Research Notes

October 1, 2020

Supply Chain Regression Analysis

Variable definitions:

- I: set of end nodes (pharmacy outlets); $|I| = n$
- J: set of intermediate nodes (suppliers); $|J| = m$
- π_i: Falsification probability for a product obtained from end node $i, i \in I$
- θ_j: Falsification probability for a product obtained from intermediate node $j, j \in I$
- $\boldsymbol{A} \in \mathbb{R}^{n \times m}$: Transition matrix between intermediate nodes and end nodes; entry (i, j) is the probability that end node i procures from intermediate node j
- $\dot{s}, \ddot{s}$: Sensitivity and specificity, respectively, of the diagnostic tool used during sampling
- $\boldsymbol{U}_i$: Total tests retrieved from end node $i, i \in I$
- $\boldsymbol{V}_i$: Total positive test results at end node $i, i \in I$
- $\hat{\boldsymbol{q}} \in \mathbb{R}^n$, with $\hat{\boldsymbol{q}} = \frac{V}{U}$: Vector of observed positive test proportions at each end node $i, i \in I$

Given a transition matrix, end node sampling results, and diagnostic sensitivity/specificity, what can we say about the falsification probabilities $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$?

We can solve a system of equations to derive an estimate of the intermediate node falsification probabilities, $\hat{\boldsymbol{\theta}}$, using the hat matrix of $\boldsymbol{A}$:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \hat{\boldsymbol{q}} \tag{1}$$

And then use $\hat{\boldsymbol{\theta}}$ to derive an estimate for $\boldsymbol{\pi}$:

$$\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{q}} - \boldsymbol{A}\hat{\boldsymbol{\theta}} \tag{2}$$

However, for any vector of observed positive test proportions, the total system, $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\pi}})$ is non-determined: for any $\hat{\boldsymbol{q}}$, there always exists a solution $(\hat{\boldsymbol{\theta}} = \boldsymbol{0}, \hat{\boldsymbol{\pi}} = \hat{\boldsymbol{q}})$.

Single-source procurement

Suppose that each end node i only procures from a single intermediate node j, such that I_j denotes the set of end nodes procuring from intermediate node j. Sorting the end nodes according to their intermediate supplier, the transition matrix $\boldsymbol{A}$ looks as follows:

$$\boldsymbol{A} = \begin{bmatrix} e_1 \\ \vdots \\ e_2 \\ \vdots \\ \vdots \\ e_m \\ \vdots \end{bmatrix}$$

Where each e_j is a standard basis vector of 0's with a 1 in the jth position. $\boldsymbol{A}^T\boldsymbol{A}$ becomes:

$$\boldsymbol{A}^T \boldsymbol{A} = diag(|I_1|, |I_2|, \ldots, |I_m|)$$

$$(\boldsymbol{A}^T\boldsymbol{A})^{-1} = diag(\frac{1}{|I_1|}, \frac{1}{|I_2|}, \dots, \frac{1}{|I_m|})$$

$$(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \begin{bmatrix} \frac{1}{|I_1|} & \frac{1}{|I_1|} & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & \frac{1}{|I_2|} & \frac{1}{|I_2|} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & \frac{1}{|I_m|} & \frac{1}{|I_m|} \end{bmatrix}$$

We observe that the product of the hat matrix and the falsification proportion observations results in an estimate $\hat{\boldsymbol{\theta}}$ that simply averages the observed proportions across the "children" of each intermediate node j.

Adjusted single-source procurement

Now suppose that each end node i procures from a primary intermediate j with probability $(1-\epsilon)$, and some secondary intermediate with probability ϵ. Let I_{ij} denote the set of end nodes with primary intermediate i and secondary intermediate j. If we retain a sorting similar to that described in the single-source case, $\boldsymbol{A}^T\boldsymbol{A}$ becomes a square matrix with $\sum_{k\in I_{ij},\forall j\in J}(1-\epsilon)^2 + \sum_{k\in I_{ji},\forall j\in J}\epsilon^2$ in the ith diagonal entry, and $\sum_{k\in I_{ij}}\epsilon(1-\epsilon) + \sum_{k\in I_{ji}}\epsilon(1-\epsilon)$ in the non-diagonal entries at row i, column j.

Balanced supply chain

Suppose further that the supply chain is perfectly balanced and non-decomposable, such that each intermediate node has an identical distribution of children of different procurement probabilities. We consider three scenarios: 1) The $(1-\epsilon), \epsilon$ case, where each end node procures from a primary intermediate node with probability $(1-\epsilon)$ and a secondary node with probability ϵ, 2) The $(1-\epsilon), \frac{\epsilon}{m-1}$ case, where each end node procures from a primary intermediate node with probability $(1-\epsilon)$ and every other intermediate node with probability $\frac{\epsilon}{m-1}$, and 3) The ρ-decay case, where for a given ρ, each end node procures from the jth-most frequently solicited intermediate node with probability $\frac{\rho^{j-1}}{(1-\rho^m)/(1-\rho)}$.

$(1-\epsilon), \epsilon$ case

For the $(1-\epsilon), \epsilon$ case, each intermediate node has $\frac{n}{m} \in \mathbb{Z}$ primary children and $\frac{n}{m(m-1)} \in \mathbb{Z}$ secondary children (so n must be a multiple of $m(m-1)$). Let a denote the ith diagonal entry of $\boldsymbol{A}^T\boldsymbol{A}$, where $a = \frac{n}{m}(1-\epsilon)^2 + \frac{n}{m}\epsilon^2 = \frac{n}{m}(1-2\epsilon+2\epsilon^2)$, and let b denote the (i,j)th non-diagonal of $\boldsymbol{A}^T\boldsymbol{A}$, where $b = \frac{2n}{m(m-1)}\epsilon(1-\epsilon) = \frac{n}{m(m-1)}(2\epsilon-2\epsilon^2)$.

Consider $(a-b)$, the "preference gap" between an intermediate node's primary children and any group of its secondary children who all possess the same primary preference. Minimizing $(a-b)$ with respect to ϵ yields:

$$a - b = \frac{n}{m}(1-2\epsilon+2\epsilon^2) - \frac{n}{m(m-1)}(2\epsilon-2\epsilon^2) = \frac{n}{m}[1-2(\frac{m}{m-1})\epsilon(1-\epsilon)] \tag{3}$$

$$\frac{d}{d\epsilon}(a-b) = 0 = \frac{-2}{m-1} - 2 + (\frac{4}{m-1}+4)\epsilon$$

$$\epsilon = \frac{1+\frac{1}{m-1}}{2+\frac{2}{m-1}} = \frac{1}{2}$$

which follows intuition. Importantly, note that when $\epsilon = \frac{1}{2}$:

$$a - b = \frac{n}{m}(1 - \frac{1}{2}\frac{m}{m-1})$$

which implies that $(a-b) \geq 0\,\forall m \geq 2$, which again is true for any supply chain we might consider.

