Maximum Likelihood Estimation for Independence Models

2026-02-12

Contents

Contents

1 Proposition 5.3.8

Let $M_{1\perp 2}$ be the model of independence of two discrete random variables, with r_1 and r_2 states respectively. Let $u \in \mathbb{N}^{r_1 \times r_2}$ be the table of counts for this model obtained from i.i.d. samples from the model. Let $u_{i_1+} = \sum_{i_2} u_{i_1 i_2}$ and $u_{+i_2} = \sum_{i_1} u_{i_1 i_2}$ be the table marginals, and $n = \sum_{i_1, i_2} u_{i_1 i_2}$ the sample size. Then the maximum likelihood estimate for a distribution $p \in M_{1\perp 2}$ given the data u is

$$\hat{p}_{i_1 i_2} = \frac{u_{i_1+} \cdot u_{+i_2}}{n^2}.$$

2 Proof

2.1 The Contingency Table Setup

Observed data

$X_1 \backslash X_2$	$X_2 = 1$	$X_2 = 2$	$\cdots$	$X_2 = r_2$	Row Totals
$X_1 = 1$	u_{11}	u_{12}	$\cdots$	u_{1r_2}	$u_{1+} = \sum_{j=1}^{r_2} u_{1j}$
$X_1 = 2$	u_{21}	u_{22}	$\cdots$	u_{2r_2}	$u_{2+} = \sum_{j=1}^{r_2} u_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$X_1 = r_1$	$u_{r_1 1}$	$u_{r_1 2}$	$\cdots$	$u_{r_1 r_2}$	$u_{r_1+} = \sum_{j=1}^{r_2} u_{r_1 j}$
Column Totals	$u_{+1} = \sum_{i=1}^{r_1} u_{i1}$	$u_{+2} = \sum_{i=1}^{r_1} u_{i2}$	$\cdots$	$u_{+r_2} = \sum_{i=1}^{r_1} u_{ir_2}$	$n = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} u_{ij}$

2.2 Notation:

- u_{ij}: observed count for $(X_1 = i, X_2 = j)$
- u_{i+}: row total for $X_1 = i$
- u_{+j}: column total for $X_2 = j$
- n: total sample size

2.3 Model Setup

Distribution $p \in \Delta_{\mathcal{R}}$ belongs to the independence model $M_{1 \perp\!\!\!\perp 2}$ if and only if:

$$P(X_1 = i, X_2 = j) = P(X_1 = i) \cdot P(X_2 = j) \quad \text{for all } i, j$$

$$p_{i_1 i_2} = \alpha_{i_1} \beta_{i_2} \quad \text{for all } i, j$$

Let:

- $\alpha_{i_1} = P(X_1 = i_1)$ for $i_1 = 1, \ldots, r_1$
- $\beta_{i_2} = P(X_2 = i_2)$ for $i_2 = 1, \ldots, r_2$
- $\Delta_{r_1 r_2 - 1}$ denotes the probability simplex:

$$\Delta_{r_1 r_2 - 1} = \left\{ P \in \mathbb{R}^{r_1 \times r_2} : \ P_{i_1 i_2} \geq 0, \ \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} P_{i_1 i_2} = 1 \right\}.$$

for some $\alpha \in \Delta_{r_1 - 1}$ and $\beta \in \Delta_{r_2 - 1}$

The constraints are:

$$\alpha_{i_1} \geq 0, \quad \beta_{i_2} \geq 0, \quad \sum_{i_1=1}^{r_1} \alpha_{i_1} = 1, \quad \sum_{i_2=1}^{r_2} \beta_{i_2} = 1.$$

2.4 Likelihood of Independence Model

$$L(\alpha, \beta \mid u) = \prod_{i_1=1}^{r_1} \prod_{i_2=1}^{r_2} (\alpha_{i_1} \beta_{i_2})^{u_{i_1 i_2}}.$$

2.4.1 Log-Likelihood Function

Taking the natural logarithm (which preserves maxima since log is monotonic increasing):

$$\ell(\alpha, \beta \mid u) = \log L(\alpha, \beta \mid u) = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} u_{i_1 i_2} \log(\alpha_{i_1} \beta_{i_2}).$$

$$\ell(\alpha, \beta \mid u) = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} u_{i_1 i_2} [\log \alpha_{i_1} + \log \beta_{i_2}].$$

Separate the double sum:

$$\ell(\alpha, \beta \mid u) = \sum_{i_1=1}^{r_1} \left(\sum_{i_2=1}^{r_2} u_{i_1 i_2} \right) \log \alpha_{i_1} + \sum_{i_2=1}^{r_2} \left(\sum_{i_1=1}^{r_1} u_{i_1 i_2} \right) \log \beta_{i_2}.$$

Recognize the marginal totals:

$$u_{i_1+} = \sum_{i_2=1}^{r_2} u_{i_1 i_2}, \quad u_{+i_2} = \sum_{i_1=1}^{r_1} u_{i_1 i_2}.$$

Thus:

$$\ell(\alpha, \beta \mid u) = \sum_{i_1=1}^{r_1} u_{i_1+} \log \alpha_{i_1} + \sum_{i_2=1}^{r_2} u_{+i_2} \log \beta_{i_2}.$$

Note: The log-likelihood separates into two independent parts: one involving only α and the other involving only β.

