The Best Choice Problem: calculating the optimal stopping rule using Markov chain theory

Cory Simon

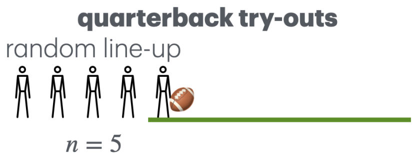
School of Chemical, Biological, and Environmental Engineering.
Oregon State University. Corvallis, OR.

February 9, 2025

Abstract

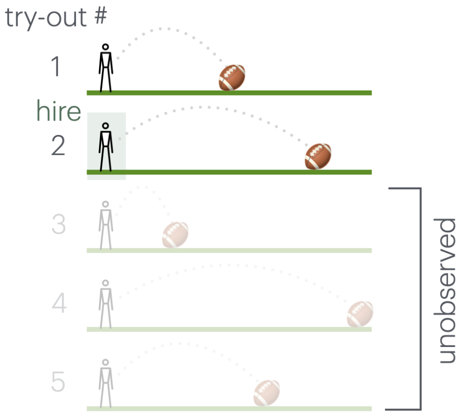
The Best Choice Problem (AKA the Secretary Problem or the Marriage Problem) is a simple, interesting, and intuitive problem of decision-making under uncertainty. Here, we explain how to calculate the optimal stopping rule for the Best Choice Problem using the theory of Markov chains.

The Best Choice Problem. [1] To hire a quarterback for our football team, we hold try-outs. A known number n of prospective quarterbacks line up randomly to sequentially throw a football as far as they can. We wish to hire the *best* prospect, i.e. the prospect who throws the football the furthest among all n prospects. (Assume each prospect's throw is reproducible.)



For each try out, we observe the range (travel distance) of the football thrown by a prospect. Thereafter, we must make a decision: either (i) hire this prospect, in which case the try-outs end, or (ii) irrevokably reject this prospect, in which case the next prospect tries out. If we get to the last (nth) prospect, we must hire them as our only option. Recall of a rejected prospect is not possible.

E.g., below is a realization of try-outs with $n = 5$ prospects, where we hired the second prospect to try out, but failed to hire the best prospect, who was set to try out later.



We want a decision-making algorithm that maximizes the probability that we hire the best prospect. A prospect who tries out and throws the football the furthest we've observed thus far in the tryouts is a *candidate*. Clearly, the optimal decision for a non-candidate is rejection. However, the optimal decision (hire or reject) for a candidate is not obvious. A candidate may or may not be the best, as we lack knowledge about the throwing ranges of the prospects who haven't yet tried out.

💡 We should be reluctant to both (i) hire a candidate early in the try-outs and (ii) reject a candidate late in the try-outs. With few (many) observed throws for us to compare against and many (few) prospects left to try out, it's unlikely (likely) an early (late) candidate is the best.

We consider the stopping rule: reject the first $t^* \in \{0, ..., n-1\}$ prospects that try out, then hire the first candidate that appears thereafter (or, the last prospect, if we get to them). The first t^* tryouts constitute the *observation phase* (exploration) to gather knowledge about the throwing capability of the pool of prospects. The remaining tryouts constitute the *selection phase*—exploitation of this knowledge to enhance our chance of hiring the best thrower.

❓ What is the optimal length of the observation phase t^*, that maximizes the probability that we hire the best thrower as quarterback for our football team?

The quarterback tryouts as a stochastic process. We represent the *a priori* unknown ranks of the prospects, in terms of their football throw range compared with all n prospects, as a random list $\mathbf{X} = (X_1, ..., X_n)$. The random variable $X_t \in [1..n] := \{1, ..., n\}$ is the rank of the prospect set to try out at the discrete try-out time $t \in [1..n]$. The list of rankings $\mathbf{X}$ follows a uniform distribution over the set of permutations $\mathcal{X}_n!$ of the list $\mathcal{X}_n := (1, ..., n)$:

$$\mathbb{P}[(X_1, ..., X_n) = (x_1, ..., x_n)] = \frac{1}{n!} \text{ for } (x_1, ..., x_n) \in \mathcal{X}_n!. \tag{1}$$

Since all rankings are equally likely, the probability that the *best* (rank-one) prospect tries out at time t is equal to the fraction of the lists $(x_1, ..., x_n)$ in $\mathcal{X}_n!$ with $x_t = 1$:

