

The Model Space of Equality Constraints

Note, this is written chronologically from how I went from A to B. This is probably not a good way to present this to a journal. I'm not entirely sure whether we should frame the problem as a graph or as the counting of partitions. Viewing it as counting of partitions opens up much more relations to combinatorics, but I understand that perspective less well than that of a graphs.

Problem Description

Given K parameters, $\theta_1, \dots, \theta_K$, we are interested in exploring all possible equality constraints. This problem can be viewed as an undirected graph where the parameters are vertices and two parameters are equal if and only if there is an edge between them. For example for $K = 3$ we have:

	θ_1	θ_2	θ_3
θ_1			
θ_2	γ_1		
θ_3	γ_2	γ_3	

where $\gamma_i \in \{0, 1\}$ for $i = 1, 2, 3$. However, if we list all possible models, it becomes clear that this parametrization gives rise to duplicate models. For example, $\gamma = (1, 1, 1)$, $\gamma = (1, 1, 0)$, $\gamma = (1, 0, 1)$, and $\gamma = (0, 1, 1)$ all represent the same model where all parameters are equal. At the same time, no single γ_i is redundant but only certain combinations of the edges are. There are multiple ways to constrain the adjacency matrix such that duplicate models are impossible and we discuss one of these here. The first constraint that we impose is that the rows of the adjacency matrix sum to one. This makes the combination $\gamma = (0, 1, 1)$ impossible. The second constraint asserts that there can only be one edge to a connected set of vertices, and that a connection can only be made with the 'first' vertex of a connected set. For example, given $\gamma_1 = 1$ it follows that $\gamma_3 = 0$.

To satisfy the first constraint, we reparametrize the binary indicator variables to categorical ones, that is, $\kappa_1 = 1, \kappa_2 \in \{1, 2\}, \kappa_3 \in \{1, 2, 3\}$.¹ Two parameters are equal whenever their indicator values are equal, for example, $\kappa = (1, 2, 2)$ implies $\theta_2 = \theta_3$. The second constraint can be formulated as $\kappa_2 \neq 2 \implies \kappa_3 \neq 2$. More generally, this means that $\kappa_i \neq i \implies \kappa_j \neq i$ for $j = i + 1, \dots, K$.

¹Note that although κ_1 is redundant we use it to ease the notation.

For illustrative purposes we list all possible model for $K = 3$ under both parametrizations and whether they violate any restrictions.

Table 1: All possible configurations for both sets of indicator variables and the whether the corresponding models violate any constraints.

Binary indicator		Categorical indicator	
γ	valid	κ	valid
0, 0, 0	✓	1, 1, 1	✓
0, 0, 1	✓	1, 1, 2	violates constraint 2
0, 1, 0	✓	1, 1, 1	✓
1, 0, 0	✓	1, 2, 1	✓
0, 1, 1	violates constraint 1	1, 2, 2	✓
1, 0, 1	violates constraint 2	1, 2, 3	✓
1, 1, 0	✓		
1, 1, 1	violates constraint 1		

From the table above we see there are in total 5 valid models. Counting the number of possible models is equivalent to counting the number of partitions into non-empty subsets of a set of size K . Therefore, given K parameters the number of possible models is given by the K^{th} Bell number, B_K . The model space grows exponentially in K , for example, $B_3 = 5$, $B_5 = 52$, $B_{10} = 115\,975$.

Sampling Models Uniformly

Here we construct a Gibbs sampler to sample uniformly from the model space. A naive approach is to sample from a uniform categorical distribution while satisfying constraint 2. However, this distribution is not uniform over the models space. From Table 1 we can see that there are 5 valid models, and that there are 2 models where $\kappa_2 = 1$ and 3 models where $\kappa_2 = 2$. Adjusting the probabilities boils down to counting how often a particular outcome occurs conditional on the previous indicator variables. These counts can be obtained using the r -Bell numbers, which are defined as (Mezo, 2011):

$$B(n, r) = \sum_{k=0}^n \left\{ \begin{matrix} n+r \\ k+r \end{matrix} \right\}_r = \sum_{k=0}^n \left(\sum_{i=0}^n \binom{n}{i} \left\{ \begin{matrix} i \\ k \end{matrix} \right\} r^{n-i} \right)$$

where $\left\{ \begin{matrix} n+r \\ k+r \end{matrix} \right\}_r$ denotes the r -Stirling number and $\left\{ \begin{matrix} i \\ k \end{matrix} \right\}$ denotes the Stirling partition number. The left definition can be interpreted as the number of the partitions of a set with $n + r$ element such that the first r elements are in distinct subsets in each partition. Let s_i denote the number of active equality constraints, e.g., $s_i = \sum_{j=1}^{i-1} I(\kappa_j = j)$, where $I(\cdot)$ is an indicator function that returns 1 if its argument is true and 0 otherwise. Then, the frequency of each

possible outcome for κ_i is given by

$$\text{Freq}(\kappa_i = j) = \begin{cases} B(K - i - 1, s_i + 1) & \text{if } i = j \\ B(K - i - 1, s_i) & \text{if } \kappa_j = j \\ 0 & \text{if } \kappa_j \neq j \end{cases}$$

For example, for $K = 3$ we have $\text{Freq}(\kappa_2 = 1) = B(1, 1) = 2$ and $\text{Freq}(\kappa_2 = 2) = B(1, 2) = 3$, which matches Table 1.

Using the frequencies we obtain the following scheme to sample uniformly from the model space:

$$\begin{aligned} \pi(\kappa_1) &= \text{Categorical}(1) \\ \pi(\kappa_2 \mid \kappa_1) &= \text{Categorical}(B(K - 2, 1), B(K - 2, 2)) \\ \pi(\kappa_3 \mid \kappa_1, \kappa_2) &= \text{Categorical}(\text{Freq}(\kappa_3 = 1), \text{Freq}(\kappa_3 = 2), \text{Freq}(\kappa_3 = 3)) \\ &\vdots \\ \pi(\kappa_K \mid \kappa_1, \dots, \kappa_{K-1}) &= \text{Categorical}(\text{Freq}(\kappa_K = 1), \dots, \text{Freq}(\kappa_K = K)) \end{aligned}$$

Here it is assumed that the probability vectors for the categorical distributions are normalized so they sum to 1.

Don: A possible problem here is that the sampler may have difficulties to transition between models that are intuitively close together. For example, given $K = 3$ and $\kappa = (1, 2, 2)$ (one equality constraint) it is not allowed to go to $(2, 2, 2)$ (two equality constraints). Instead we'd need to do $(1, 2, 3) \rightarrow (1, 1, 3) \rightarrow (1, 1, 1)$. Note that $(1, 1, 2)$ is also an impossible state. A solution for this is to let $\kappa_i \in \{1, \dots, K\} \forall i$. However, that introduces a large number of duplicate models we'd need to adjust for. So far I've not encountered any issues with this yet, but maybe we will when K grows larger.

A Beta-Binomial Penalty

As in Linear regression, a uniform distribution over models is not uniform over the number of equality constraints. To achieve this, we implement a beta-binomial penalty. First, we need to count the number of models with k number of equality constraints for $k = 1, \dots, K$.

References

Mezo, I. (2011). The r -bell numbers. *Journal of Integer Sequences*, 14(1), 1–14.