1 Solving parameters using Gauss-Jordan elimination

Standard approach to solving determined sets of equations requires just enough constrains to ensure uniqueness of the solution, without it being overdetermined (which can lead to contradictions). In our case, the set of available information is greater than required number of equations that satisfy above conditions. Out of 2^n possible binary vectors representing joint probability of events, we are forced to pick n that introduce the fewest error in further calculations. This approach can yield good results when done properly, but the question of which equations to choose remains unanswered. Preserving linear independence of vectors invoke additional complexity to the rules by which we decide the final set of equations.

Using Gauss-Jordan elimination, we can avoid this problem entirely, since it allows us to work with both overdetermined and underdetermined systems of equations. The order of equations is also taken into account, so in cases of contradictions, certain combinations are preferred to others. General rules of Gauss elimination apply, but are slightly modified to fit our product-equations (rather than linear equations). The outcome of our procedure for Gauss-Jordan elimination is a reduced row echelon form.

1.1 Example of Gauss-Jordan elimination

Let us work with this simple example written in a standard matrix form $A \cdot X = b$.

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \tag{1}$$

We will use symbols for vector b, so we can keep track of operations with constants. Turning that into augmented matrix $[A|b]$ yields

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array}\right] \tag{2}$$

Notice that this equation set is both contradictory (x_3 can be calculated in two ways), and underdetermined (There is no way to distinguish between x_1 and x_2 parameters). We will now perform Gauss-Jordan elimination steps in order to show that certain properties we care about (such as preserving preference of equations determined by their order) apply.

Since we do not focus on any particular column, so we will try to solve the whole equation. We will distinguish pivot elements with colors red (currently selected pivot element) and blue (previous pivot elements).

1. We choose the our first pivot element (in red), and use it to zero-out remaining coefficients in first column.

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array}\right] \quad r_2 = r_2 - r_1 \quad \sim \quad \left[\begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 0 & 0 & -1 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array}\right] \tag{3}$$

2. We select the second pivot element - note that no two pivot elements can share the same row. The first non-zero element that satisfies this condition is coefficient of x_3 in second row. Choosing the first element from the top guarantees that the order of preference of equations in taken into account.

$$
\left[\begin{array}{cccc|c}
1 & 1 & 1 & 0 & b_1 \\
0 & 0 & -1 & 0 & b_2 - b_1 \\
0 & 0 & 0 & 1 & b_3 \\
0 & 0 & 1 & 0 & b_4
\end{array}\right]
\begin{array}{l} \\ r_2 = r_2 \cdot (-1) \\ \\ \end{array}
\sim
\left[\begin{array}{cccc|c}
1 & 1 & 1 & 0 & b_1 \\
0 & 0 & 1 & 0 & b_1 - b_2 \\
0 & 0 & 0 & 1 & b_3 \\
0 & 0 & 1 & 0 & b_4
\end{array}\right]
\tag{4}
$$

$$
\left[\begin{array}{cccc|c}
1 & 1 & 1 & 0 & b_1 \\
0 & 0 & 1 & 0 & b_1 - b_2 \\
0 & 0 & 0 & 1 & b_3 \\
0 & 0 & 1 & 0 & b_4
\end{array}\right]
\begin{array}{l} r_1 = r_1 - r_2 \\ \\ \\ r_4 = r_4 - r_2 \end{array}
\sim
\left[\begin{array}{cccc|c}
1 & 1 & 0 & 0 & b_1 - (b_1 - b_2) \\
0 & 0 & 1 & 0 & b_1 - b_2 \\
0 & 0 & 0 & 1 & b_3 \\
0 & 0 & 0 & 0 & b_4 - (b_1 - b_2)
\end{array}\right]
\tag{5}
$$

3. The last pivot element is going to be coefficient at x_4 in fourth row. Since the remaining coefficients are all zeros in fourth column, no changes are made. We can simplify the values in new vector b. Notice that fourth row is a zero-vector - we can eliminate that from the equation set.

$$
\left[\begin{array}{cccc|c}
1 & 1 & 0 & 0 & b_1 - (b_1 - b_2) \\
0 & 0 & 1 & 0 & b_1 - b_2 \\
0 & 0 & 0 & 1 & b_3 \\
0 & 0 & 0 & 0 & b_4 - (b_1 - b_2)
\end{array}\right]
\sim
\left[\begin{array}{cccc|c}
1 & 1 & 0 & 0 & b_2 \\
0 & 0 & 1 & 0 & b_1 - b_2 \\
0 & 0 & 0 & 1 & b_3
\end{array}\right]
\tag{6}
$$

Let's compare our end-result with the initial matrix:

$$
\left[\begin{array}{cccc|c}
1 & 1 & 1 & 0 & b_1 \\
1 & 1 & 0 & 0 & b_2 \\
0 & 0 & 0 & 1 & b_3 \\
0 & 0 & 1 & 0 & b_4
\end{array}\right]
\tag{7}
$$

As we can see, x_3 was calculated using first and second row $(b_1 - b_2)$. Let's see what happens after we move third row on the top position, indicating that our preferred ordering of equation changed. (We expect now to calculate x_3 solely by first row).