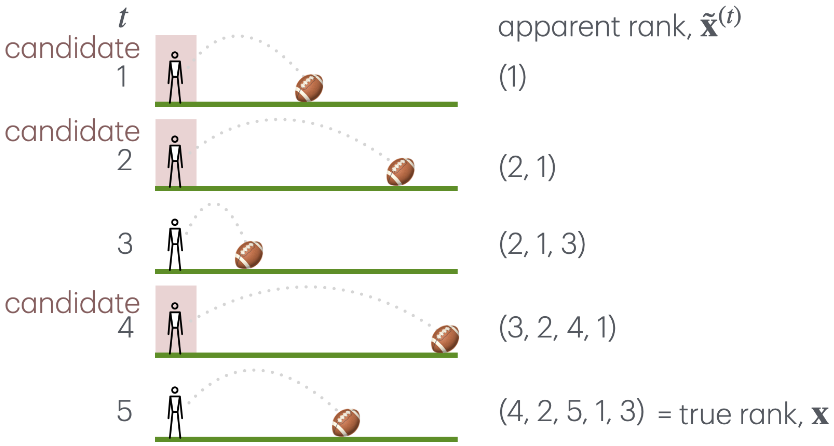
$$\mathbb{P}[X_t = 1] = \frac{(n-1)!}{n!} = \frac{1}{n} \text{ for } t \in [1..n]. \tag{2}$$

Immediately after try-out t, we observe only the *apparent*-ranks of the prospects, in terms on their football throw range compared with only the first t prospects who have tried out thus far. The list of apparent ranks $\tilde{\mathbf{X}}^{(t)} := (\tilde{X}_1^{(t)}, ..., \tilde{X}_t^{(t)})$ at time t also follows a uniform distribution:

$$\mathbb{P}[(\tilde{X}_1^{(t)}, ..., \tilde{X}_t^{(t)}) = (\tilde{x}_1^{(t)}, ..., \tilde{x}_t^{(t)})] = \frac{1}{t!} \text{ for } (\tilde{x}_1^{(t)}, ..., \tilde{x}_t^{(t)}) \in \mathcal{X}_t!. \tag{3}$$

Arrival of candidates. Prospect t is a *candidate* quarterback iff at time t they have apparent-rank of one, i.e. iff, $\tilde{x}_t^{(t)} = 1$. A candidate quarterback is an apparent-best, but may not be the best (unless, $t = n$); a prospect with a higher football throw range may try out later.

As a checkpoint to explain our notation and terminology, below is an example realization of try-outs for $n = 5$ with the apparent-ranks shown and candidates highlighted at each try-out time.


t
apparent rank, $\tilde{\mathbf{x}}^{(t)}$
candidate
1
(1)
candidate
2
(2, 1)
3
(2, 1, 3)
candidate
4
(3, 2, 4, 1)
5
(4, 2, 5, 1, 3) = true rank, $\mathbf{x}$

At time t, we can write the conditional probability that a candidate is the best:

$$\mathbb{P}[X_t = 1 \mid \tilde{X}_t^{(t)} = 1] = \frac{\mathbb{P}[(X_t = 1) \cap (\tilde{X}_t^{(t)} = 1)]}{\mathbb{P}[\tilde{X}_t^{(t)} = 1]} = \frac{1/n}{1/t} = \frac{t}{n} = 1 - \frac{n-t}{n} \text{ for } t \in [1..n]. \tag{4}$$

Explaining the numerator, the event $(X_t = 1) \cap (\tilde{X}_t^{(t)} = 1)$ is the same as the event $X_t = 1$ because the best will certainly be a candidate. The last equality aligns with our intuition: the probability that a candidate is the best is simply the probability that the best does not try out later.

💡 Eqn. 4 tells us that candidates appearing later in the tryout sequence are more likely to be the best than candidates appearing earlier. The reason for this is that it's harder to be the apparent-best thrower when there are more prospects to compare (compete) against. This intuition underlies our reasoning for having the observation/exploration phase. We should reject candidates early in the try-outs, as an apparent-rank of one is less likely to imply a true-rank of one, and hire a candidate that appears later in the try-outs, when an apparent-rank of one is more likely to coincide with a true-rank of one.

