, terminate. Otherwise, repeat from step (II) with $i = i + 1$.

Let X_i be the event that the algorithm reaches round i, and Y_{iv} be the even that candidate i is hired with value v. Then the expected total value obtained by the algorithm is

$$\sum_{i=1}^{n} \sum_{v} \mathbf{Pr}[q_i = v]\mathbf{Pr}[Y_{iv}]v.$$

Next we try to give a lower bound for $\mathbf{Pr}[Y_{iv}]$. We can write

$$\begin{aligned}
\mathbf{Pr}[Y_{iv}] &= \mathbf{Pr}[Y_{iv}|X_i]\mathbf{Pr}[X_i] + \mathbf{Pr}\left[Y_{iv}|\overline{X_i}\right]\mathbf{Pr}\left[\overline{X_i}\right] \\
&= \mathbf{Pr}[Y_{iv}|X_i]\mathbf{Pr}[X_i] + 0 \cdot \mathbf{Pr}\left[\overline{X_i}\right] \\
&= cdy_{iv}^{*}\mathbf{Pr}[X_i]
\end{aligned}$$

To give a lower bound on $\mathbf{Pr}[Y_{iv}]$, we give upper bound on $\mathbf{Pr}\left[\overline{X_i}\right]$. The event $\overline{X_i}$ (the algorithm does not reach round i) happens if A_1: we have hired someone before round i, or A_2: we have interviewed k candidates before round i.

Consider A_1. Using a union bound, we know

$$\mathbf{Pr}[A_1] \leq \sum_{j<i} \sum_{v} \mathbf{Pr}[q_j = v]\mathbf{Pr}[Y_{jv}] \leq cd \sum_{j=1}^{n} \sum_{v} \mathbf{Pr}[q_j = v]y_{jv}^{*}\mathbf{Pr}[X_j] \leq cd.$$

Consider A_2. Let Z_i be an indicator random variable that denotes if the algorithm interviews candidate i or not and let $Z = \sum_{i=1}^{n} Z_i$. Since $\mathbf{Pr}[Z_i = 1] = cx_i^{*}\mathbf{Pr}[X_i]$, we know that

$$\mathbf{E}[Z] = \sum_{i=1}^{n} \mathbf{E}[Z_i] = \sum_{i=1}^{n} \mathbf{Pr}[Z_i = 1] \leq \sum_{i=1}^{n} cx_i^{*} \leq ck.$$

By Markov's inequality, we have

$$\mathbf{Pr}[A_2] \leq \mathbf{Pr}[Z \geq k] \leq \frac{\mathbf{E}[Z]}{k} \leq c.$$

Combining what we have,

$$\begin{aligned}
\mathbf{Pr}[Y_{iv}] &= cdy_{iv}^{*}\mathbf{Pr}[X_i] \\
&= cdy_{iv}^{*}(1 - \mathbf{Pr}\left[\overline{X_i}\right]) \\
&= cdy_{iv}^{*}(1 - \mathbf{Pr}[A_1] - \mathbf{Pr}[A_2]) \\
&\geq cdy_{iv}^{*}(1 - cd - c)
\end{aligned}$$

Thus the expected value obtained by the algorithm is

$$\sum_{i=1}^{n} \sum_{v} \mathbf{Pr}[q_i = v]\mathbf{Pr}[Y_{iv}]v \geq \sum_{i=1}^{n} \sum_{v} \mathbf{Pr}[q_i = v]y_{iv}^{*}vcd(1 - cd - c) = cd(1 - cd - c)\mathrm{OPT}.$$

To get a constant ratio $\frac{1}{9}$, one can choose $c = \frac{1}{3}$ and $d = 1$.

Q2

Denote by $w_{i,t}$ the weight of expert i at the beginning of step t. Let $W_t = \sum_{i=1}^{n} w_{i,t}$. Denote by $x_{i,t}$ the prediction of expert i at step t and by y_t the actual result at step t. Then we have

$$\begin{aligned}
W_{t+1} &= \sum_{i=1}^{n} w_{i,t+1} \\
&= \sum_{i:x_{i,t}=y_t} w_{i,t+1} + \sum_{i:x_{i,t}\neq y_t} w_{i,t+1} \\
&= \sum_{i:x_{i,t}=y_t} w_{i,t+1} + \sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}\geq \frac{W_t}{4n}}} w_{i,t+1} + \sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}< \frac{W_t}{4n}}} w_{i,t+1} \\
&= \sum_{i:x_{i,t}=y_t} w_{i,t} + \sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}\geq \frac{W_t}{4n}}} \frac{1}{2} w_{i,t} + \sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}< \frac{W_t}{4n}}} w_{i,t} \\
&= W_t - \frac{1}{2} \sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}\geq \frac{W_t}{4n}}} w_{i,t}.
\end{aligned}$$