1. We choose the our first pivot element, and use it to zero-out remaining coefficients in first column.

$$
\left[\begin{array}{cccc|c}
0 & 0 & 1 & 0 & b_1 \\
1 & 1 & 1 & 0 & b_2 \\
1 & 1 & 0 & 0 & b_3 \\
0 & 0 & 0 & 1 & b_4
\end{array}\right]
\begin{array}{l} \\ \\ r_3 = r_3 - r_2 \\ \end{array}
\sim
\left[\begin{array}{cccc|c}
0 & 0 & 1 & 0 & b_1 \\
1 & 1 & 1 & 0 & b_2 \\
0 & 0 & -1 & 0 & b_3 - b_2 \\
0 & 0 & 0 & 1 & b_1
\end{array}\right]
\tag{8}
$$

2. Again, no candidate for pivot element in second column, coefficient at x_3 in first row is the next pivot element

$$
\left[\begin{array}{cccc|c}
0 & 0 & 1 & 0 & b_1 \\
1 & 1 & 1 & 0 & b_2 \\
0 & 0 & -1 & 0 & b_3 - b_2 \\
0 & 0 & 0 & 1 & b_1
\end{array}\right]
\begin{array}{l} \\ r_2 = r_2 - r_1 \\ r_3 = r_3 + r_1 \\ \end{array}
\sim
\left[\begin{array}{cccc|c}
0 & 0 & 1 & 0 & b_1 \\
1 & 1 & 0 & 0 & b_2 - b_1 \\
0 & 0 & 0 & 0 & b_3 - b_2 + b_1 \\
0 & 0 & 0 & 1 & b_4
\end{array}\right]
\tag{9}
$$

.

3. Last item fit for a pivot element is a coefficient at x_4 in fourth row. After getting rid of zero vectors, we achieve the following reduced row echelon form matrix

$$\left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 0 & b_3 - b_2 + b_1 \\ 0 & 0 & 0 & 1 & b_4 \end{array}\right] \sim \left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 1 & b_4 \end{array}\right] \tag{10}$$

.

As expected, x_3 was calculated using the most preferred set of equations, as dictated by their order. What this method does not take into account is the relative weight (or preference) of each row. As of yet, all we could rely on was simple ordering of equations, without using the information about quantity or frequency of each type of equation in our learning set. If our method was to provide that feature to us as well, we could talk about very complete and solid solution that can be expected to perform optimally.

1.2   Properties of zero vectors

In this section we will describe how solving equation set without any particular ordering does not DEPRIVE[FIXME] us from reproducing other ways to calculate given parameter. As we saw previously, different order of equation leads to different outcomes for certain parameters. This variety came from contradictions in equation set. Zero vectors that emerge during Gauss-Jordan elimination contain information about other ways to calculate given value. Instead of removing them in the process, we can store them, and utilize them later. Let's see how previous example holds to that theory.

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 1 & 0 & 0 & b_1 - (b_1 - b_2) \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 0 & 0 & b_4 - (b_1 - b_2) \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array}\right] \tag{11}$$

This particular order of equation lead to x_3 being calculated from two top-most equations in a set. If we add our final solution for x_3 (vector in red), to the zero vector we ought to remove in a penultimate step of our algorithm (vector in blue), we obtain previously abandoned solution:

$$\left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 - b_2 \end{array}\right] + \left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & b_4 - (b_1 - b_2) \end{array}\right] = \left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_4 \end{array}\right] \tag{12}$$

Using the same method we can start from the solution obtained after rearranging the order of equations in initial matrix.

$$\left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 1 & 0 & b_2 \\ 1 & 1 & 0 & 0 & b_3 \\ 0 & 0 & 0 & 1 & b_4 \end{array}\right] \sim \left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 0 & b_3 - b_2 + b_1 \\ 0 & 0 & 0 & 1 & b_4 \end{array}\right] \sim \left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 1 & b_4 \end{array}\right] \tag{13}$$

Linear combination of two vectors again yields a different solution

$$\left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \end{array}\right] + (-1) \cdot \left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & b_3 - b_2 + b_1 \end{array}\right] = \left[\begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_2 - b_3 \end{array}\right] \tag{14}$$

Above property is crucial in showing that although the method itself does not take relative frequency, or weight between the equations in a set, it does provide us with a collection of solutions for given parameter. This way we can iterate over the set of possible solutions and choose the one that minimizes the relative error, or take a weighted average as our solution.

Let's treat fourth column that responds to quantity of each equation in a dataset. Whole equation set adds up to 2500 records. We can calculate a frequency for each of the probabilities, and treat them as weights.

$$
\begin{array}{cccc|c|c|c}
 & & & & \text{probability} & \text{quantity} & \text{frequency } Fq(b_i) \\
\hline
0 & 0 & 1 & 0 & b_1 & 980 & 0.392 \\
1 & 1 & 1 & 0 & b_2 & 760 & 0.304 \\
1 & 1 & 0 & 0 & b_3 & 440 & 0.176 \\
0 & 0 & 0 & 1 & b_4 & 320 & 0.128
\end{array}
\tag{15}
$$

Using this data we can come up with few heuristics for calculating final value of x_3 or compare different solutions. Let's propose a fitness function for a solution:

$$
F(s) = \frac{\displaystyle\prod_{b_i \in s} Fq(b_i)}{\displaystyle\sum_{b_i \in s} 1},
\tag{16}
$$

, where $b_i \in s$ is true when b_i is taken into account (adding or subtracting) in given solution.

$$
\begin{array}{c|c|c}
 & \text{probability} & \text{fitness function F(s)} \\
\hline
[0 \quad 0 \quad 1 \quad 0]_1 & b_1 & 0.392 \\
[0 \quad 0 \quad 1 \quad 0]_2 & b_2 - b_3 & \frac{0.304 \cdot 0.176}{2} = 0.24
\end{array}
\tag{17}
$$

At this point we can pick a solution with a higher fitness value, or take weighted average of each solution as our final answer:

$$
\begin{array}{c|c}
 & \text{probability} \\
\hline
[0 \quad 0 \quad 1 \quad 0] & \frac{0.392 \cdot b_1 + 0.24 \cdot (b_2 - b_3)}{0.392 + 0.24}
\end{array}
\tag{18}
$$