Combinatorics and the hiring problem

Conrado Martínez
U. Politècnica Catalunya

Joint work with M. Archibald, U. Cape Town, South Africa

Ottawa, May 2009

The hiring problem

- Originally introduced by Broder *et al.* (SODA 2008)
- A (potentially infinite) sequence of i.i.d. random variables Q_i uniformly distributed in $[0, 1]$
- At step i you either hire or discard candidate i with score Q_i
- Decisions are irrevocable
- Goals: hire candidates at some reasonable rate, improve the "mean" quality of the company's staff

The hiring problem

- Originally introduced by Broder *et al.* (SODA 2008)
- A (potentially infinite) sequence of i.i.d. random variables Q_i uniformly distributed in $[0, 1]$
- At step i you either hire or discard candidate i with score Q_i
- Decisions are irrevocable
- Goals: hire candidates at some reasonable rate, improve the "mean" quality of the company's staff

The hiring problem

- Originally introduced by Broder *et al.* (SODA 2008)
- A (potentially infinite) sequence of i.i.d. random variables Q_i uniformly distributed in $[0, 1]$
- At step i you either hire or discard candidate i with score Q_i
- Decisions are irrevocable
- Goals: hire candidates at some reasonable rate, improve the "mean" quality of the company's staff

The hiring problem

- Originally introduced by Broder *et al.* (SODA 2008)
- A (potentially infinite) sequence of i.i.d. random variables Q_i uniformly distributed in $[0, 1]$
- At step i you either hire or discard candidate i with score Q_i
- Decisions are irrevocable
- Goals: hire candidates at some reasonable rate, improve the “mean” quality of the company’s staff

The hiring problem

- Originally introduced by Broder *et al.* (SODA 2008)
- A (potentially infinite) sequence of i.i.d. random variables Q_i uniformly distributed in $[0, 1]$
- At step i you either hire or discard candidate i with score Q_i
- Decisions are irrevocable
- Goals: hire candidates at some reasonable rate, improve the "mean" quality of the company's staff

The hiring problem

- Here: a permutation σ of length n, candidate i has score $\sigma(i)$; the permutation is actually presented as a sequence of unknown length $S = s_1, s_2, s_3, \ldots$ with $1 \leq s_i \leq i+1$, s_i is the rank of the ith candidate relative to the candidates seen so far (i included)
- Our model is equivalent after "normalization", but is amenable to techniques from analytic combinatorics
- $\mathcal{H}(\sigma)$ = the set of candidates hired in permutation σ
- $h(\sigma) = \#\mathcal{H}(\sigma)$

The hiring problem

- Here: a permutation σ of length n, candidate i has score $\sigma(i)$; the permutation is actually presented as a sequence of unknown length $S = s_1, s_2, s_3, \ldots$ with $1 \le s_i \le i+1$, s_i is the rank of the ith candidate relative to the candidates seen so far (i included)
- Our model is equivalent after "normalization", but is amenable to techniques from analytic combinatorics
- $\mathcal{H}(\sigma)$ = the set of candidates hired in permutation σ
- $h(\sigma) = \#\mathcal{H}(\sigma)$

The hiring problem

- Here: a permutation σ of length n, candidate i has score $\sigma(i)$; the permutation is actually presented as a sequence of unknown length $S = s_1, s_2, s_3, \ldots$ with $1 \leq s_i \leq i + 1$, s_i is the rank of the ith candidate relative to the candidates seen so far (i included)
- Our model is equivalent after "normalization", but is amenable to techniques from analytic combinatorics
- $\mathcal{H}(\sigma) =$ the set of candidates hired in permutation σ
- $h(\sigma) = \#\mathcal{H}(\sigma)$

The hiring problem

- Here: a permutation σ of length n, candidate i has score $\sigma(i)$; the permutation is actually presented as a sequence of unknown length $S = s_1, s_2, s_3, \ldots$ with $1 \leq s_i \leq i+1$, s_i is the rank of the ith candidate relative to the candidates seen so far (i included)
- Our model is equivalent after "normalization", but is amenable to techniques from analytic combinatorics
- $\mathcal{H}(\sigma) =$ the set of candidates hired in permutation σ
- $h(\sigma) = \#\mathcal{H}(\sigma)$

Rank-based hiring

A hiring strategy is rank-based if and only if it only depends on the relative rank of the current candidate compared to the candidates seen so far.