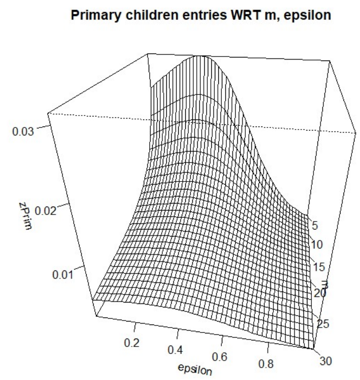

Figure 1: Relative change in entry values of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for primary children of intermediate node j with respect to m and ϵ

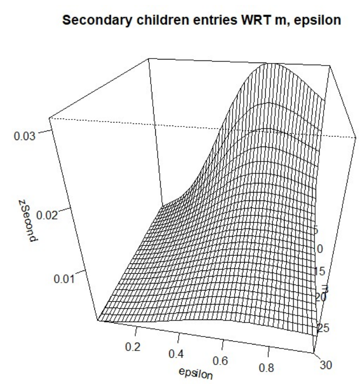

Figure 2: Relative change in entry values of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for secondary children of intermediate node j with respect to m and ϵ

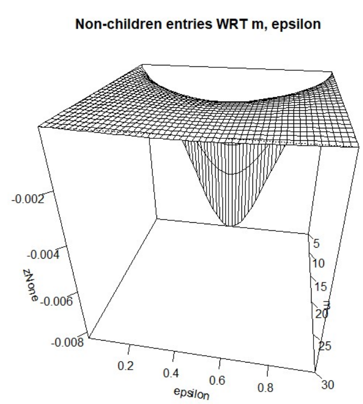

Figure 3: Relative change in entry values of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for non-children of intermediate node j with respect to m and ϵ

To generate the inverse of $\boldsymbol{A}^T\boldsymbol{A}$, consider that $\boldsymbol{A}^T\boldsymbol{A} = bP + (a-b)I$, where P is an all-ones matrix. It can be shown that the $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ is then $\frac{-b}{(a-b)(bm+a-b)}P + \frac{1}{a-b}I$. The diagonal entries of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ are then $\frac{m-1-2\epsilon+2\epsilon^2}{n(\frac{m-1}{m}-2\epsilon+2\epsilon^2)}$, and the non-diagonal entries become $\frac{-2\epsilon+2\epsilon^2}{n(\frac{m-1}{m}-2\epsilon+2\epsilon^2)}$.

The $m \times n$ matrix $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$, then, has three element types, where row i signifies an intermediate node, and column j signifies an end node:

1. j is a **primary** child of i: $\frac{(1-\epsilon)(m-1)-2\epsilon(1-\epsilon)}{n(\frac{m-1}{m}-2\epsilon(1-\epsilon))}$

2. j is a **secondary** child of i: $\frac{\epsilon(m-1)-2\epsilon(1-\epsilon)}{n(\frac{m-1}{m}-2\epsilon(1-\epsilon))}$

3. j is a **no** child of i: $\frac{-2\epsilon(1-\epsilon)}{n(\frac{m-1}{m}-2\epsilon(1-\epsilon))}$

See Figures 1, 2 and 3. As anticipated, we observe that values of ϵ nearer to 0.5 increase the weight in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ given to secondary children. Interestingly, these same ϵ values result in a greater off-setting/negative effect for non-children. This effect translates to a tendency to reduce the $\hat{\theta}_i$ estimate for intermediate node i when the supply chain becomes significantly convoluted. The result is that the $\hat{\boldsymbol{\theta}}$ values tend to equalize with ϵ values near 0.5.

These effects become drastically less pronounced as m, the number of intermediate nodes, increases. However, bearing in mind that we are analyzing a supply chain where each end node receives only 2 preferences, it seems intuitive that increasing the availability of preferences, in conjunction with m, should result in correspondingly more dramatic curves for the elements of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$. Importantly, we note that there exists an overall *interactive* impact of ϵ and m on $\hat{\boldsymbol{\theta}}$.

$(1-\epsilon)$, $\frac{\epsilon}{m-1}$ case

In the $(1-\epsilon)$, $\frac{\epsilon}{m-1}$ case, each end node procures from a primary intermediate node with probability $(1-\epsilon)$ and every other intermediate node with probability $\frac{\epsilon}{m-1}$, where $\epsilon \neq \frac{m-1}{m}$ (details below). Observe that for $\epsilon > \frac{m-1}{m}$, the "primary" intermediate node is then procured from less than all other suppliers. Each intermediate node has $\frac{n}{m} \in \mathbb{Z}$ primary children and $\frac{n}{m}(m-1) \in \mathbb{Z}$ secondary children (so now n need only be a multiple of m).

As with the previous case, letting a denote the ith diagonal entry of $\boldsymbol{A}^T\boldsymbol{A}$, with $a = \frac{n}{m}(1-2\epsilon+\frac{m}{m-1}\epsilon^2)$, and letting b denote the (i,j)th non-diagonal of $\boldsymbol{A}^T\boldsymbol{A}$, with $b = \frac{n}{m(m-1)}[2\epsilon - \frac{m}{(m-1)}\epsilon^2]$, allows the calculation of the entries of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ as follows:

- **Diagonal**: $\frac{1-\frac{2}{m-1}\epsilon+\frac{m}{(m-1)^2}\epsilon^2}{\frac{n}{m}\left[1-\frac{2m}{m-1}\epsilon+\frac{m^2}{(m-1)^2}\epsilon^2\right]}$

- **Non-Diagonal**: $\frac{\frac{-2}{m-1}\epsilon+\frac{m}{(m-1)^2}\epsilon^2}{\frac{n}{m}\left[1-\frac{2m}{m-1}\epsilon+\frac{m^2}{(m-1)^2}\epsilon^2\right]}$

Figures 4 and 5 illustrate these entries. Compare these expressions with those for the $(1-\epsilon), \epsilon$ case and note they align for $m = 2$: $\frac{m-1-2\epsilon+2\epsilon^2}{n(\frac{m-1}{m}-2\epsilon+2\epsilon^2)}$ and $\frac{-2\epsilon+2\epsilon^2}{n(\frac{m-1}{m}-2\epsilon+2\epsilon^2)}$ for the diagonal and non-diagonal entries, respectively. Additionally, observe that the effect of the non-diagonal entries of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ diminishes with increasing m, and that the weight of the diagonal entry decreases with increasing ϵ, as we should anticipate. Importantly, $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ is undefined when $1 - \frac{2m}{m-1}\epsilon + \frac{m^2}{(m-1)^2}\epsilon^2 = 0$, i.e., when $\epsilon = \frac{m-1}{m}$. Note that in the case of $\epsilon = \frac{m-1}{m}$, each end node procures from every supplier node with probability $\frac{1}{m}$, indicating a perfectly distributed likelihood across suppliers, and implying that designating the effect of falsification detection at any end node upon a supplier node would assign likelihood evenly across all suppliers. At values near $\epsilon = \frac{m-1}{m}$, we see strongly positive/negative values for diagonal/non-diagonal entries of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$, which suggests much more volatility of $\hat{\boldsymbol{\theta}}$ estimates with respect to detection at end nodes.

