Quan Nguyen
quan@wustl.edu

Research note
April 3, 2020

Agenda:

- Section 1: behavior of PoF under specific assumptions
- Section 2: distributions of unnormalized probabilities in individual columns

# 1 Behavior of PoF in special matrices

We are interested in analyzing how the PoF quantity would behave under the following assumptions:

- The individual values in each column following a specific column-specific distribution, and there is a specific order in the expected cost among the columns. This corresponds to the idea that some intervention types are better than others. We assume:

  $$\mathbb{E}[X_1] \leq \mathbb{E}[X_2] \leq ... \leq \mathbb{E}[X_m]$$

  where $X_j$ is the random variable for the cost of an agent when assigned to intervention $j$. So intervention 1 is the most effective, while intervention $j$ is the least.

- The values in each column are sorted in increasing order. This implies that there is a common order in cost among the agents in every intervention: if one agent will do well in one intervention, they will also do well in the others. This assumption is related to the fact that the real costs are highly correlated among the interventions.

Again, we'd like to study the behavior of $C(E)/C(L)$, potentially as a function of the probability distribution of costs in each intervention, $m$ (the number of intervention types), and the capacity combination of these interventions.

In the following subsection, we consider the problem when we make stronger assumptions on the cost matrix: when the column distributions are separated uniform distributions.

## 1.1 Separated uniform distributions

In this section, we assume that the costs in column $j = 1, ..., m$ are order statistics drawn from $U[a_j, b_j]$, where $(a_j, b_j) \cap (a_k, b_k) = \emptyset, \forall j \neq k$. Since there is a specific order among the expected values in individual columns, we also see that $a_j \geq b_k, \forall j > k$.

The uniform distribution assumption allows the expected value of each cell in the matrix to have simple expressions, while the separation among the distributions causes the leximin assignment to be deterministic. For example, say we have the following cost matrix:

| | **1** | **2** | **3** | **...** | **n - 1** | **n** |
|---|---|---|---|---|---|---|
| **1** | $X_{1,(1)}$ | $X_{2,(1)}$ | $X_{3,(1)}$ | ... | $X_{n-1,(1)}$ | $X_{n,(1)}$ |
| **2** | $X_{1,(2)}$ | $X_{2,(2)}$ | $X_{3,(2)}$ | ... | $X_{n-1,(2)}$ | $X_{n,(2)}$ |
| **3** | $X_{1,(3)}$ | $X_{2,(3)}$ | $X_{3,(3)}$ | ... | $X_{n-1,(3)}$ | $X_{n,(3)}$ |
| **...** | ... | ... | ... | ... | ... | ... |
| **n - 1** | $X_{1,(n-1)}$ | $X_{2,(n-1)}$ | $X_{3,(n-1)}$ | ... | $X_{n-1,(n-1)}$ | $X_{n,(n-1)}$ |
| **n** | $X_{1,(n)}$ | $X_{2,(n)}$ | $X_{3,(n)}$ | ... | $X_{n-1,(n)}$ | $X_{n,(n)}$ |

Table 1: Cost matrix with order statistics

where $X_{j,(i)}$ is the $i^{th}$ smallest out of $n$ samples drawn from $U[a_j, b_j]$. Thus the expected value of $X_{j,(i)}$ can be computed:

$$\mathbb{E}[X_{j,(i)}] = a_j + (b_j - a_j)\frac{i}{n+1}$$

We now consider the expected cost of a given valid assignment $\Pi$, where agent $i$ is assigned to intervention $\Pi(i)$:

$$\begin{aligned}
\mathbb{E}[C(\Pi)] &= \mathbb{E}\left[\sum_i X_{\Pi(i),(i)}\right] \\
&= \sum_i \mathbb{E}\left[X_{\Pi(i),(i)}\right] \\
&= \sum_i a_{\Pi(i)} + (b_{\Pi(i)} - a_{\Pi(i)})\frac{i}{n+1} \\
&= \sum_i a_i + \frac{1}{n+1}\sum_i (b_{\Pi(i)} - a_{\Pi(i)})i
\end{aligned}$$

Note that $\sum_i a_i$ is fixed, while the sum $\sum_i (b_{\Pi(i)} - a_{\Pi(i)})i$ consists of pair-wise products of $(1, 2, ..., n)$ and a permutation of $(r_1, r_2, ..., r_n)$, where $r_j = b_j - a_j$. To minimize this expected assignment cost, we pair 1 with the largest $r_j$, 2 with the second largest $r_j$, and so on.

