Smooth Sorting

Aresh Pourkavoos

October 11, 2023

softmax : $\mathbb{R}^n \to [0,1]^n$, $\text{softmax}(v)_i = \frac{\exp(v_i)}{\sum_{j=1}^n \exp(v_j)}$
Gives smoothed version of argmax, which returns vector with 1 in position of largest element and 0 elsewhere
Entries have sum 1: may be interpreted as probability that given element of v is the largest
Want to generalize to $\text{soft}_k(v)$ for $k \in \{1, \ldots, n\} = [n]$: probability that v_j is the k^{th}-largest
Desired properties:

1. $\text{soft}_1 = \text{softmax}$
2. $\text{soft}_{n+1-k}(v) = \text{soft}_k(-v)$
3. The largest entry of $\text{soft}_k(v)$ is in the same place as the k^{th}-largest entry of v
4. $\sum_{j=1}^n \text{soft}_k(v)_j = 1$ (exactly one element of v is at position k)
5. $\sum_{k=1}^n \text{soft}_k(v)_j = 1$ (v_j is in exactly one position)
6. soft_k is smooth
7. $\text{soft}_k(v) = \text{soft}_k(u)$ if there is $c \in \mathbb{R}$ such that for all i, $u_i = v_i + c$,
   i.e. soft_k does not change when adding a constant to all elements of the input

Define $S \in \mathbb{R}^{n \times n}$ by $S_{jk} = \text{soft}_k(v)_j$
Properties 4 and 5: S is doubly stochastic
Birkhoff's theorem: $S = \sum_\pi P(\pi) M_\pi$ over permutations π of $[n]$,
M_π has 1 in position $(\pi(i), i)$ and 0 elsewhere, $P(\pi) \in [0,1]$, $\sum_\pi P(\pi) = 1$
Interpretation: $P(\pi)$ = probability of $v_{\pi(k)}$ being k^{th}-largest for all k
Justification: $S_{jk} = \sum_{\pi(k)=j} P(\pi)$: v_j is k^{th}-largest iff a permutation where v_j is k^{th}-largest is chosen
Reframe problem: define a distribution over permutations, get 4 and 5 for free
Equivalently, define procedure to pick a random permutation: "smooth" sorting

First attempt: property 1 may be rewritten $P(\pi(1) = j) = \text{softmax}(v)_j$
Can ensure by picking $\pi(1) = a$ first, then permuting the remaining $n-1$ elements recursively:

- Define $\sigma : [n] \setminus \{a\} \leftrightarrow [n-1]$ by $\sigma(i) = i$ if $i < a$ and $i - 1$ if $i > a$: map between indices
- Let $v' \in \mathbb{R}^{n-1}$ where $v'_i = v_{\sigma(i)}$ (i.e. v' is v with v_a removed)
- Find permutation π' of $[n-1]$ based on v'
- Define $\pi(i) = a$ if $i = 1$ and $\sigma^{-1}(\pi'(i-1))$ otherwise

Base case $n = 0$: identity permutation
Since softmax satisfies independence of irrelevant alternatives (IIA), this is equivalent to:

- Repeatedly pick $i \in [n]$ with probability $\text{softmax}(v)_i$
- If already seen, discard; else, add to list

- Once list has length n, $\pi(k)$ = element at k^{th} position

However, this procedure does not satisfy property 2:

- Let $n = 3$, $v_1 = 2$, $v_2 = 1$, $v_3 = 0$
- Then $\text{soft}_3(v)_3 = \text{soft}_1(-v)_3 = \frac{\exp(0)}{\exp(-2)+\exp(-1)+\exp(0)} \approx 0.67$
- Procedure picks v_3 last iff either (v_1 first, then v_2) or (v_2 first, then v_1)
- Probability $= \frac{\exp(2)}{\exp(2)+\exp(1)+\exp(0)} \frac{\exp(1)}{\exp(1)+\exp(0)} + \frac{\exp(1)}{\exp(2)+\exp(1)+\exp(0)} \frac{\exp(2)}{\exp(2)+\exp(0)} \approx 0.70$

New procedure: should act on v "symmetrically" to satisfy property 2
Idea: pick $\pi(1) = a$ and $\pi(n) = b$ at once, then find middle recursively using v without v_a or v_b
(Base cases are now $n = 0$ or 1: still identity)
First attempt: pick a by softmax(v) and b by softmax($-v$) independently
Equivalently:

- Define matrix E by $E_{ab} = \exp(v_a - v_b)$
- Pick (a, b) with probability proportional to E_{ab}, i.e. equal to E_{ab}/(sum of entries of E)

E stores non-normalized joint distribution of (a, b)
Non-normalized marginal distribution of a given by $\sum_b E_{ab} = \exp(v_a) \sum_b \exp(-v_b)$
Proportional to $\exp(v_a)$, which is proportional to $\text{softmax}(v)_a$, so property 1 satisfied
Problem: a could equal b since $E_{aa} = \exp(v_a - v_a) = 1 \neq 0$: can't have $\pi(1) = \pi(n)$
Need to fix E by making diagonal 0 without changing marginal distributions (ratios of row/column sums)
First attempt:

- Just zero out diagonal
- Equivalently, try to pick a and b again with the same probabilities until $a \neq b$
- Sum of each row decreases by 1, but since rows can have different sums, their ratio can change
- Thus marginal distribution of a not preserved (same for b)
- Scaling E to restore sum of entries doesn't help since ratios of row/column sums don't change

Second attempt:

- Zero out diagonal, add $\frac{1}{n-1}$ to all other entries
- Equivalently, pick a and b as usual, but if $a = b$, pick a distinct pair *uniformly* instead
- Row/column sums decrease by 1 in 1 spot and increase by $\frac{1}{n-1}$ in $n - 1$ spots: no change
- Thus marginal distributions of a and b preserved: success!

Recall this is a single iteration: overall process finds $\pi(1)$ and $\pi(n)$, then $\pi(2)$ and $\pi(n-1)$, ...
Use probabilities of picking each π to define soft_k as described by Birkhoff's thm.
Satisfies properties 1, 2, 4, 5, 6, and 7 by virtue of its construction
However, procedure suggests that closed form of soft_k must case on whether $k < n/2$
(Also expensive by naive method, but not currently a concern)