In contrast with the $(1-\epsilon), \epsilon$ case, because every end node procures from every supplier, the $m \times n$ matrix $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ only has two element types for row i and column j, corresponding to supplier i and end node j, as follows:

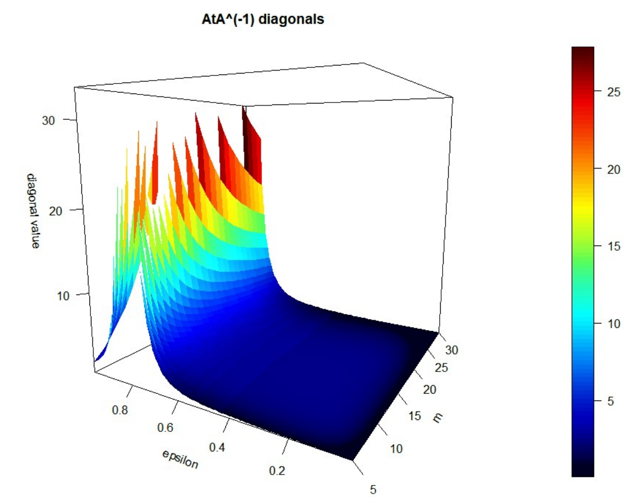

Figure 4: Relative change in diagonal values of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ with respect to m and ϵ for the $(1-\epsilon)$, $\frac{\epsilon}{m-1}$ case

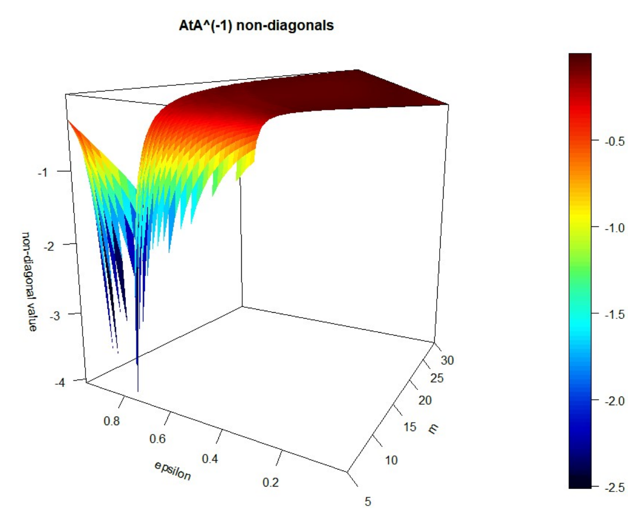

Figure 5: Relative change in non-diagonal values of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ with respect to m and ϵ for the $(1-\epsilon)$, $\frac{\epsilon}{m-1}$ case

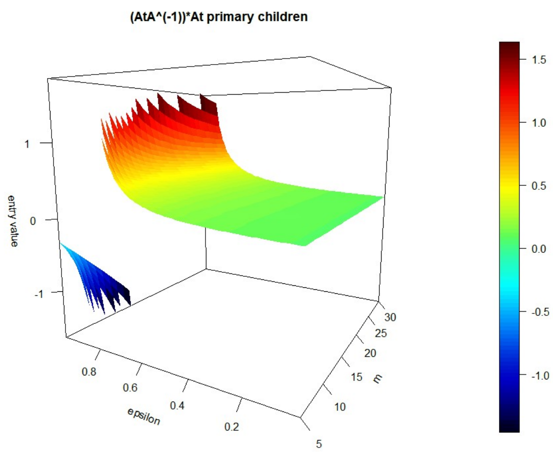

Figure 6: Relative change in primary children values of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ with respect to m and ϵ for the $(1-\epsilon)$, $\frac{\epsilon}{m-1}$ case

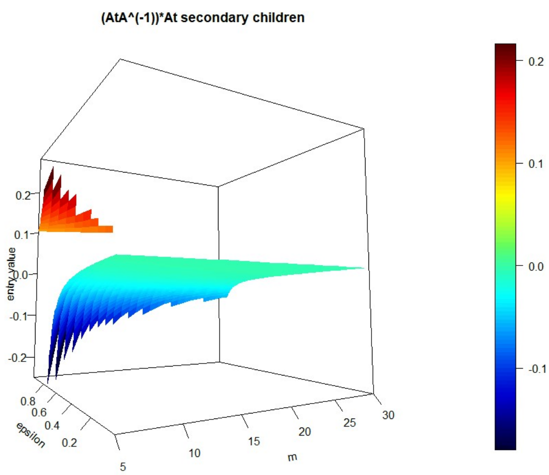

Figure 7: Relative change in secondary children values of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ with respect to m and ϵ for the $(1-\epsilon)$, $\frac{\epsilon}{m-1}$ case

- j is a primary child of i: $\frac{(1-\epsilon)-\frac{2\epsilon}{m-1}+\frac{m\epsilon^2}{(m-1)^2}}{\frac{n}{m}[1-\frac{2m\epsilon}{m-1}+\frac{m^2\epsilon^2}{(m-1)^2}]}$
- j is a secondary child of i: $\frac{\frac{-\epsilon}{m-1}+\frac{m\epsilon^2}{(m-1)^2}}{\frac{n}{m}[1-\frac{2m\epsilon}{m-1}+\frac{m^2\epsilon^2}{(m-1)^2}]}$

Figures 6 and 7 illustrate these entries of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$. As one might anticipate, more balanced distributions of procurement likelihood across intermediate nodes translates to much more volatility in computing $\hat{\boldsymbol{\theta}}$ estimates. Under these ϵ conditions, $\hat{\boldsymbol{\theta}}$ is subject to significant adjustments upon the detection of falsification at any end node.

ρ-decay case

In the ρ-decay case, for a given ρ each end node procures from the jth-most frequently solicited intermediate node with probability $\rho^{j-1}\frac{1-\rho}{1-\rho^m}$. Low ρ values correspond with strong preferences for few suppliers, while high ρ values correspond with more even procurement distributions.