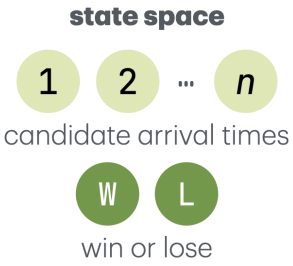
Seeking the optimal stopping rule. Our stopping rule, in our attempt to hire the best thrower, is parameterized by $t^* \in [0...n-1]$:

1. observation phase (exploitation): reject all prospects in try-out $t \leq t^*$ (even if they are candidates) to calibrate our assessment of the football-throwing capability of the pool of applicants
2. selection phase (exploitation): hire the first candidate that appears in tryout $t > t^*$. (If we get to the last, nth prospect, we must hire them, whether or not a candidate.)

The outcome of our stopping rule is either: (1) *win* (W), if we hire the best thrower, or (2) *lose* (L), if we hire a thrower who is not the best. Our objective is to find the parameter $t^* = t^*_{\text{opt}}$ defining the optimal stopping rule, which yields the maximal probability of a W outcome.

To find t^*_{opt}, we must calculate the probability of a W outcome under any stopping rule t^*. For this, we use the theory of Markov chains [2]—first used to analyze the Best Choice Problem by Dynkin in 1963 [3] (according to [1]; I could not confirm, as Dynkin's paper is written in Russian).

Formulating the try-outs, our stopping rule, and the outcome as a Markov chain. We formulate the arrival (try-out) times of candidates, the stopping rule parameterized by t^*, and the resulting W or L outcome as a discrete-time, homogeneous, finite Markov chain (MC) $T_0, T_1, T_2, ...$ with state space $\mathcal{T} := \{1, ..., n, W = n+1, L = n+2\}$. Each event in the sequence of events constituting the MC represents one of: (1) observing a candidate at some try-out time, (2) winning (W), by hiring or having hired the best thrower, or (3) losing (L), by hiring or having hired a prospect who is not the best thrower.



Accordingly, the random variable T_i of the MC denotes the try-out time at which we observe the ith candidate (case: $T_i \in [1..n]$), the state of winning (case: $T_i = W$), or the state of losing (case: $T_i = L$). Since the first prospect in try-out $t = 1$ has apparent-rank of one, with no other prospects to compare against, and thus is a candidate, the initial state of the MC is $T_0 = 1$ with certainty.

Note, states W and L are absorbing. Once a MC enters an absorbing state, it never leaves.

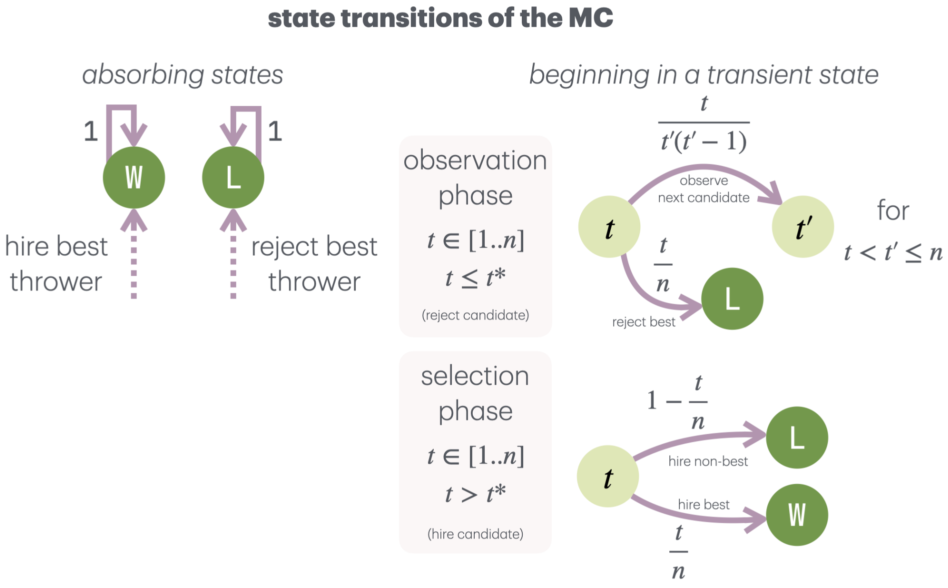
💡 For a given stopping rule t^*, we wish to calculate the probability that the MC eventually absorbs in the state W. The optimal stopping rule $t^* = t^*_{\text{opt}}$ follows from the t^* that maximizes this probability.