If the algorithm makes a mistake at step t, then

$$\sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}\geq \frac{W_t}{4n}}} w_{i,t+1} + \sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}< \frac{W_t}{4n}}} w_{i,t+1} \geq \frac{1}{2} W_t.$$

Also for any step t (whether a mistake happens or not) we know

$$\sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}< \frac{W_t}{4n}}} w_{i,t} < \frac{1}{4} W_t.$$

Combining the above two facts, we have

$$\sum_{\substack{i:x_{i,t}\neq y_t \\ w_{i,t}\geq \frac{W_t}{4n}}} w_{i,t} \geq \frac{1}{4} W_t$$

and therefore

$$W_{t+1} \leq \frac{7}{8} W_t$$

if a mistake happens at step t.

If a mistake does not happen at step t, we note that $W_{t+1} \leq W_t$ as the weight on every expert never increases.

Next we show by induction that for any expert i at step t, $w_{i,t} \geq \frac{W_t}{8n}$. When $t = 1$, $w_{i,1} = \frac{W_1}{n} \leq \frac{W_1}{8n}$. At step t, if expert i does not make a mistake or if $w_{i,t} < \frac{W_t}{4n}$ then her weight would not change and thus

$$w_{i,t+1} = w_{i,t} \geq \frac{W_t}{8n} \geq \frac{W_{t+1}}{8n};$$

if expert i makes a mistake and $w_{i,t} \geq \frac{W_t}{4n}$ then

$$w_{i,t+1} = \frac{1}{2} w_{i,t} \geq \frac{W_t}{8n}.$$

In a contiguous block t to $t+j$, if the best expert i^* makes m mistakes, then

$$W_{t+j} \geq w_{i^*,t+j} \geq \frac{1}{2^m} w_{i^*,t} \geq \frac{1}{2^m} \frac{W_t}{8n}.$$

Denote by M the number of mistakes made by the algorithm from step t to step $t+j$, then we have $W_{t+j} \leq \left(\frac{7}{8}\right)^m W_t$. Combining the upper bound and the lower bound on W_{t+j}, we have

$$\frac{1}{2^m} \frac{W_t}{8n} \leq W_{t+j} \leq \left(\frac{7}{8}\right)^m W_t$$

which gives us $M = O\left(m + \log n\right)$.

Q3

(a) **Fact 1** *For* $0 \leq x \leq 1$*,* $(1+\epsilon)^x \leq 1 + \epsilon x$*; for* $-1 \leq x \leq 0$*,* $(1+\epsilon)^x \leq 1 + \frac{\epsilon}{1+\epsilon} x$*.*

Suppose $0 \leq c_{i,t} \leq 1$ for any i, t. Given $R_{i,t} = \frac{1}{1+\epsilon} \sum_j p_{j,t} c_{j,t} - c_{i,t}$, one can see $-1 < R_{i,j} < 1$.

Consider step $t+1$. Denote by A the set of experts that are awake at step t and $P \subset A$ those with $R_{i,t} \geq 0$. Write $L = \sum_j p_{j,t} c_{j,t}$ and thus $R_{i,t} = \frac{\epsilon}{1+\epsilon} L - c_{i,t}$. We have