To construct the $\boldsymbol{A}$ matrix, note that each row corresponding to an end node i is a permutation γ_i of the row $\frac{1-\rho}{1-\rho^m}\begin{bmatrix}\rho^0 & \rho^1 & \dots & \rho^{m-1}\end{bmatrix}$. Since we have n rows each summing to 1, each column of $\boldsymbol{A}$ must sum to $\frac{n}{m}$. Combinatorially, there are many ways to construct such a matrix (on the order of $O((n!)^m)$). We consider matrices where $n = m(m-1)$, so that $\boldsymbol{A}$ consists of blocks of $\frac{n}{m}$ rows, where each block consists of a column of ρ^0 entries and $(m-1)$ columns of a doubly-stochastic permutation matrix of size $(m-1)\times(m-1)$. For the entries of $\boldsymbol{A}^T\boldsymbol{A}$ to be balanced, it must be the case that there are $\frac{n}{m}$ of each $\rho^k, k \in \{0,1,...,m-1\}$ in each column of $\boldsymbol{A}$. $\boldsymbol{A}$ appears as follows:

$$\boldsymbol{A} = \frac{1-\rho}{1-\rho^m}\begin{bmatrix}
\rho^0 & \rho^1 & \dots & \rho^{m-2} & \rho^{m-1} \\
\rho^0 & \rho^2 & \dots & \rho^{m-1} & \rho^1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\rho^0 & \rho^{m-1} & \dots & \rho^1 & \rho^{m-2} \\
\\
\rho^1 & \rho^0 & \dots & \rho^{m-2} & \rho^{m-1} \\
\rho^2 & \rho^0 & \dots & \rho^{m-1} & \rho^1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\rho^{m-1} & \rho^0 & \dots & \rho^1 & \rho^{m-2} \\
\\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\\
\rho^{m-1} & \rho^{m-2} & \dots & \rho^1 & \rho^0 \\
\rho^1 & \rho^{m-1} & \dots & \rho^2 & \rho^0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\rho^{m-2} & \rho^1 & \dots & \rho^{m-1} & \rho^0
\end{bmatrix}$$

With this structure, any two columns of $\boldsymbol{A}$ are aligned such that, pairwise, there exists two of each pair $\rho^i, \rho^j, \forall i \neq j$. The resulting entries of $\boldsymbol{A}^T\boldsymbol{A}$ are as follows:

- **Diagonal**: $\frac{(1-\rho)^2}{(1-\rho^m)^2}\cdot\frac{n}{m}\sum_{j=0}^{m-1}\rho^{2j}$
- **Non-Diagonal**: $\frac{(1-\rho)^2}{(1-\rho^m)^2}\cdot 2\sum_{\substack{i<j \\ i,j\in\{0,1,\dots,m-1\}}}\rho^i\rho^j$

Denote $S_1 = \sum_{j=0}^{m-1}\rho^{2j}$ and $S_2 = 2\sum_{\substack{i<j \\ i,j\in\{0,1,\dots,m-1\}}}\rho^i\rho^j$. Figures 8 and 9 depict these summations with respect to m and ρ. Note that both S_1 and S_2 increase approximately linearly in m and exponentially in ρ. Following the ideas of the previous cases yields the following for the entries of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$, illustrated in Figures 10 and 11:

- **Diagonal**: $\frac{(m-2)S_2+\frac{n}{m}S_1}{(\frac{n}{m})^2(S_1)^2+\frac{n(m-2)}{m}(S_1)(S_2)-(m-1)(S_2)^2}\cdot\frac{(1-\rho^m)^2}{(1-\rho)^2}$
- **Non-Diagonal**: $\frac{-S_2}{(\frac{n}{m})^2(S_1)^2+\frac{n(m-2)}{m}(S_1)(S_2)-(m-1)(S_2)^2}\cdot\frac{(1-\rho^m)^2}{(1-\rho)^2}$

We can take advantage of the fact that $n = m(m-1)$ to simplify these entries for $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$:

- **Diagonal**: $\frac{\frac{m-2}{m-1}S_2+S_1}{[(m-1)S_1-S_2][S_1+S_2]}\cdot\frac{(1-\rho^m)^2}{(1-\rho)^2}$

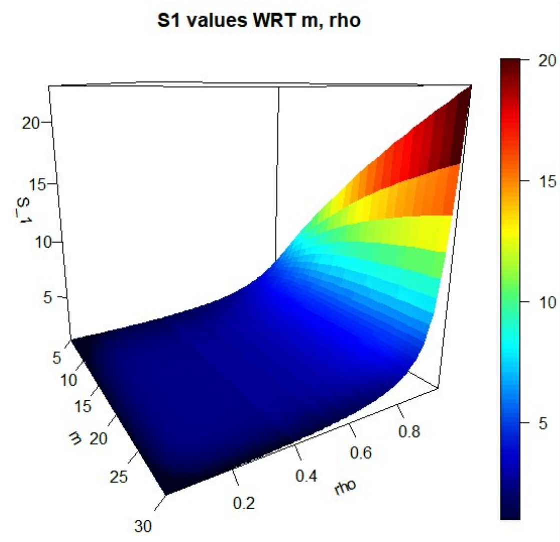

Figure 8: Relative change in S_1 values with respect to m and ρ

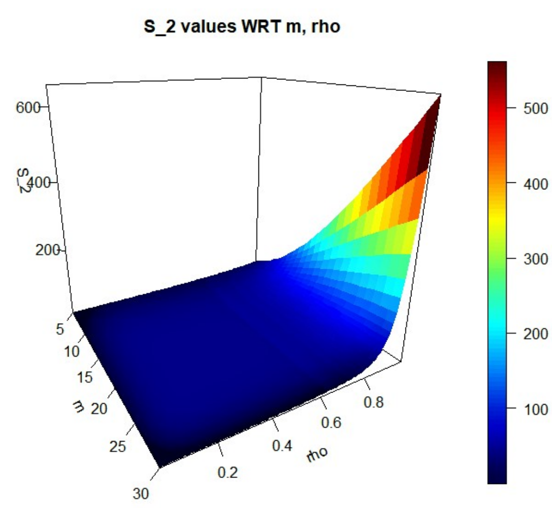

Figure 9: Relative change in S_1 values with respect to m and ρ

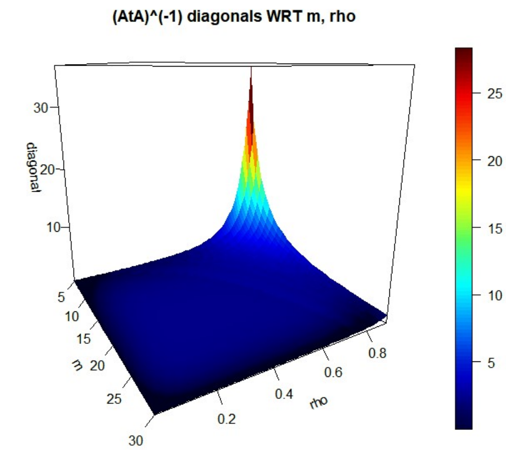

Figure 10: Relative change in diagonals of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ with respect to m and ρ