We denote $(r_{1'}, r_{2'}, ..., r_{n'})$ is the permutation where $r_j$'s are sorted in decreasing order:

$$r_{1'} \geq r_{2'} \geq ... \geq r_{n'}$$

So the minimized expected cost is:

$$\min_{\Pi} \mathbb{E}[C(\Pi)] = \sum_i a_i + \frac{1}{n+1}\sum_i i\ r_{i'}$$

The expected cost of the efficient assignment, or in other words, the expected minimum cost, is at most the minimum expected cost above. Additionally, the cost of any assignment is guaranteed to be larger than $\sum_i a_i$, so we have the following bounds for $\mathbb{E}[C(E)]$:

$$\mathbb{E}[C(E)] \leq \sum_i a_i + \frac{1}{n+1} \sum_i i \; r_{i'}$$

$$\mathbb{E}[C(E)] > \sum_i a_i$$

As for the leximin assignment, to minimize the largest cost, agent 1 needs to be assigned to intervention $n$ (any other agent would have a higher cost, thus making the assignment to have a higher bottleneck). Similarly, agent $i$ needs to be assigned to intervention $n - i + 1$, for all $i = 2, ..., n$. In short, the leximin assignment corresponds to the anti-diagonal ($AD$) assignment.

We can compute its expected cost:

$$\mathbb{E}[C(L)] = \mathbb{E}[C(AD)] = \sum_i a_i + \frac{1}{n+1} \sum_i i \; r_{n-i+1}$$

So the ratio $R = \mathbb{E}[C(E)]/\mathbb{E}[C(L)]$ is bounded as follows:

$$R < \frac{\sum_i a_i + \frac{1}{n+1} \sum_i i \; r_{n-i+1}}{\sum_i a_i} = 1 + \frac{\sum_i i \; r_{n-i+1}}{(n+1) \sum_i a_i}$$

$$R \geq \frac{\sum_i a_i + \frac{1}{n+1} \sum_i i \; r_{n-i+1}}{\sum_i a_i + \frac{1}{n+1} \sum_i i \; r_{i'}} = \frac{(n+1) \sum_i a_i + \sum_i i \; r_{n-i+1}}{(n+1) \sum_i a_i + \sum_i i \; r_{i'}}$$

If the number of intervention types $m < n$ and $c_j$ is the capacity for intervention type $j$, the two assignments behave similarly. Under leximin, the first $c_m$ agents with lowest cost will be assigned to intervention $m$, then the next $c_{m-1}$ agents will be assigned to intervention $m - 1$, and so on. For the assignment minimizing the expected cost, the first $c_{1'}$ agents will be assigned to intervention $1'$, then the next $c_{2'}$ will be assigned to intervention $2'$, and so on.

We thus have the following:

$$\begin{aligned}\mathbb{E}[C(L)] = \frac{1}{2(n+1)}[&r_m\ c_m(c_m+1)+\\ &r_{m-1}\ c_{m-1}(c_m+c_{m-1}+1)+\\ &r_{m-2}\ c_{m-2}(c_m+c_{m-1}+c_{m-2}+1)+\\ &...+\\ &+r_1\ c_1(\sum_j c_j+1)]+\sum_j c_j\ a_j\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[C(E)] \le \frac{1}{2(n+1)}[&r_{1'}c_{1'}(c_{1'}+1)+\\ &r_{2'}c_{2'}(c_{2'}+c_{2'}+1)+\\ &r_{(m-2)'}c_{(m-2)'}(c_{m'}+c_{(m-1)'}+c_{(m-2)'}+1)+\\ &...+\\ &r_{m'}c_{m'}(\sum_j c_{j'}+1)]+\sum_j c_j\ a_j\end{aligned}$$

From these formulas, a few notes can be made about the behavior of the contrast between $\mathbb{E}[C(L)]$ and $\mathbb{E}[C(E)]$:

- Keeping $n$, $m$, and the capacity constraints $c_j$ fixed, $\mathbb{E}[C(L)]$ grows larger than $\mathbb{E}[C(E)]$ when $(r_m, r_{m-1}, ..., r_1)$ is in reverse order of $(r_{1'}, r_{2'}, ..., r_{m'})$, or in other words, when $r_1 \ge r_2 \ge ... \ge r_m$. This is because under $L$, the aforementioned sum of pair-wise products will contain terms that are products of two large numbers.