Rank-based hiring

- Rank-based strategies modelize actual restrictions to measure qualities
- Many natural strategies are rank-based, e.g.,
  - above the best
  - above the mth best
  - above the median
  - above the P% best
- Other interesting strategies are not, e.g., above the average, above a threshold.

Rank-based hiring

- Rank-based strategies modelize actual restrictions to measure qualities
- Many natural strategies are rank-based, e.g.,
  - above the best
  - above the mth best
  - above the median
  - above the $P\%$ best
- Other interesting strategies are not, e.g., above the average, above a threshold.

Rank-based hiring

- Rank-based strategies modelize actual restrictions to measure qualities
- Many natural strategies are rank-based, e.g.,
  - above the best
  - above the mth best
  - above the median
  - above the $P\%$ best
- Other interesting strategies are not, e.g., above the average, above a threshold.

Rank-based hiring

- Rank-based strategies modelize actual restrictions to measure qualities
- Many natural strategies are rank-based, e.g.,
  - above the best
  - above the mth best
  - above the median
  - above the $P\%$ best
- Other interesting strategies are not, e.g., above the average, above a threshold.

Rank-based hiring

- Rank-based strategies modelize actual restrictions to measure qualities
- Many natural strategies are rank-based, e.g.,
    - above the best
    - above the mth best
    - above the median
    - above the $P\%$ best
- Other interesting strategies are not, e.g., above the average, above a threshold.

Rank-based hiring

- Rank-based strategies modelize actual restrictions to measure qualities
- Many natural strategies are rank-based, e.g.,
  - above the best
  - above the *m*th best
  - above the median
  - above the *P*% best
- Other interesting strategies are not, e.g., above the average, above a threshold.

Rank-based hiring

- Rank-based strategies modelize actual restrictions to measure qualities
- Many natural strategies are rank-based, e.g.,
  - above the best
  - above the mth best
  - above the median
  - above the $P\%$ best
- Other interesting strategies are not, e.g., above the average, above a threshold.

Rank-based hiring

The recursive decomposition of permutations

$$\mathcal{P} = \epsilon + \mathcal{P} \star Z$$

is the natural choice for the analysis of rank-based strategies.

Rank-based hiring

- Let $\sigma \star j$ denote the permutation one gets after relabelling j, $j+1, \ldots, n = |\sigma|$ to $j+1, j+2, \ldots, n+1$ and appending j at the end.
- Ex: $32451 \star 3 = 425613$, $32451 \star 2 = 435612$
- Let $X_j(\sigma) = 1$ if candidate with score j is hired after σ and $X_j(\sigma) = 0$ otherwise.
- $h(\sigma \star j) = h(\sigma) + X_j(\sigma)$

Rank-based hiring

- Let $\sigma \star j$ denote the permutation one gets after relabelling j, $j+1, \ldots, n = |\sigma|$ to $j+1, j+2, \ldots, n+1$ and appending j at the end.
- Ex: $32451 \star 3 = 425613$, $32451 \star 2 = 435612$
- Let $X_j(\sigma) = 1$ if candidate with score j is hired after σ and $X_j(\sigma) = 0$ otherwise.
- $h(\sigma \star j) = h(\sigma) + X_j(\sigma)$