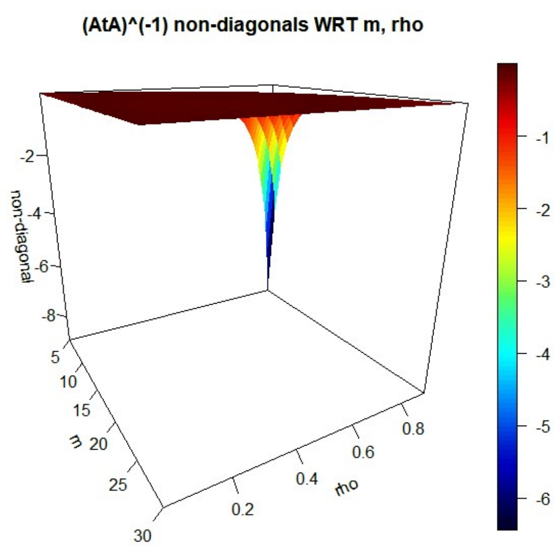

Figure 11: Relative change in non-diagonals of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ with respect to m and ρ

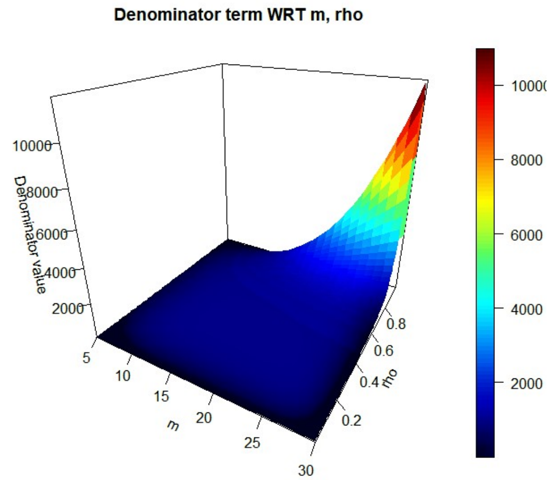

Figure 12: Relative change in denominator term of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ with respect to m and ρ

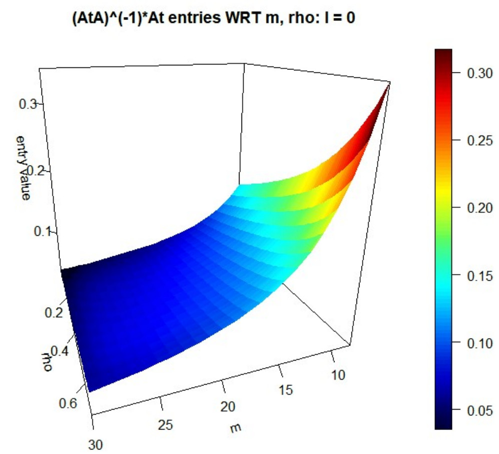

Figure 13: Relative change in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for a 1st child with respect to m and ρ

- **Non-Diagonal**: $\frac{\frac{-S_2}{m-1}}{[(m-1)S_1-S_2][S_1+S_2]} \cdot \frac{(1-\rho^m)^2}{(1-\rho)^2}$

Figure 12 depicts the denominator term of the $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ entries. Observe that this term is never less than 0.

We then have m types of entries for $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$, where row i corresponds to a supplier node and column j corresponds to an end node. Somewhat differently than the $1-\epsilon, \epsilon$ and $1-\epsilon, \frac{\epsilon}{m-1}$ cases, the $(l+1)$th child of a supplier node refers to an end node with procurement likelihood $\rho^l \frac{1-\rho}{1-\rho^m}, \forall l \in \{0, 1, ..., m-1\}$. For supplier i, an $(l+1)$th child j has a corresponding entry in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ as follows:

$$\frac{\rho^l \frac{1-\rho}{1-\rho^m}(S_1+S_2) - \frac{S_2}{m-1}}{[(m-1)S_1-S_2][S_1+S_2]} \cdot (\frac{1-\rho^m}{1-\rho})^2 \tag{4}$$

Figures 13, 14, 15, 16 and 17 depict how these entries change with respect to m and ρ for $l \in \{0, 1, 2, 3, m\}$. Note that for many combinations of m and ρ, the projection matrix terms are negative for $l \geq 2$. From 4, the projection matrix terms are negative whenever $\rho^l < \frac{S_2 \frac{1-\rho^m}{1-\rho}}{(m-1)(S_1+S_2)}$.

Additionally, note in the projection matrix figures that the ranges of m and ρ are restricted in order to depict the interesting regions of the space. For large values of ρ, the entries of $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ are generally

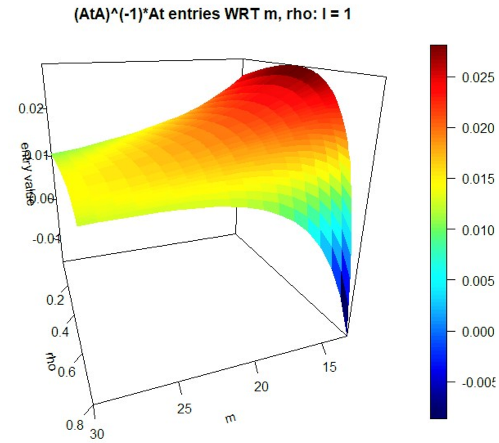

Figure 14: Relative change in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for a 2nd child with respect to m and ρ

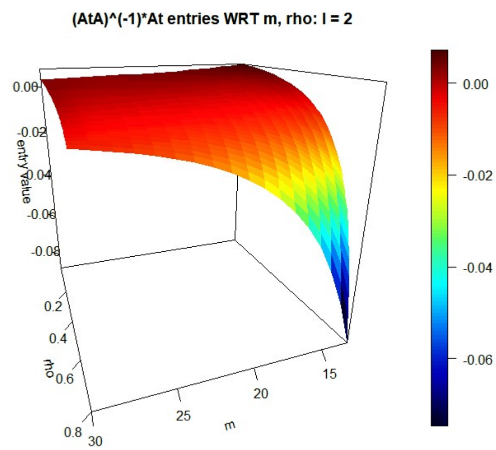

Figure 15: Relative change in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for a 3rd child with respect to m and ρ