  Roughly speaking, if there is more uncertainty in terms of cost from the more effective interventions than from the less effective interventions, the leximin assignment will be more likely to be costly.

- Keeping $n$, $m$, and the ranges $r_j$ fixed, the capacity constraints follow the same trend: if $c_1 \ge c_2 \ge ... \ge c_m$, the last few terms $r_i\ c_i(\sum_{j=m}^{i} c_j + 1)$ in $E[C(L)]$ will grow larger than their counterparts in $\mathbb{E}[C(E)]$.

  A possible interpretation is when the effective interventions are able to host many agents while the less effective ones are not, implementing leximin can lead to a much higher cost.

- We are also interested in the behavior of these two bounds as functions of the number of intervention types $m$. Due to the strong assumptions of this model, when $m$ becomes sufficiently large, the sum $\sum_j c_j\ a_j$ will dominate the two expressions, making the upper bound close to 1 (in most cases).

The following figure visualizes the behavior of the two bounds as a function of $m$ under different contexts: `Increasing ranges` denotes $2^{m-1}\ r_1 = 2^{m-2}\ r_2 = ... = r_m$, while `Decreasing ranges` denotes the opposite : $r_1 = 2r_2 = ... = 2^{m-1}\ r_m$; the same goes for the $c_j$'s in `Increasing capacities` and `Decreasing capacities`. The number of agents $n$, which is equal to $\sum_j c_j$, is fixed at 10,000.