MC transition probabilities. First, we assemble the transition probabilities for our Markov chain into an $n+2 \times n+2$ stochastic matrix $\mathbf{P}$ whose entry (t, t') gives the probability of transitioning to state $t' \in \mathcal{T}$ of the Markov chain when the current state is $t \in \mathcal{T}$:

$$p_{t,t'} := \mathbb{P}[T_{i+1} = t' \mid T_i = t]. \tag{5}$$

We emphasize that these transition probabilities depend on the stopping rule t^*.

The possible state transitions and their probabilities are depicted below, broken down into when we begin in an absorbing state versus a transient state. The transient states are further broken down into whether we begin in the observation phase or the selection phase.



First, consider starting in either of the two absorbing states, W or L. We have $p_{W,W} = p_{L,L} = 1$ because, once we win or lose, we cannot exit this state. Our hiring decision is irrevocable.

Second, consider starting in the state $t \in [1..n]$ that represents observing a candidate at try-out time t. I.e., the prospect who just tried out at time t threw the football the furthest we've seen so far. According to our stopping rule, we reject this candidate if $t \leq t^*$ and hire this candidate if $t > t^*$. For $t > t^*$, we hire the candidate and enter (i) the W state with the probability that this candidate is the best, t/n, and (ii) the L state with the probability that this candidate is not the best, $1 - t/n$. For $t \leq t^*$, we reject the candidate and enter (i) the L state with the probability that this candidate is the best, t/n, and (ii) state $t' \in [t+1..n]$, corresponding to observing the next candidate, with probability:

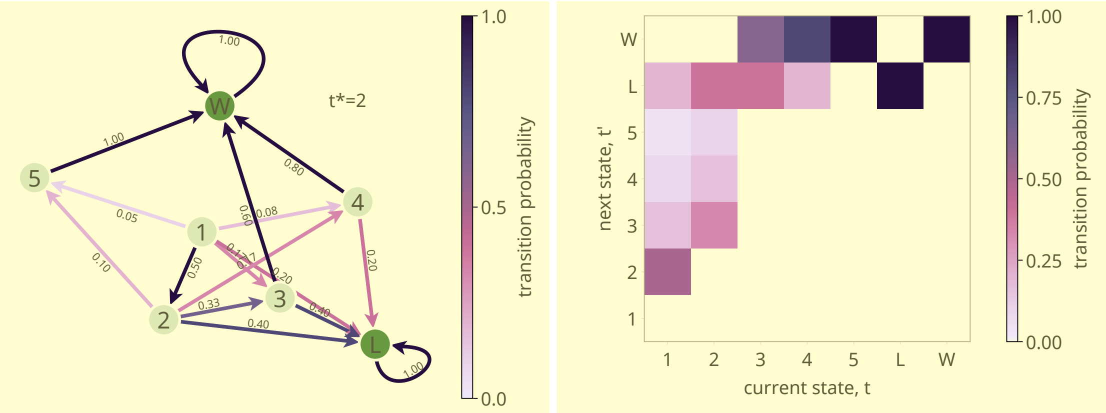
$$\mathbb{P}[T_{i+1} = t' \mid T_i = t] = \frac{\mathbb{P}[(T_{i+1} = t') \cap (T_i = t)]}{\mathbb{P}[T_i = t]} = \frac{1/[t'(t'-1)]}{1/t}. \tag{6}$$

The first equality follows from Bayes' theorem. The denominator is the probability that the prospect who tries out at time t will be a candidate, i.e. will have apparent-rank of one. The numerator is the probability that both prospects at time t' and at time t will be candidates. This is equal to the fraction of apparent-ranking lists $\tilde{\mathbf{x}}^{(t')}$ in the set of lists $\mathcal{X}_{t'}$! with $x_{t'}^{(t')} = 1$ and $x_t^{(t')} = 2$, $(t'-2)!/t'!$.

Putting it all together, the entries of the stochastic matrix $\mathbf{P}$ specifying the transition probabilities of the MC are:

$$p_{t,t'} := \mathbb{P}[T_{i+1} = t' \mid T_i = t] =$$
$$\begin{cases} 1 & t = t' = W \vee t = t' = L \\ \begin{cases} t/n & t' = L \quad \text{reject best} \\ t/[t'(t'-1)] & t < t' \leq n \quad \text{next candidate} \end{cases} & t \in [1..t^*] \quad \text{observation phase} \\ \begin{cases} t/n & t' = W \quad \text{hire best} \\ 1 - t/n & t' = L \quad \text{hire non-best} \end{cases} & t \in [t^*+1..n] \quad \text{selection phase} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

As an example, we visualize both the transition graph and matrix of the MC for $n = 5$ try-outs and a stopping rule of $t^* = 2$ below.