$$
\begin{aligned}
W_{t+1} &= \sum_{i \in A} w_{i,t+1} + \sum_{i \notin A} w_{i,t+1} \\
&= \sum_{i \in A \cap P} w_{i,t+1} + \sum_{i \in A \setminus P} w_{i,t+1} + \sum_{i \notin A} w_{i,t} \\
&= \sum_{i \in A \cap P} w_{i,t} (1+\epsilon)^{R_{i,t}} + \sum_{i \in A \setminus P} w_{i,t} (1+\epsilon)^{R_{i,t}} + \sum_{i \notin A} w_{i,t} \\
&\leq \sum_{i \in A \cap P} w_{i,t} (1 + \epsilon R_{i,t}) + \sum_{i \in A \setminus P} w_{i,t} \left(1 + \frac{\epsilon}{1+\epsilon} R_{i,t}\right) + \sum_{i \notin A} w_{i,t} \\
&= \sum_{i \in A} w_{i,t} + \sum_{i \notin A} w_{i,t} + \sum_{i \in A \cap P} w_{i,t} \epsilon R_{i,t} + \sum_{i \in A \setminus P} w_{i,t} \frac{\epsilon}{1+\epsilon} R_{i,t} \\
&= W_t + \sum_{i \in A \cap P} w_{i,t} \epsilon \left(\frac{\epsilon}{1+\epsilon} L - c_{i,t}\right) + \sum_{i \in A \setminus P} w_{i,t} \frac{\epsilon}{1+\epsilon} \left(\frac{\epsilon}{1+\epsilon} L - c_{i,t}\right) \\
&= W_t + \frac{1}{1+\epsilon} L \left(\sum_{i \in A \cap P} w_{i,t} \epsilon + \sum_{i \in A \setminus P} w_{i,t} \frac{\epsilon}{1+\epsilon}\right) - \left(\epsilon \sum_{i \in A \cap P} w_{i,t} c_{i,t} + \frac{\epsilon}{1+\epsilon} \sum_{i \in A \setminus P} w_{i,t} c_{i,t}\right) \\
&= W_t + \frac{\epsilon}{1+\epsilon} \frac{\sum_{i \in A} w_{i,t} c_{i,t}}{\sum_{i \in A} w_{i,t}} \left(\sum_{i \in A \cap P} w_{i,t} + \sum_{i \in A \setminus P} w_{i,t} \frac{1}{1+\epsilon}\right) - \left(\epsilon \sum_{i \in A \cap P} w_{i,t} c_{i,t} + \frac{\epsilon}{1+\epsilon} \sum_{i \in A \setminus P} w_{i,t} c_{i,t}\right) \\
&< W_t + \frac{\epsilon}{1+\epsilon} \sum_{i \in A} w_{i,t} c_{i,t} - \left(\epsilon \sum_{i \in A \cap P} w_{i,t} c_{i,t} + \frac{\epsilon}{1+\epsilon} \sum_{i \in A \setminus P} w_{i,t} c_{i,t}\right) \\
&= W_t + \sum_{i \in A \cap P} \left(\frac{\epsilon}{1+\epsilon} - \epsilon\right) w_{i,t} c_{i,t} \\
&\leq W_t.
\end{aligned}
$$

By induction, this means $W_t \leq n$ for any step t as $W_0 = n$.

(b) Let $T_i = (t_1, t_2, \cdots, t_k)$ and denote by C_t the cost of the algorithm at step t. By definition $\text{cost}_i(\text{ALG}) = \mathbf{E}\left[\sum_{t_j \in T_i} C_{t_j}\right] = \sum_{t_j \in T_i} \mathbf{E}\left[C_{t_j}\right]$ and $\text{cost}_i(i) = \sum_{t_j \in T_i} c_{i,t_j}$. We can write

$$
w_{i,t_{j+1}} = w_{i,t_j} (1+\epsilon)^{\frac{1}{1+\epsilon} \mathbf{E}\left[C_{t_j}\right] - c_{i,t_j}}.
$$

Therefore, at the last step T, we have

$$
\begin{aligned}
w_{i,T} &= w_{i,0} (1+\epsilon)^{\sum_{t_j \in T_i} \frac{1}{1+\epsilon} \mathbf{E}\left[C_{t_j}\right] - c_{i,t_j}} \\
&= (1+\epsilon)^{\frac{1}{\epsilon} \text{cost}_i(\text{ALG}) - \text{cost}_i(i)}.
\end{aligned}
$$

By part (a) we know $w_{i,T} \leq W_T \leq n$, so we have

$$
(1+\epsilon)^{\frac{1}{\epsilon} \text{cost}_i(\text{ALG}) - \text{cost}_i(i)} \leq n
$$

which gives us

$$\mathrm{cost}_i\left(\mathrm{ALG}\right) \leq \left(1+\epsilon\right)\mathrm{cost}_i\left(i\right) + O\left(\frac{1}{\epsilon}\log n\right).$$