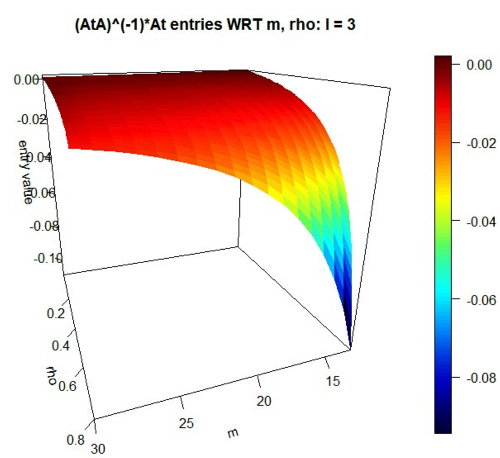

Figure 16: Relative change in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for a 4th child with respect to m and ρ

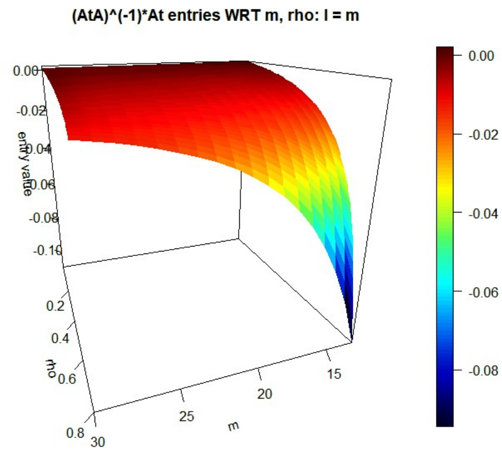

Figure 17: Relative change in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ for a mth child with respect to m and ρ

very negative or very positive. As with the $(1-\epsilon)$, $\frac{\epsilon}{m-1}$ case, highly distributed procurement likelihoods can cause very high variability in $\hat{\boldsymbol{\theta}}$ estimates as falsification samples are detected.

Unbalanced supply chain

Note that $\boldsymbol{A}^T\boldsymbol{A}$ can be formed iteratively by the following process. Initialize $\alpha \leftarrow u_1 \otimes u_1$, where $u_1 = (1-\epsilon)e_i + \epsilon e_j$, with i and j denoting the primary and secondary intermediate nodes for the first end node under consideration. A $(1-\epsilon)^2$ term is added to the ith diagonal, a ϵ^2 term is added to the jth diagonal, and $\epsilon(1-\epsilon)$ terms are added to the (i,j)th and (j,i)th locations of α. Form u_2 similarly, and set $\alpha \leftarrow \alpha + (u_2 \otimes u_2)$. Iterating through all $k \in I$, the resulting α is then $\boldsymbol{A}^T\boldsymbol{A}$.

DOESN'T WORK - A NEEDS TO BE INVERTIBLE AT EVERY STEP

Try initializing $\alpha = I$ ("dummy" end node children) and removing the children at the end (how to adjust at the end?). After one step, where $u_1 = (1-\epsilon)e_{i_1} + \epsilon e_{j_1}$, we get:

$$\alpha_1 \leftarrow I + u_1 \otimes u_1 = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots \\ 0 & 1 & \dots & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1+(1-\epsilon)^2 & \dots & \epsilon(1-\epsilon) & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \epsilon(1-\epsilon) & \dots & 1+\epsilon^2 & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & 1 \end{bmatrix}$$

Using Sherman-Morrison to find the inverse gives us:

$$\alpha_1^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots \\ 0 & 1 & \dots & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1-\frac{(1-\epsilon)^2}{2-2\epsilon+2\epsilon^2} & \dots & \frac{-\epsilon(1-\epsilon)}{2-2\epsilon+2\epsilon^2} & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{-\epsilon(1-\epsilon)}{2-2\epsilon+2\epsilon^2} & \dots & 1-\frac{\epsilon^2}{2-2\epsilon+2\epsilon^2} & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & 1 \end{bmatrix}$$

For α_2, we have 7 potential cases *[A lot of algebra involved for the other cases, does it make sense to follow them all?]*, depending on whether $i_2, j_2 \in \{i_1, j_1, J \setminus \{i_1, j_1\}\}$.

- **Case 1**: The simplest case is when $i_2, j_2 \in \{J \setminus \{i_1, j_1\}\}$, where only the $\{i_2, j_2\}$ indices change:

$$\alpha_2^{-1} = \begin{array}{cc} & \begin{matrix} & & & i_1 & & j_1 & & i_2 & & j_2 & & \end{matrix} \\ \begin{matrix} \\ \\ \\ i_1 \\ \\ j_1 \\ \\ i_2 \\ \\ j_2 \\ \\ \\ \end{matrix} & \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1-\frac{(1-\epsilon)^2}{2-2\epsilon+2\epsilon^2} & \dots & \frac{-\epsilon(1-\epsilon)}{2-2\epsilon+2\epsilon^2} & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{-\epsilon(1-\epsilon)}{2-2\epsilon+2\epsilon^2} & \dots & 1-\frac{\epsilon^2}{2-2\epsilon+2\epsilon^2} & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & \dots & 1-\frac{(1-\epsilon)^2}{2-2\epsilon+2\epsilon^2} & \dots & \frac{-\epsilon(1-\epsilon)}{2-2\epsilon+2\epsilon^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & \dots & \frac{-\epsilon(1-\epsilon)}{2-2\epsilon+2\epsilon^2} & \dots & 1-\frac{\epsilon^2}{2-2\epsilon+2\epsilon^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \end{array}$$

Note that the update terms have limited ranges, as $\epsilon \in (0, 1)$.
Formulas to play with:
Woodbury:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

Sherman-Morrison:

$$(A + uu^T)^{-1} = A^{-1} - \frac{A^{-1}uu^TA^{-1}}{1 + u^TA^{-1}u}$$

Transition Matrix Metrics

Given a transition matrix, $\boldsymbol{A}$, we can form metrics regarding the dispersal and identifiability levels.

$tr(\boldsymbol{A}^T\boldsymbol{A})$ can be used as a measure of the dispersal, or distribution, of procurement probabilities across outlets. The squared nature of $\boldsymbol{A}^T\boldsymbol{A}$ means that the entries are dominated by the maximum row values (recall that each row of $\boldsymbol{A}$ sums to 1), so the trace is a good measure of how dependent a normal outlet is on its primary importer. The trace is bounded above by n, and bounded below by $m\frac{1}{m^2}n = \frac{n}{m}$.