Probability distribution over state space. Now, we calculate the MC's probability distribution over states at each MC-time $i \geq 0$ (distinct from try-out time $t \in [1..n]$). This will enable us to calculate the probability of winning under a given stopping rule t^*.

At any given discrete MC time $i \geq 0$, let $\boldsymbol{\pi}^{(i)} \in [0,1]^{n+2}$ be a probability vector whose entry t is the probability the MC resides in state $t \in \mathcal{T}$ at MC-time i. That is:

$$\pi_t^{(i)} := \mathbb{P}[T_i = t]. \tag{8}$$

Again, we emphasize that $\boldsymbol{\pi}^{(i)}$ depends on the stopping rule t^*.

The initial distribution over states is:

$$\boldsymbol{\pi}^{(0)} = (1, 0, ..., 0) \tag{9}$$

since we begin the MC by observing a candidate at try-out time $t = 1$.

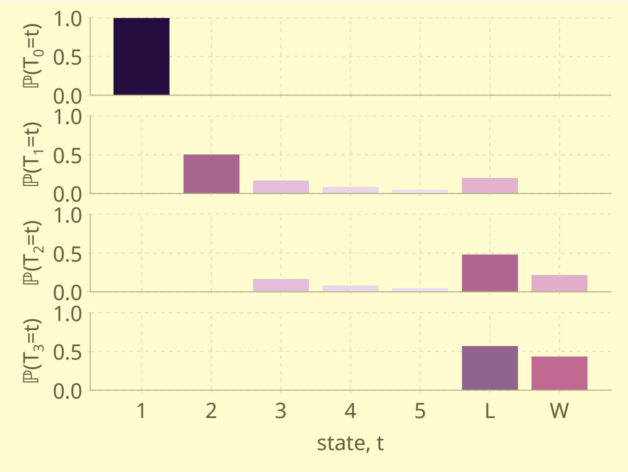
After a single state transition, the probability that the MC resides in state $t' \in \mathcal{T}$ sums, over all initial states, the probability of being in that initial state and then transitioning to state t':

$$\pi_{t'}^{(1)} = \sum_{t \in \mathcal{T}} \mathbb{P}[T_1 = t' \mid T_0 = t] \mathbb{P}[T_0 = t]. \tag{10}$$

In matrix-vector form, we have $\boldsymbol{\pi}^{(1)\intercal} = \boldsymbol{\pi}^{(0)\intercal}\mathbf{P}$. Continuing this logic, the probability distribution over states for the MC after i state transitions involves powers of the transition matrix:

$$\boldsymbol{\pi}^{(i)\intercal} = \boldsymbol{\pi}^{(0)\intercal}\mathbf{P}^i. \tag{11}$$

For example, we use this equation to calculate $\boldsymbol{\pi}^{(i)}$ for $i \in [0..3]$ under stopping rule $t^* = 2$, visualized below.



Probability of the MC absorbing in the W state. At MC-time $i = t^* + 1$, the MC is certain to reside in one of the two absorbing states W or L. (Here is a proof by contradiction. Suppose at MC-time $i = t^* + 1$, the MC resides in a state $t \in [1..n]$ instead. This means we have observed a candidate at try-out time t. Because the MC begins in state $t_0 = 1$ and W and L are absorbing states, the previously realized MC states $t_0, t_1, ..., t_{t^*}$ must have all corresponded to the observation of [distinct] candidates, as well. Together, we must have observed, then, $t^* + 2$ candidates for the MC to reside in state $t \in [1..n]$ at MC-time $i = t^* + 1$. However, this violates our stopping rule, which instructs us to have hired the $(t^* + 1)$th candidate that appeared during the selection phase and prohibits us from observing a $(t^* + 2)$nd candidate.) Consequently,

$$\boldsymbol{\pi}^{(t^*+1)} =: [0, ..., 0, p_W, p_L] \tag{12}$$

where p_W is the probability we hire the best and p_L is the probability we hire a prospect that is not the best—both, a function of our stopping rule t^*.