Rank-based hiring

- Let $\sigma \star j$ denote the permutation one gets after relabelling j, $j+1, \ldots, n = |\sigma|$ to $j+1, j+2, \ldots, n+1$ and appending j at the end.
- Ex: $32451 \star 3 = 425613$, $32451 \star 2 = 435612$
- Let $X_j(\sigma) = 1$ if candidate with score j is hired after σ and $X_j(\sigma) = 0$ otherwise.
- $h(\sigma \star j) = h(\sigma) + X_j(\sigma)$

Rank-based hiring

- Let $\sigma \star j$ denote the permutation one gets after relabelling j, $j+1, \ldots, n = |\sigma|$ to $j+1, j+2, \ldots, n+1$ and appending j at the end.
- Ex: $32451 \star 3 = 425613$, $32451 \star 2 = 435612$
- Let $X_j(\sigma) = 1$ if candidate with score j is hired after σ and $X_j(\sigma) = 0$ otherwise.
- $h(\sigma \star j) = h(\sigma) + X_j(\sigma)$

Rank-based hiring

Theorem

Let $H(z,u) = \sum_{\sigma \in \mathcal{P}} \frac{z^{|\sigma|}}{|\sigma|!} u^{h(\sigma)}$.
Then

$$(1-z)\frac{\partial}{\partial z} H(z,u) - H(z,u) = (u-1) \sum_{\sigma \in \mathcal{P}} X(\sigma) \frac{z^{|\sigma|}}{|\sigma|!} u^{h(\sigma)},$$

where $X(\sigma)$ *the number of* j *such that* $X_j(\sigma) = 1$.

Rank-based hiring

We can write $h(\sigma) = 0$ if σ is the empty permutation and $h(\sigma \star j) = h(\sigma) + X_j(\sigma)$.

$$\begin{aligned}
H(z, u) = \sum_{\sigma \in \mathcal{P}} \frac{z^{|\sigma|}}{|\sigma|!} u^{h(\sigma)} &= 1 + \sum_{n>0} \sum_{\sigma \in \mathcal{P}_n} \frac{z^{|\sigma|}}{|\sigma|!} u^{h(\sigma)} \\
&= 1 + \sum_{n>0} \sum_{1 \le j \le n} \sum_{\sigma \in \mathcal{P}_{n-1}} \frac{z^{|\sigma \star j|}}{|\sigma \star j|!} u^{h(\sigma \star j)} \\
&= 1 + \sum_{n>0} \sum_{1 \le j \le n} \sum_{\sigma \in \mathcal{P}_{n-1}} \frac{z^{|\sigma|+1}}{(|\sigma| + 1)!} u^{h(\sigma) + X_j(\sigma)} \\
&= 1 + \sum_{n>0} \sum_{\sigma \in \mathcal{P}_{n-1}} \frac{z^{|\sigma|+1}}{(|\sigma| + 1)!} u^{h(\sigma)} \sum_{1 \le j \le n} u^{X_j(\sigma)}.
\end{aligned}$$

Rank-based hiring

Since $X_j(\sigma)$ is either 0 or 1 for all j and all σ, we have

$$\sum_{1 \leq j \leq n} u^{X_j(\sigma)} = (|\sigma| + 1 - X(\sigma)) + uX(\sigma),$$

where $X(\sigma) = \sum_{1 \leq j \leq |\sigma|+1} X_j(\sigma)$.

$$H(z,u) = 1 + \sum_{n>0} \sum_{\sigma \in \mathcal{P}_{n-1}} \frac{z^{|\sigma|+1}}{(|\sigma|+1)!} u^{h(\sigma)} \Big((|\sigma| + 1 - X(\sigma)) + uX(\sigma) \Big).$$

The theorem follows after differentiation and a few additional algebraic manipulations.

