3-category pairwise joints parameterization

Sophie Li

March 2025

1 Notation

We want to paramaterize the marginals and pairwise joints of a categorical distribution with c categories and n variables. For illustration purposes, we make $n = 2$, $c = 3$ (and can hopefully generalize the logic). For notation, let

$$P(X_1 = 1) = a_{11}, P(X_1 = 2) = a_{12}, P(X_2 = 1) = a_{21}, P(X_2 = 2) = a_{22}$$

Denote the 9 pairwise marginals as such

$$\beta_{00} = P(X_1 = 0, X_2 = 0)...\beta_{22} = P(X_1 = 2, X_2 = 2)$$

2 Consistency (Linear System of Eqns)

Assume we have set the marginals already. We want a general solution for the joints.

They must satisfy 6 equations (3 row equations, 3 column equations). The first "row" equation corresponds to $P(X_1 = 0)$:

$$(1) P(X_1 = 0) = P(X_1 = 0, X_2 = 0) + P(X_1 = 0, X_2 = 1) + P(X_1 = 0, X_2 = 2)$$

$$P(X_1 = 0, X_2 = 0) + \beta_{01} + \beta_{02} = 1 - a_{11} - a_{12}$$

The third row eqn is

$$P(X_1 = 2) = P(X_1 = 2, X_2 = 0) + P(X_1 = 2, X_2 = 1) + P(X_1 = 2, X_2 = 2)$$

The column equations are analogous with X_2.

For consistency, we set the 6 off-diagonal β's, and isolate the diagonal joints.

Row eqn 1 implies

$$P(X_1 = 0, X_2 = 0) = 1 - a_{11} - a_{12} - (\beta_{01} + \beta_{02})$$

Col eqn 1 implies

$$P(X_1 = 0, X_2 = 0) = 1 - a_{21} - a_{22} - (\beta_{10} + \beta_{20})$$

These expressions must be equal. Similarly, there are 2 corresponding eqns for $P(X_1 = 1, X_2 = 1)$ and $P(X_1 = 2, X_2 = 2)$. (I'll skip the details and write all 3 here):

$$1 - a_{11} - a_{12} - (\beta_{01} + \beta_{02}) - 1 - a_{21} - a_{22} - (\beta_{10} + \beta_{20})$$

$$\beta_{01} + \beta_{21} - (\beta_{10} + \beta_{12}) = a_{21} - a_{11}$$

$$\beta_{02} + \beta_{12} - (\beta_{20} + \beta_{21}) = a_{22} - a_{12}$$

It turns out if we add equations 2 and 3 and simplify the expression, we get equation 1. Since eqn 1 is redundant, this is a linear system of 2 equations in 6 variables.

Skipping the matrix algebra, we get 4 free parameters $\beta_{10}, \beta_{12}, \beta_{20}, \beta_{21}$ (notice, these are from the last 2 rows). The remaining two parameters are

$$\beta_{01} = a_{21} - a_{11} + \beta_{10} + \beta_{12} - \beta_{21}$$

$$\beta_{02} = a_{22} - a_{12} - \beta_{12} + \beta_{20} + \beta_{21}$$

Thus, we can first set the marginals, then choose 4 free betas, then everything else is determined and will be consistent.

3 Inequalities

However, using the solution above, it is possible to get negative probabilities. This section devises an exact procedure to choose the β's in an appropriate range, depending on the previous parameters, to ensure we get NON-negative values.

First, obviously the marginals have to follow

$$a_{11} + a_{12} \leq 1$$

$$a_{21} + a_{22} \leq 1$$

Now, assume we have fixed the marginals. We have 9 joints to consider. Four of them are the free β's. Evidently, they must be ≥ 0.

Now, we have 5 inequalities left, corresponding to

$$\beta_{00}, \beta_{01}, \beta_{02}, \beta_{11}, \beta_{22}$$

We write them in terms of the free params and set them to be ≥ 0, in the above order.

In the first row, we must have $\beta_{00}, \beta_{01}, \beta_{02} \geq 0$, which corresponds to

$$\beta_{10} + \beta 20 \leq 1 - (a_{21} + a_{22})$$

$$\beta_{21} - (\beta_{10} + \beta_{12}) \leq a_{21} - a_{11}$$

$$\beta_{12} - (\beta_{20} + \beta_{21}) \leq a_{22} - a_{12}$$

Then, β_{11}, β_{22} correspond to the following:

$$(\beta_{10} + \beta_{12}) \leq a_{11}$$

$$(\beta_{20} + \beta_{21}) \leq a_{12}$$

4 Exact Procedure

I now derive a procedure for a general parameterization, satisfying the 5 inequalities. First, we set β_{10} freely:

$$0 \leq \beta_{10} \leq min(a_{11}, 1 - a_{21} - a_{22})$$

Note that $\beta_{10} = P(1, 0)$ must be smaller than both $P(X_1 = 1)$ and $P(X_2 = 0)$ hence why we take the min.