Voilá, here's how we can calculate the probability of hiring the best quarterback under a stopping rule t^*:

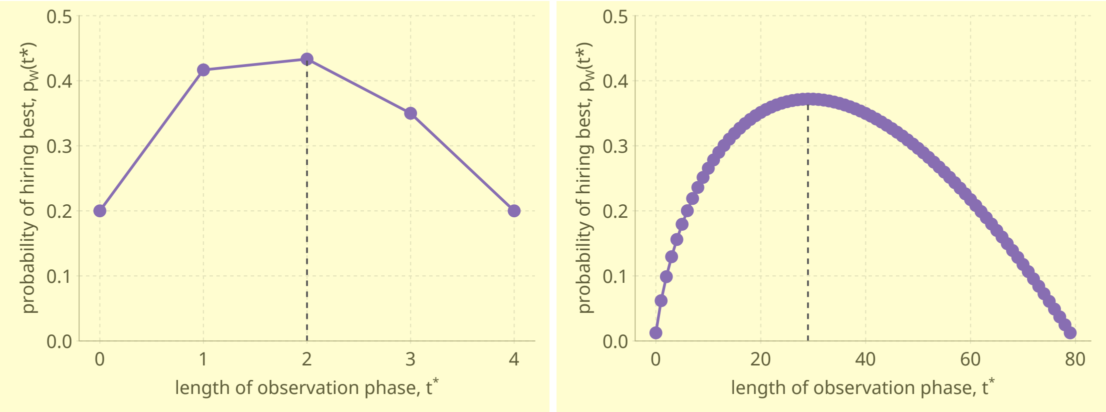
1. build the transition matrix $\mathbf{P}$ of the MC in eqn. 7
2. build the initial state probability vector $\boldsymbol{\pi}^{(0)}$ in eqn. 9
3. compute the state probability vector after $t^* + 1$ MC-steps, $\boldsymbol{\pi}^{(t^*+1)\intercal} = \boldsymbol{\pi}^{(0)\intercal}\mathbf{P}^{t^*+1}$
4. grab the probability of hiring the best quarterback, p_W, from the second-to-last entry of $\boldsymbol{\pi}^{(t^*+1)}$.

Finding the optimal stopping rule. We calculate the optimal stopping rule

$$t^*_{\text{opt}} = \underset{t^* \in [0..n-1]}{\arg\max}\ p_W(t^*) \tag{13}$$

via a brute-force, exhaustive search.

For example, we plot the probability of hiring the best quarterback $p_W(t^*)$ against the length of the observation phase t^* for $n = 5$ and $n = 80$ try-outs below. For $n = 2$, we find the optimal stopping rule $t^*_{\text{opt}} = 2$ gives a probability of hiring the best quarterback of ≈43.3%. For $n = 80$, we find the optimal stopping rule $t^*_{\text{opt}} = 29$ gives a probability of hiring the best of ≈37.2%.



Why is the arrival time of candidates Markovian? We glossed over why the sequence of arrival times of candidates is Markovian, i.e. why:

$$\mathbb{P}[T_{i+1} = t_{i+1} \mid T_0 = 1, ..., T_i = t_i] = \mathbb{P}[T_{i+1} = t_{i+1} \mid T_i = t_i] \text{ for } t_1 < t_2 < \cdots < t_{i+1} \leq n. \tag{14}$$

The left is the probability that the prospect at try-out time t_{i+1} throws the football further than all prospects at try-out time $t \in [1..t_{i+1} - 1]$ given candidates were observed at try-out times $1, t_1, ..., t_i$. After we observed the candidate at try-out time t_i, we're only comparing the next prospects' football throw ranges with this candidate's football throw range when determining the next prospect to label a candidate. After all, this candidate at try-out t_i throws the football further than all preceding candidates. So, knowledge of the arrival times of preceding candidates $t_1, ..., t_{i-1}$ is not needed.

Modifications and extensions to the Best Choice Problem. Many modifications and extensions of the Best Choice Problem have been framed and solved, including uncertain employment, recall of rejected candidates, adoption of different utility functions from the ’the best or nothing’ mindset here, an unknown number of prospects, try-out costs, finite memory, and multiple hires. For some extensions, the MC formulation is fruitful. [1]

References

[1] PR Freeman. The secretary problem and its extensions: A review. *International Statistical Review/Revue Internationale de Statistique*, pages 189–206, 1983.

[2] Robert P Dobrow. *Introduction to stochastic processes with R*. John Wiley & Sons, 2016.

[3] Evgenii Borisovich Dynkin. Optimal choice of the stopping moment of a markov process. In *Doklady Akademii Nauk*, volume 150, pages 238–240. Russian Academy of Sciences, 1963.