Pragmatic strategies

A hiring strategy is pragmatic if and only if

- Whenever it would hire a candidate with score j, it would hire a candidate with a larger score

$$X_j(\sigma) = 1 \implies X_{j'}(\sigma) = 1 \qquad \text{for all } j' \geq j$$

- The number of scores it would potentially hire increases at most by one if and only if the candidate in the previous step was hired

$$X(\sigma \star j) \leq X(\sigma) + X_j(\sigma)$$

Pragmatic strategies

A hiring strategy is pragmatic if and only if

- Whenever it would hire a candidate with score j, it would hire a candidate with a larger score

$$X_j(\sigma) = 1 \implies X_{j'}(\sigma) = 1 \qquad \text{for all } j' \geq j$$

- The number of scores it would potentially hire increases at most by one if and only if the candidate in the previous step was hired

$$X(\sigma \star j) \leq X(\sigma) + X_j(\sigma)$$

Pragmatic strategies

- The first condition is very natural and reasonable; the second one is technically necessary for several results we discuss later
- Above the best, above the mth best, above the $P\%$ best, ... are all pragmatic

Pragmatic strategies

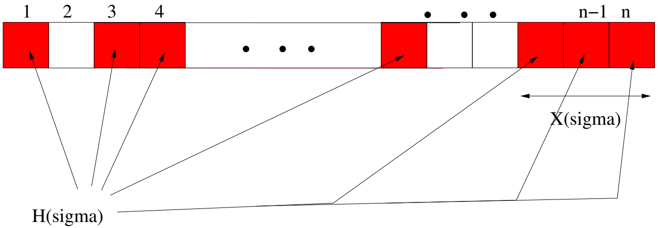
- The first condition is very natural and reasonable; the second one is technically necessary for several results we discuss later
- Above the best, above the mth best, above the P% best, ... are all pragmatic

Pragmatic strategies

Theorem

For any pragmatic hiring strategy and any permutation σ, the $X(\sigma)$ best candidates of σ have been hired (and possibly others).

Pragmatic strategies



Pragmatic strategies

Let r_n denote the rank of the last hired candidate in a random permutation, and

$$g_n = 1 - \frac{r_n}{n}$$

is called the gap.

Theorem

For any pragmatic hiring strategy,

$$\mathbb{E}[g_n] = \frac{1}{2n}(\mathbb{E}[X_n] - 1),$$

where $\mathbb{E}[X_n] = [z^n] \sum_{\sigma \in \mathcal{P}} X(\sigma) z^{|\sigma|} / |\sigma|!$.

Hiring above the maximum

Candidate i is hired if and only if her score is above the score of the best currently hired candidate.

- $X(\sigma) = 1$
- $\mathcal{H}(\sigma) = \{i : i \text{ is a left-to-right maximum}\}$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = \ln n + O(1)$
- Variance of h_n is also $\ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0, 1)$

Hiring above the maximum

Candidate i is hired if and only if her score is above the score of the best currently hired candidate.

- $X(\sigma) = 1$
- $\mathcal{H}(\sigma) = \{i : i \text{ is a left-to-right maximum}\}$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = \ln n + O(1)$
- Variance of h_n is also $\ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0,1)$

Hiring above the maximum

Candidate i is hired if and only if her score is above the score of the best currently hired candidate.

- $X(\sigma) = 1$
- $\mathcal{H}(\sigma) = \{i : i \text{ is a left-to-right maximum}\}$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = \ln n + O(1)$
- Variance of h_n is also $\ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0,1)$

Hiring above the maximum

Candidate i is hired if and only if her score is above the score of the best currently hired candidate.

- $X(\sigma) = 1$
- $\mathcal{H}(\sigma) = \{i : i \text{ is a left-to-right maximum}\}$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = \ln n + O(1)$
- Variance of h_n is also $\ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0, 1)$

Hiring above the mth best

Candidate i is hired if and only if her score is above the score of the mth best currently hired candidate.