$det(\boldsymbol{A}^T\boldsymbol{A})$ is a related measure to $tr(\boldsymbol{A}^T\boldsymbol{A})$, but is more oriented towards how discernible importers are vis a vis outlets.

Alternative Estimation Models

The inputs used for estimating $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\pi}}$ are as follows:

- $\boldsymbol{A} \in \mathbb{R}^{n\times m}$: The (estimated) transition matrix from intermediate nodes to end nodes
- $\boldsymbol{U}_i$: Total tests retrieved from end node $i, i \in I$
- $\boldsymbol{V}_i$: Total positive test results at end node $i, i \in I$
- $\hat{\boldsymbol{q}} \in \mathbb{R}^n$, with $\hat{\boldsymbol{q}} = \frac{V}{U}$: Vector of observed positive test proportions at each end node $i, i \in I$
- $\dot{s}, \ddot{s}$: Sensitivity and specificity, respectively, of the diagnostic tool used during sampling

Logistic regression

We can form Bernoulli logistic regression estimates using iteratively re-weighted least squares to maximize the log-likelihood. Let $\hat{\boldsymbol{\beta}}$ signify the coefficients corresponding to the importer SF effects within the logit function (as opposed to $\hat{\boldsymbol{\theta}}$). The $\hat{\boldsymbol{\beta}}$ estimates are found using the following algorithm at step k:

$$\hat{\boldsymbol{\beta}}_{k+1} = (\boldsymbol{A}^T\boldsymbol{S}_k\boldsymbol{A})^{-1}\boldsymbol{A}^T(\boldsymbol{S}_k\boldsymbol{A}\hat{\boldsymbol{\beta}}_k + \hat{\boldsymbol{q}} - \boldsymbol{\mu}_k)$$

where $\boldsymbol{\mu}_i = \frac{1}{1+e^{-\hat{\boldsymbol{\beta}}^T\boldsymbol{A}_i}}$ signifies the mean SF effect estimate from all importers on outlet i, and $\boldsymbol{S}$ is a diagonal matrix with entry i equal to $\boldsymbol{\mu}_i(1-\boldsymbol{\mu}_i)$, the variance of the current mean SF effect estimate for outlet i. We can then calculate $\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{q}} - \boldsymbol{\mu}$. This modeling approach typically produces estimates very close to those produced via the linear projection, and as such does not result in any improvements in estimate variance.

Maximizing likelihood

Alternatively, we can use the Bernoulli model to generate likelihood estimates. We retain the logit link function, where β_j denotes the coefficient for importer $j \in J$ and β_i denotes the coefficient for outlet $i \in I$. First, we establish the underlying outlet SF probability, $\boldsymbol{\phi}$, as a function of importer and outlet SF tendencies:

$$\boldsymbol{\phi} = \left[1 - \frac{e^{\beta_j}}{e^{\beta_j}+1}\right]_{n\times 1} \bullet \boldsymbol{A}\left[\frac{e^{\beta_i}}{e^{\beta_i}+1}\right]_{m\times 1} + \left[\frac{e^{\beta_j}}{e^{\beta_j}+1}\right]_{n\times 1}$$

Where the first term denotes the probability of procuring an SFP via the importer, and the second term denotes the probability of the outlet itself procuring an SFP. We then adjust $\boldsymbol{\phi}$ for the sensitivity and specificity of our diagnostics:

$$\dot{\boldsymbol{\phi}} = \dot{s}\boldsymbol{\phi} + (1-\ddot{s})(1-\boldsymbol{\phi})$$

Finally, we derive the likelihood and log-likelihood via the binomial likelihood function:

$$L = \prod_{j\in J} \dot{\boldsymbol{\phi}}_j^{V_j}(1-\dot{\boldsymbol{\phi}}_j)^{U_j - V_j}$$

$$l = \sum_{j\in J} V_j \ln \dot{\boldsymbol{\phi}}_j + (U_j - V_j)\ln(1-\dot{\boldsymbol{\phi}}_j)$$

We add a regularization term $\bar{w} \sum_{j \in J} |\beta_j - \beta_{j,0}|$ to the likelihood function before maximizing, where $\bar{w}$ is a regularization weight, and $\beta_{j,0}$ signifies an initial value for the SFP contribution of outlet $j, j \in J$. We then use the L-BFGS-B method of `scipy.optimize` to maximize the regularized likelihood function.

Posterior sampling with NUTS

The gradient of the likelihood function with respect to a given $\boldsymbol{\beta}$ can be expressed in terms of β_j and β_i:

$$\frac{\partial}{\partial \beta_j} \boldsymbol{\phi} = -\frac{e^{\beta_j}}{(e^{\beta_j}+1)^2} A_j \left[\frac{e^{\beta_i}}{e^{\beta_i}+1}\right]_{m \times 1} + \frac{e^{\beta_j}}{(e^{\beta_j}+1)^2}$$

$$\frac{\partial}{\partial \beta_j} \dot{\boldsymbol{\phi}} = \dot{s} \frac{\partial}{\partial \beta_j} \boldsymbol{\phi} - (1-\ddot{s})(1 - \frac{\partial}{\partial \beta_j} \boldsymbol{\phi})$$

$$\frac{\partial}{\partial \beta_j} l = \frac{y_j}{\boldsymbol{\phi}_j} \frac{\partial}{\partial \beta_j} \dot{\boldsymbol{\phi}} - \frac{n_j - y_j}{1 - \boldsymbol{\phi}_j} \frac{\partial}{\partial \beta_j} \dot{\boldsymbol{\phi}}$$

Posterior samples as a function of $\boldsymbol{A}$, $\boldsymbol{A}$, and $\boldsymbol{A}$ can be drawn using the likelihood function for $\boldsymbol{\beta}$, along with this likelihood gradient, via the No U-Turn Sampler (NUTS) described by [1]. The NUTS algorithm requires a number of burn-in samples (M_{adapt}) as well as a target rejection rate (δ) before returning a number of samples (M) from the posterior distribution. The means of these generated distributions can then be used to provide $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\pi}}$.