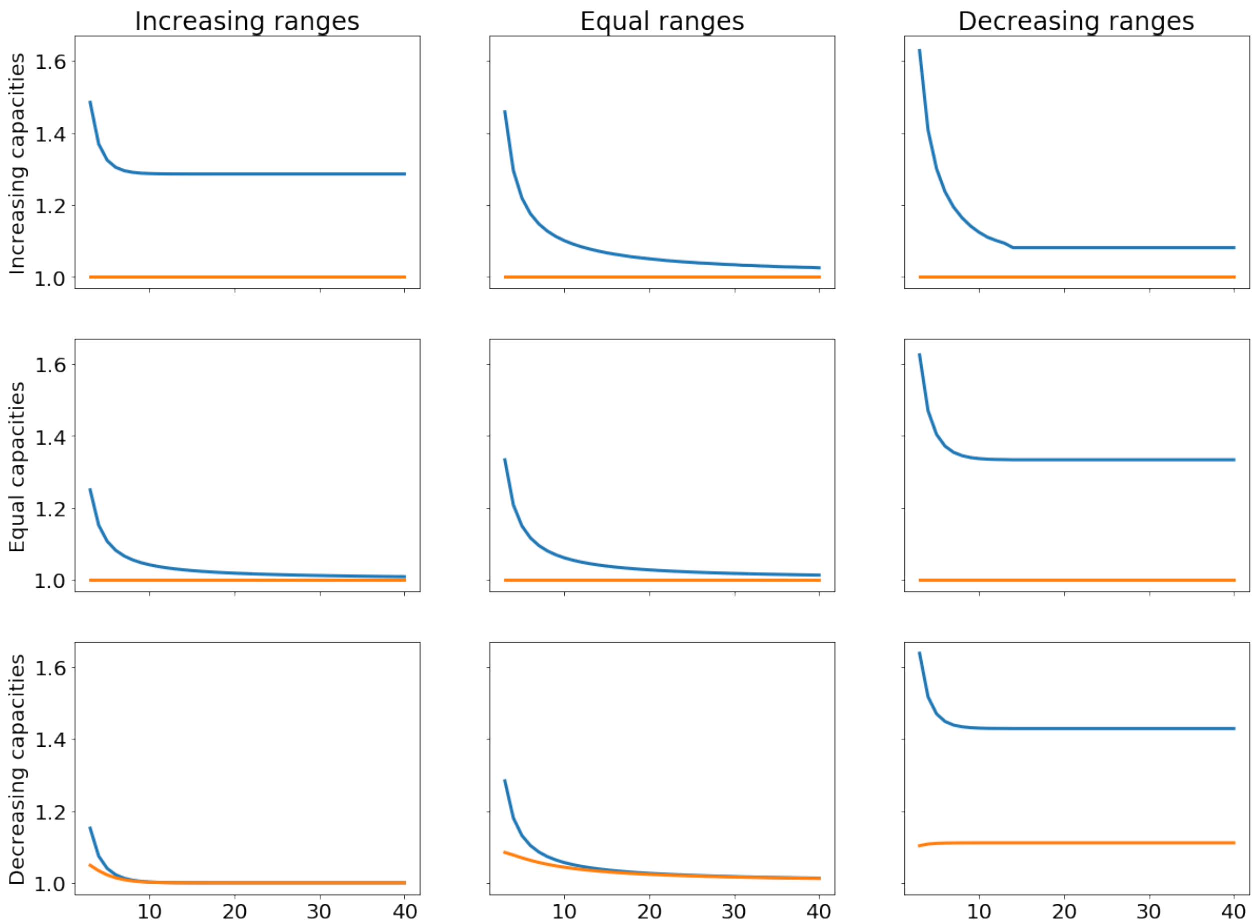


Figure 1: Behavior of PoF bounds

In many cases, as $m$ grows large, the two bounds converge near 1, indicating the low PoF for leximin. Again, this might be the artifact of the cost distributions being separated, casuing $\sum_j c_j\ a_j$ to be the dominating term in both of numerator and the denominator of each bound.

However, when the ranges decrease, the upper bound does not do this, and in the special case where the capacities also decrease, the lower bound for PoF stays higher than in other situations.

## 1.2 Leximin with general distributions

When the cost distributions of individual interventions are not separated, the leximin assignment and its cost cannot be approximated easily. One of the reasons for this is the leximin assignment only considers the relative ranking of the values in the cost matrix, as opposed to the actual values.

A potential idea is to simply assume that:

$$\mathbb{E}[X_{j,(i)}] \geq \mathbb{E}[X_{j-1,(i+1)}], \forall j > 1, i < n \tag{1}$$

In other words, the expected value of an element in the matrix is larger than the expected value of the element below it to the left (the next element in the anti-diagonal direction). We then approximate the leximin assignment using the $AD$ assignment again.

If the leximin assignment is not equivalent to $AD$ in some cost matrix, we can argue that there exists some element in the last column whose value is less than the value of an element on the anti-diagonal (otherwise $AD$ would have a better bottleneck/sub-bottleneck):

$$\begin{aligned}\mathbb{I}[L \not\equiv AD] &= \mathbb{I}[X_{n,(k)} \leq X_{j,(n-j+1]} \\ &= \mathbb{I}\left[X_{n,(k)} \leq \mathbb{E}[X_{n,(1)}] \ \vee \ X_{j,(n-j+1)} \geq \mathbb{E}[X_{n,(1)}]\right]\end{aligned}$$

So:

$$\begin{aligned}\Pr[L \not\equiv AD] &= \Pr\left[X_{n,(k)} \leq \mathbb{E}[X_{n,(1)}] \ \vee \ X_{j,(n-j+1)} \geq \mathbb{E}[X_{n,(1)}]\right] \\ &= \Pr\left[X_{n,(k)} \leq \mathbb{E}[X_{n,(1)}]\right] + \Pr\left[X_{j,(n-j+1)} \geq \mathbb{E}[X_{n,(1)}]\right]\end{aligned}$$

We can potentially upper bound this quantity using the Chernoff bound or Markov's inequality to argue that the probability that the leximin assignment deviates from $AD$ is not high under assumption (1). As long as leximin can be approximated using $AD$, we can inspect what special structure of the cost matrix will cause PoF to be large.