Similarly, set β_{20}:

$$0 \leq \beta_{20} \leq min(a_{12}, 1 - a_{21} - a_{22} - \beta_{10})$$

Then, noting the last 2 inequalities in section 3, β_{12} and β_{21} must satisfy:

$$\beta_{12} \leq a_{11} - \beta_{10}$$

$$\beta_{21} \leq a_{12} - \beta_{20}$$

Note that by now, we have satisfied 3 of the 5 inequalities. We turn our attention to the remaining two, corresponding to β_{01}, β_{02}. Isolating our unknowns, we obtain

$$\beta_{21} - \beta_{12} \leq \beta_{10} + a_{21} - a_{11}$$
$$\beta_{12} - \beta_{21} \leq \beta_{20} + a_{22} - a_{12}$$

Flipping the last inequality and combining, this means

$$-(\beta_{20} + a_{22} - a_{12}) \leq \beta_{21} - \beta_{12} \leq \beta_{10} + a_{21} - a_{11}$$

For this to hold, we must have LHS $\leq$ RHS:

$$-(\beta_{20} + a_{22} - a_{12}) \leq \beta_{10} + a_{21} - a_{11}$$

This imposes an additional constraint when we pick β_{20} after fixing β_{10}:

$$max(0, a_{11} - a_{21} - a_{22} + a_{12} - \beta_{10}) \leq \beta_{20}$$

Now, turning back to the remaining 3 inequalities for β_{12}, β_{21}, after β_{10}, β_{20} have been fixed:

$$\beta_{12} \leq a_{11} - \beta_{10}$$

$$\beta_{21} \leq a_{12} - \beta_{20}$$

$$-(\beta_{20} + a_{22} - a_{12}) \leq \beta_{21} - \beta_{12} \leq \beta_{10} + a_{21} - a_{11}$$

Case 1: $a_{12} - \beta_{20} \geq a_{11} - \beta_{10}$ (i.e. the upper bound for β_{21} exceeds that of β_{12}.

First, we choose β_{12} freely in its range. Then, we choose β_{21} s.t.

$$0 \leq \beta_{21} \leq a_{12} - \beta_{20}$$

$$\beta_{12} - (\beta_{20} + a_{22} - a_{12}) \leq \beta_{21} \leq \beta_{12} + \beta_{10} + a_{21} - a_{11}$$

Note: it is not hard to justify that these two ranges always overlap (i.e. there will always be valid solutions for β_{21}).

The intuition for why we split it into cases is that since β_{12} has a lesser upper bound, it has less "flexibility" so we must fix it first. There may not always be valid solutions if we choose β_{21} first in this case (if we choose it too high).

Case 2: $a_{12} - \beta_{20} \leq a_{11} - \beta_{10}$ (i.e. the upper bound for β_{12} exceeds that of β_{21}.

This case is analogous. First, we choose β_{21} freely in its range. Then, we choose β_{12} s.t.

$$0 \leq \beta_{12} \leq a_{11} - \beta_{10}$$

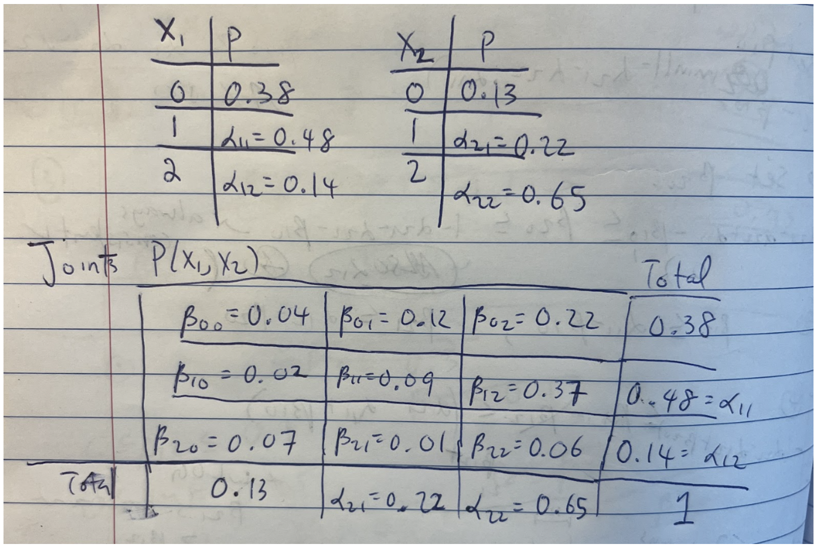
$$-(\beta_{20} + a_{22} - a_{12}) \leq \beta_{21} - \beta_{12} \leq \beta_{10} + a_{21} - a_{11}$$