- $X(\sigma) = |\sigma| + 1$ if $|\sigma| < m$; $X(\sigma) = m$ if $|\sigma| \geq m$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = m \ln n + O(1)$ for fixed m
- Variance of h_n is also $m \ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0,1)$
- The case of arbitrary m can be studied by introducing $\mathrm{H}(z,u,v) = \sum_{m \geq 1} v^m H^{(m)}(z,u)$, where $H^{(m)}(z,u)$ is the GF that corresponds to a given particular m.
- We can show that $\mathbb{E}[h_n] = m(H_n - H_m + 1) \sim m \ln(n/m) + m + O(1)$, with H_n the nth harmonic number

Hiring above the mth best

Candidate i is hired if and only if her score is above the score of the mth best currently hired candidate.

- $X(\sigma) = |\sigma| + 1$ if $|\sigma| < m$; $X(\sigma) = m$ if $|\sigma| \geq m$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = m \ln n + O(1)$ for fixed m
- Variance of h_n is also $m \ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0, 1)$
- The case of arbitrary m can be studied by introducing $\mathrm{H}(z, u, v) = \sum_{m \geq 1} v^m H^{(m)}(z, u)$, where $H^{(m)}(z, u)$ is the GF that corresponds to a given particular m.
- We can show that $\mathbb{E}[h_n] = m(H_n - H_m + 1) \sim m \ln(n/m) + m + O(1)$, with H_n the nth harmonic number

Hiring above the mth best

Candidate i is hired if and only if her score is above the score of the mth best currently hired candidate.

- $X(\sigma) = |\sigma| + 1$ if $|\sigma| < m$; $X(\sigma) = m$ if $|\sigma| \geq m$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = m \ln n + O(1)$ for fixed m
- Variance of h_n is also $m \ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0, 1)$
- The case of arbitrary m can be studied by introducing $\mathrm{H}(z, u, v) = \sum_{m \geq 1} v^m H^{(m)}(z, u)$, where $H^{(m)}(z, u)$ is the GF that corresponds to a given particular m.
- We can show that $\mathbb{E}[h_n] = m(H_n - H_m + 1) \sim m \ln(n/m) + m + O(1)$, with H_n the nth harmonic number

Hiring above the mth best

Candidate i is hired if and only if her score is above the score of the mth best currently hired candidate.

- $X(\sigma) = |\sigma| + 1$ if $|\sigma| < m$; $X(\sigma) = m$ if $|\sigma| \geq m$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = m \ln n + O(1)$ for fixed m
- Variance of h_n is also $m \ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0,1)$
- The case of arbitrary m can be studied by introducing $\mathrm{H}(z,u,v) = \sum_{m\geq 1} v^m H^{(m)}(z,u)$, where $H^{(m)}(z,u)$ is the GF that corresponds to a given particular m.
- We can show that $\mathbb{E}[h_n] = m(H_n - H_m + 1) \sim m \ln(n/m) + m + O(1)$, with H_n the nth harmonic number

Hiring above the mth best

Candidate i is hired if and only if her score is above the score of the mth best currently hired candidate.

- $X(\sigma) = |\sigma| + 1$ if $|\sigma| < m$; $X(\sigma) = m$ if $|\sigma| \geq m$
- $\mathbb{E}[h_n] = [z^n] \left.\frac{\partial H}{\partial u}\right|_{u=1} = m \ln n + O(1)$ for fixed m
- Variance of h_n is also $m \ln n + O(1)$ and after proper normalization h_n^* converges to $\mathcal{N}(0,1)$
- The case of arbitrary m can be studied by introducing $\mathrm{H}(z,u,v) = \sum_{m\geq 1} v^m H^{(m)}(z,u)$, where $H^{(m)}(z,u)$ is the GF that corresponds to a given particular m.
- We can show that $\mathbb{E}[h_n] = m(H_n - H_m + 1) \sim m \ln(n/m) + m + O(1)$, with H_n the nth harmonic number

Hiring above the median

Candidate i is hired if and only if her score is above the score of the median of the scores of currently hired candidates.