## 1.3 Distributions with "sharp turns"

We consider another type of matrices with a special structure that can cause PoF for leximin to be large. Say we have the following matrix:

| | **1** | **2** | **3** | **...** | **n - 1** | **n** |
|---|---|---|---|---|---|---|
| **1** | $\epsilon_{1,1}$ | $\epsilon_{1,2}$ | $\epsilon_{1,3}$ | ... | $\epsilon_{1,n-1}$ | $C_{1,n}$ |
| **2** | $\epsilon_{2,1}$ | $\epsilon_{2,2}$ | $\epsilon_{2,3}$ | ... | $C_{2,n-1}$ | $C_{2,n}$ |
| **3** | $\epsilon_{3,1}$ | $\epsilon_{3,2}$ | $\epsilon_{3,3}$ | ... | $C_{3,n-1}$ | $C_{3,n}$ |
| **...** | ... | ... | ... | ... | ... | ... |
| **n - 1** | $\epsilon_{n-1,1}$ | $C_{n-1,2}$ | $C_{n-1,3}$ | ... | $C_{n-1,n-1}$ | $C_{n-1,n}$ |
| **n** | $C_{n,1}$ | $C_{n,2}$ | $C_{n,3}$ | ... | $C_{n,n-1}$ | $C_{n,n}$ |

Table 2: Cost matrix with order statistics

where $\mathbb{E}[\epsilon_{i,j}] \ll \mathbb{E}[C_{k,l}]$. We further assume:

$$\gamma \max_{i,j} \mathbb{E}[\epsilon_{i,j}] \leq \min_{k,l} \mathbb{E}[C_{k,l}], \text{ for some } \gamma \gg 1$$

We also assume that the bottom-right element of the matrix is not much larger than the other $C_{i,j}$'s:

$$\mathbb{E}[C_{n,n}] \leq \lambda \min_{i,j} \mathbb{E}[C_{i,j}], \text{ for some } \lambda > 1$$

We again consider $AD$ as an approximation of the leximin assignment:

$$\begin{aligned}
\mathbb{E}[C(L)] &\approx \mathbb{E}[C(AD)] \\
&= \mathbb{E}\left[\sum_i C_{i,n-i+1}\right] \\
&= \sum_i \mathbb{E}[C_{i,n-i+1}] \\
&\geq n \min_{i,j} \mathbb{E}[C_{i,j}]
\end{aligned}$$

We also consider the assignment corresponding to the cells immediately above anti-diagonal combined with the bottom-right cell $AD'$:

$$\begin{aligned}
\mathbb{E}[C(E)] &\leq \mathbb{E}[C(AD')] \\
&= \mathbb{E}\left[C_{n,n} + \sum_{i=1}^{n-1} \epsilon_{i,n-i}\right] \\
&= \mathbb{E}[C_{n,n}] + \sum_{i=1}^{n-1} \mathbb{E}[\epsilon_{i,n-i}] \\
&\leq \lambda \min_{i,j} \mathbb{E}[C_{i,j}] + (n-1)\frac{1}{\gamma} \min_{i,j} \mathbb{E}[C_{i,j}]
\end{aligned}$$

So the ratio of the two expectations can be upper bounded as follows:

$$R \geq \frac{n}{\lambda + (n-1)\frac{1}{\gamma}} = \left(\frac{n}{\lambda\gamma + n - 1}\right)\gamma \approx \gamma$$

For our assumptions to hold, there needs to be a large decrease in expected value going from an element in the anti-diagonal to the element directly above it. In other words, the probability distribution of cost in intervention $j$ needs to have the expected values of its $(n-j)^{th}$ and $(n-j+1)^{th}$ order statistics far apart from each other.

Beta distributions with "U-shaped" density functions come to mind, but there does not seem to be analytical expressions for expected values of order statistics of a Beta distribution. (There is an R package that numerically computes these values[1].)

# 2 More on the real data set

We consider the original, unnormalized reentry probabilities in the homelessness data set. Specifically, we'd like to fit some probability distribution defined on $(0, 1)$ on the values in individual columns (probabilities from individual intervention types).