$$\beta_{21} - (\beta_{10} + a_{21} - a_{11}) \leq \beta_{12} \leq \beta_{21} + \beta_{20} + a_{22} - a_{12}$$

With this, we have gotten four free parameters to ensure the joints are not only consistent w/the marginals, but also non-negative!

5 Example of Procedure Execution

As an example, let us recover the following joint distribution, applying this algorithm. Imagine we have fixed the marginals (alphas), and want to get a fully general paramaterization for the possible joints (betas).



X_1	P
0	0.38
1	$\alpha_{11}=0.48$
2	$\alpha_{12}=0.14$

X_2	P
0	0.13
1	$\alpha_{21}=0.22$
2	$\alpha_{22}=0.65$

Joints $P(X_1, X_2)$

				Total
	$\beta_{00}=0.04$	$\beta_{01}=0.12$	$\beta_{02}=0.22$	0.38
	$\beta_{10}=0.02$	$\beta_{11}=0.09$	$\beta_{12}=0.37$	$0.48=\alpha_{11}$
	$\beta_{20}=0.07$	$\beta_{21}=0.01$	$\beta_{22}=0.06$	$0.14=\alpha_{12}$
Total	0.13	$\alpha_{21}=0.22$	$\alpha_{22}=0.65$	1

Step 1. Set β_{10} s.t. $0 \leq \beta_{10} \leq min(a_{11}, 1 - a_{21} - a_{22}) = min(0.13, 0.48)$, so let $\beta_{10} = 0.02$

Step 2. Set β_{20} s.t. $max(0, a_{11} - a_{21} - a_{22} + a_{12} - \beta_{10}) \leq \beta_{20} \leq min(a_{12}, 1 - a_{21} - a_{22} - \beta_{10})$

Here, we have $max(0, -0.27) \leq \beta_{20} \leq min(0.14, 0.11)$, so let $\beta_{20} = 0.07$.

Step 3. Determine the upper bounds of β_{12}, β_{21}:

$$0 \leq \beta_{12} \leq a_{11} - \beta_{10} = 0.46$$

$$0 \leq \beta_{21} \leq a_{12} - \beta_{20} = 0.07$$

Step 4. Execute the casework described above. Since β_{12}'s bound is bigger than β_{21}'s, this is case 2. Set $\beta_{21} = 0.01$ first. Then, set β_{12} s.t.

$$0 \leq \beta_{12} \leq a_{11} - \beta_{10} = 0.46$$

$$\beta_{21} - (\beta_{10} + a_{21} - a_{11}) = 0.25 \leq \beta_{12} \leq \beta_{21} + \beta_{20} + a_{22} - a_{12} = 0.59$$

Thus, we choose $\beta_{12} = 0.37$, our last free parameter.

Step 5. Calculate the remaining joints based on the free variables. Recall in our system of equations,

$$\beta_{01} = a_{21} - a_{11} + \beta_{10} + \beta_{12} - \beta_{21} = 0.12$$

$$\beta_{02} = a_{22} - a_{12} - \beta_{12} + \beta_{20} + \beta_{21} = 0.22$$

which correctly matches our original distribution. Lastly, we can recover the diagonals as $\beta_{00} = 0.04, \beta_{11} = 0.09, \beta_{22} = 0.06$, which completes the procedure.

6 Generalizing

This is the simple case of 2 variables, 3 categories.

If we increase the no. of **variables**, I believe the problem won't get much harder. For instance, if we had 3 variables, we would have 3 different joint distributions. I believe one can apply this procedure to each, with a total of $3 * 4 = 12$ free parameters.

If we increase the no. of **categories**, we basically get an $(n + 1)$ by $(n + 1)$ grid. My intuition is that the no. of parameters (betas) increases quadratically, whereas the no. of constraints only increases linearly, thus the system becomes "less" constrained, which might make the general case not so much worse. However, it is difficult by hand so I have not tried it. But the general intuition should be the same.