- $X(\sigma) = \lceil (h(\sigma) + 1)/2 \rceil$
- $\sqrt{\frac{n}{\pi}}(1 + O(n^{-1})) \leq \mathbb{E}[h_n] \leq 3\sqrt{\frac{n}{\pi}}(1 + O(n^{-1}))$
- This result follows easily by using previous theorem with $X_L(\sigma) = (h(\sigma) + 1)/2$ and $X_U(\sigma) = (h(\sigma) + 3)/2$ to lower and upper bound

Hiring above the median

Candidate i is hired if and only if her score is above the score of the median of the scores of currently hired candidates.

- $X(\sigma) = \lceil (h(\sigma) + 1)/2 \rceil$
- $\sqrt{\frac{n}{\pi}}(1 + O(n^{-1})) \le \mathbb{E}[h_n] \le 3\sqrt{\frac{n}{\pi}}(1 + O(n^{-1}))$
- This result follows easily by using previous theorem with $X_L(\sigma) = (h(\sigma) + 1)/2$ and $X_U(\sigma) = (h(\sigma) + 3)/2$ to lower and upper bound

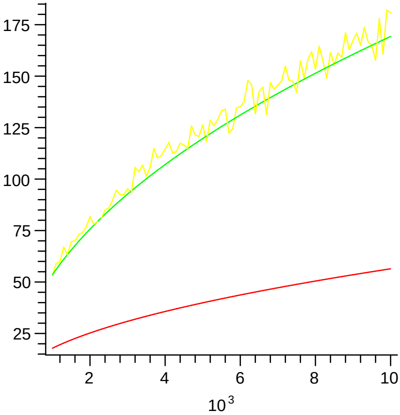
Hiring above the median

Candidate i is hired if and only if her score is above the score of the median of the scores of currently hired candidates.

- $X(\sigma) = \lceil (h(\sigma) + 1)/2 \rceil$
- $\sqrt{\frac{n}{\pi}}(1 + O(n^{-1})) \leq \mathbb{E}[h_n] \leq 3\sqrt{\frac{n}{\pi}}(1 + O(n^{-1}))$
- This result follows easily by using previous theorem with $X_L(\sigma) = (h(\sigma) + 1)/2$ and $X_U(\sigma) = (h(\sigma) + 3)/2$ to lower and upper bound

Hiring above the median

$n \in \{1000, \ldots, 10000\}$, $M = 100$ random permutations for each n



In red: $\mathbb{E}[h_n]$ with X_L; in green: $\mathbb{E}[h_n]$ with X_U; in yellow: simulation

Final remarks

- Other quantities, e.g. time of the last hiring, etc. can also be analyzed using techniques from analytic combinatorics
- We have also analyzed hiring above the $P\%$ best candidate with the same machinery, actually we have explicit solutions for $H(z, u)$
- We have extensions of these results to cope with randomized hiring strategies
- Many variants of the problem are interesting and natural; for instance, include firing policies

Final remarks

- Other quantities, e.g. time of the last hiring, etc. can also be analyzed using techniques from analytic combinatorics
- We have also analyzed hiring above the $P\%$ best candidate with the same machinery, actually we have explicit solutions for $H(z, u)$
- We have extensions of these results to cope with randomized hiring strategies
- Many variants of the problem are interesting and natural; for instance, include firing policies

Final remarks

- Other quantities, e.g. time of the last hiring, etc. can also be analyzed using techniques from analytic combinatorics
- We have also analyzed hiring above the $P\%$ best candidate with the same machinery, actually we have explicit solutions for $H(z, u)$
- We have extensions of these results to cope with randomized hiring strategies
- Many variants of the problem are interesting and natural; for instance, include firing policies

Final remarks

- Other quantities, e.g. time of the last hiring, etc. can also be analyzed using techniques from analytic combinatorics
- We have also analyzed hiring above the $P\%$ best candidate with the same machinery, actually we have explicit solutions for $H(z, u)$
- We have extensions of these results to cope with randomized hiring strategies
- Many variants of the problem are interesting and natural; for instance, include firing policies

Thanks for your attention!