ANOVA

First we consider the variance of the observation vector, $\hat{\boldsymbol{q}}$. Each element $\hat{q}_i$ is the result of a binomial trial with t_i samples, a probability q_i of a positive test, and an underlying probability π_i of a falsified sample, where π_i is sampled from some distribution $\Pi(\cdot), \forall i \in I$, with mean μ_Π and variance σ^2_Π. Taking into account the diagnostic sensitivity and specificity, where $\dot{s}$ denotes the sensitivity (bounded below by 0.5) and $\ddot{s}$ denotes 1 less the specificity (bounded above by 0.5), we calculate q_i as:

$$q_i = \dot{s}\pi_i + \ddot{s}(1-\pi_i) = \ddot{s} + \pi_i(\dot{s} - \ddot{s})$$

The variance of each observed $\hat{q}_i$ is then:

$$\begin{aligned}\mathbb{V}[\hat{q}_i|\boldsymbol{\pi} \in \Pi] = \frac{q_i(1-q_i)}{t_i} &= \frac{2\pi_i^2\dot{s}\ddot{s} + \dot{s}\pi_i + 2\ddot{s}^2\pi_i - \dot{s}\pi_i^2 - \ddot{s}\pi_i^2 - 2\dot{s}\ddot{s}\pi_i - \ddot{s}\pi_i - \ddot{s}^2 + \ddot{s})}{t_i} \\ &= \frac{\ddot{s}(1-\ddot{s}) + 2\dot{s}\ddot{s}\pi_i(\pi_i - 1) + \dot{s}\pi_i(1-\pi_i) + 2\ddot{s}^2\pi_i - \ddot{s}\pi_i(\pi_i - 1)}{t_i}\end{aligned}$$

With perfect specificity ($\ddot{s} = 0$), we obtain:

$$\mathbb{V}[\hat{q}_i|\boldsymbol{\pi} \in \Pi] = \frac{\dot{s}\pi_i(1-\pi_i)}{t_i}$$

With perfect sensitivity ($\dot{s} = 1$), we obtain:

$$\mathbb{V}[\hat{q}_i|\boldsymbol{\pi} \in \Pi] = \frac{\ddot{s}(1-\ddot{s}) + \pi_i(1-\pi_i)(1-\ddot{s}) + 2\ddot{s}^2\pi_i}{t_i}$$

The variance of $\hat{\boldsymbol{\theta}}$ then becomes, from Equation 1:

$$\mathbb{V}[\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{q}}] = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\Sigma(\hat{\boldsymbol{q}})((\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T)^T$$

The mean square error (MSE) term:

$$MSE_{\hat{\boldsymbol{q}}} = \frac{||\hat{\boldsymbol{\pi}}||^2}{n-(m+1)} = \frac{||\hat{\boldsymbol{q}} - \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\hat{\boldsymbol{q}}||^2}{n-(m+1)}$$

being an unbiased estimator of $\mathbb{V}[\boldsymbol{q}] = \mathbb{V}[\ddot{s} + \boldsymbol{\pi}(\dot{s} - \ddot{s})] = (\dot{s} - \ddot{s})^2\sigma^2_\Pi$, and where $\frac{(n-(m+1))MSE_{\hat{\boldsymbol{q}}}}{(\dot{s}-\ddot{s})^2\sigma^2_\Pi}$ follows a χ^2 distribution with $n-(m+1)$ degrees of freedom.

In the case of single-source procurement, the resulting variance matrix for $\hat{\boldsymbol{\theta}}$ becomes:

$$\mathbb{V}[\hat{\boldsymbol{\theta}}] = \begin{bmatrix} \frac{1}{|I_1|^2}\sum\limits_{i \in I_1}\mathbb{V}[\hat{q}_i] & 0 & \dots & 0 \\ 0 & \frac{1}{|I_2|^2}\sum\limits_{i \in I_2}\mathbb{V}[\hat{q}_i] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{|I_m|^2}\sum\limits_{i \in I_m}\mathbb{V}[\hat{q}_i] \end{bmatrix}$$

Simulation Model

[See Word document and Powerpoint graphic for a more succinct description of the simulation model]

Testing Policies

The following is a list of testing policies that can be used within the simulation model. Each test policy uses a variety of different outputs, and outputs a schedule of days and outlets for conducting sampling.

- **Deterministic**: Rotates through each outlet in numerical/indexed order until the sampling budget is exhausted, such that Day 1 features Outlet 1, Day 2 features Outlet 2, etc., cycling back to Outlet 1 once Outlet $|J|$ is reached
- **Randomized**: Upon determining a number of samples to collect each day such that no day features more than a difference of 1 extra sample from any other day, chooses random outlets on each day until the sampling budget is exhausted
- **ϵ-greedy**: ϵ is the desired exploration ratio, such that the outlet j with the highest positive proportion $\hat{q}_j$ is exploited with probability $(1-\epsilon)$
- **Exponential-decay** ϵ: Similar to ϵ-greedy, except that ϵ decays exponentially over time
- **ϵ-first**: ϵ now signifies the proportion of our budget devoted to exploration before moving towards exploitation
- **Thompson-sampling**: Uses random samples from β distributions formed from outlet positives and negatives thus far to determine the next outlet to sample
- **Every-other**: Exploit on even days, and explore on odd days
- $\sin \epsilon$: Exploration follows a sine function of the number of days that have elapsed
- **Thompson-sampling with NUTS**: Thompson-sampling is conducted by projecting a NUTS posterior sample of importer SFP rates onto the outlets using $\boldsymbol{A}$, from which the outlet with the highest rate is selected
- **Exploration with NUTS**: The aim is to reduce importer posterior distribution variances. First, generate importer and outlet SFP distributions via NUTS. Identify importer node distribution variances. Multiply by $\boldsymbol{A}$ to get a weighted vector corresponding to the outlets. Pick an outlet (or set of outlets) to sample according to this vector.
- **Threshold exploration with NUTS**: The aim is to classify importers' SFP rates as above or below a particular threshold. First, generate importer and outlet SFP distributions via NUTS. Identify importer distribution spread over a designated threshold value, t, by calculating $1/|F(t)-0.5|$, where $F()$ is a beta distribution fit to the posterior distribution estimates. Weight the distribution spreads over t, then multiply by $\boldsymbol{A}$ to get a weighted vector corresponding to outlets. Use this vector to choose an outlet (or set of outlets) to sample.

Questions & Thoughts

- Add weights to observations by number of data points collected (Q matrix we discussed previously)?
- How closely does the simulation model verify align with these calculations?
- Is there an impact of variable market shares for different importers? $|I_1| \geq |I_2| \geq \cdots \geq |I_m|$
- Correct that the process of forming $\boldsymbol{A}^T\boldsymbol{A}$ for adjusted single-source procurement above can be generalized for general distribution vector u?

Future Research Ideas

- Expanding to three-tier supply chains, so large countries and global suppliers might be included? Might be implications for value in cross-country coordination and data-sharing
- How might these models affect the choice of detection device investment?
- What are the implications for decisions of which region and/or specific outlets to sample from? (Potentially the most direct implications from this analysis)

- What types of statistics + hypothesis testing can be conducted under different supply chains and test results?
- What if we have continuous data on pharmaceutical quality (i.e., API levels), instead of go/no-go information?

References

[1] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.