Here we consider two techniques: kernel density estimation using a Gaussian kernel and fitting a Beta distribution to the data (using MLE to find the parameters[2]). The results are visualized below:

[1]https://rdrr.io/cran/binGroup/man/beta.dist.html
[2]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.rv_continuous.fit.html

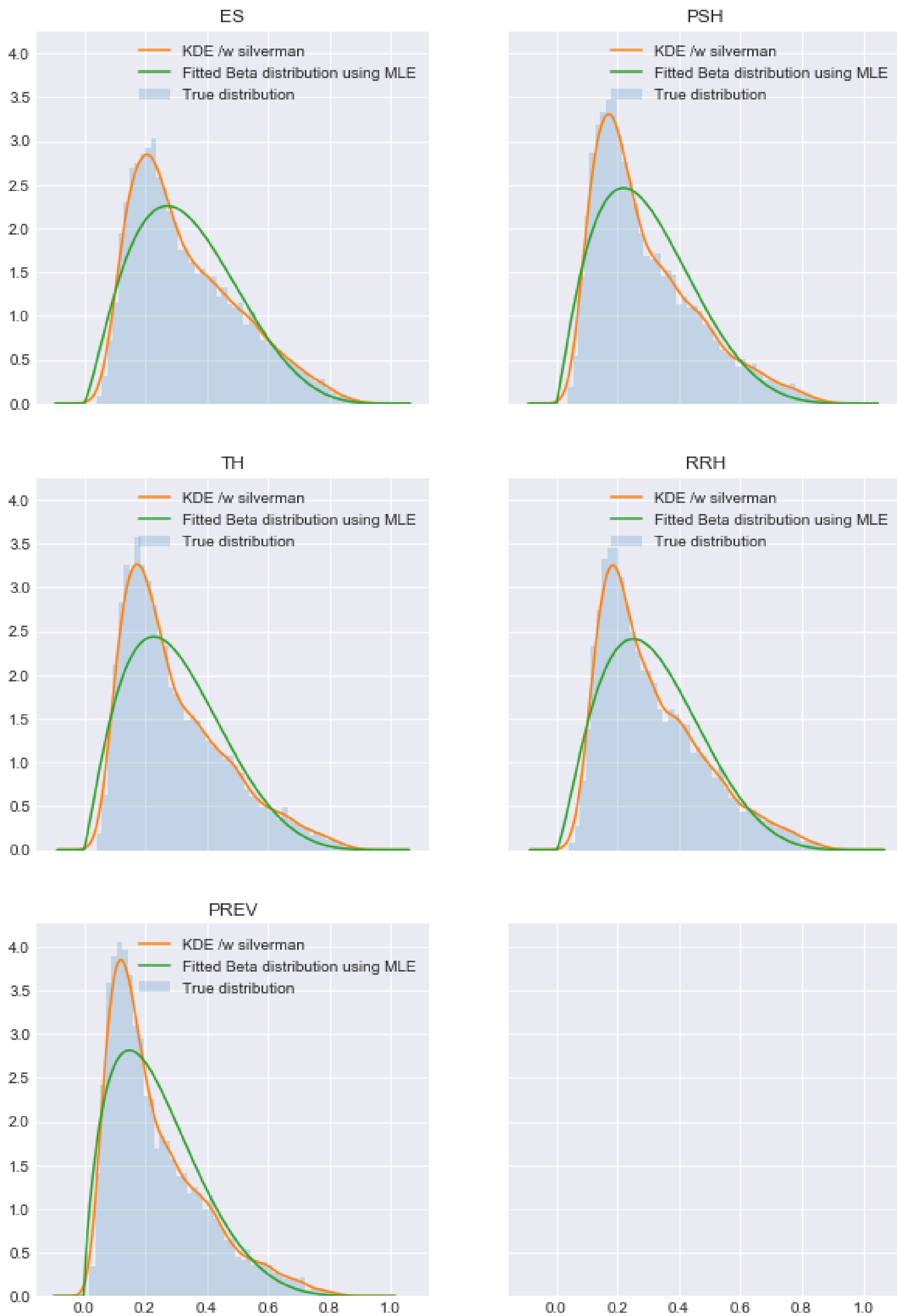


Figure 2: KDE vs. Fitting a Beta distribution

Additional notes:

- KDEs will need to be truncated to only give positive mass to points between 0 and 1.
- Fitting a mixture of Beta distributions will need to be implemented.

# References