Maximization w.r.t α and β, respectively subject to the constraints define earlier.

2.4.2 Maximizing with Respect to α

To maximize:

$$\ell_\alpha = \sum_{i_1=1}^{r_1} u_{i_1+} \log \alpha_{i_1}$$

subject to the constraints $\alpha_{i_1} \geq 0$ and $\sum_{i_1=1}^{r_1} \alpha_{i_1} = 1$.

We use the method of Lagrange multipliers. Define the Lagrangian:

$$\mathcal{L}_\alpha(\alpha, \lambda) = \sum_{i_1=1}^{r_1} u_{i_1+} \log \alpha_{i_1} + \lambda \left(1 - \sum_{i_1=1}^{r_1} \alpha_{i_1} \right),$$

where λ is the Lagrange multiplier.

Take partial derivatives with respect to α_{i_1}:

$$\frac{\partial \mathcal{L}_\alpha}{\partial \alpha_{i_1}} = \frac{u_{i_1+}}{\alpha_{i_1}} - \lambda = 0 \quad \text{for } i_1 = 1, \ldots, r_1.$$

Solving for α_{i_1}:

$$\frac{u_{i_1+}}{\alpha_{i_1}} = \lambda \quad \Rightarrow \quad \alpha_{i_1} = \frac{u_{i_1+}}{\lambda}.$$

From the constraint $\sum_{i_1=1}^{r_1} \alpha_{i_1} = 1$:

$$\sum_{i_1=1}^{r_1} \alpha_{i_1} = \sum_{i_1=1}^{r_1} \frac{u_{i_1+}}{\lambda} = \frac{1}{\lambda} \sum_{i_1=1}^{r_1} u_{i_1+} = 1.$$

But $\sum_{i_1=1}^{r_1} u_{i_1+} = n$. Hence,

$$\frac{n}{\lambda} = 1 \quad \Rightarrow \quad \lambda = n.$$

Therefore,

$$\hat{\alpha}_{i_1} = \frac{u_{i_1+}}{n} \quad \text{for } i_1 = 1, \ldots, r_1.$$

2.4.3 Maximizing with Respect to β

Similarly, we maximize:

$$\ell_\beta = \sum_{i_2=1}^{r_2} u_{+i_2} \log \beta_{i_2}$$

subject to $\beta_{i_2} \geq 0$ and $\sum_{i_2=1}^{r_2} \beta_{i_2} = 1$.

Define the Lagrangian:

$$\mathcal{L}_\beta(\beta, \mu) = \sum_{i_2=1}^{r_2} u_{+i_2} \log \beta_{i_2} + \mu \left(1 - \sum_{i_2=1}^{r_2} \beta_{i_2} \right),$$

where μ is the Lagrange multiplier.

Take partial derivatives:

$$\frac{\partial \mathcal{L}_\beta}{\partial \beta_{i_2}} = \frac{u_{+i_2}}{\beta_{i_2}} - \mu = 0 \quad \text{for } i_2 = 1, \ldots, r_2.$$

Solving:

$$\beta_{i_2} = \frac{u_{+i_2}}{\mu}.$$

Using the constraint $\sum_{i_2=1}^{r_2} \beta_{i_2} = 1$:

$$\sum_{i_2=1}^{r_2} \beta_{i_2} = \sum_{i_2=1}^{r_2} \frac{u_{+i_2}}{\mu} = \frac{1}{\mu} \sum_{i_2=1}^{r_2} u_{+i_2} = 1.$$

But $\sum_{i_2=1}^{r_2} u_{+i_2} = n$. Therefore:

$$\frac{n}{\mu} = 1 \quad \Rightarrow \quad \mu = n.$$

Sub. $\mu = n$:

$$\hat{\beta}_{i_2} = \frac{u_{+i_2}}{n} \quad \text{for } i_2 = 1, \ldots, r_2.$$

2.4.4 MLE for Joint Probabilities

Since $p_{i_1 i_2} = \alpha_{i_1}\beta_{i_2}$ under the independence model, the maximum likelihood estimate is:

$$\hat{p}_{i_1 i_2} = \hat{\alpha}_{i_1} \cdot \hat{\beta}_{i_2} = \frac{u_{i_1+}}{n} \cdot \frac{u_{+i_2}}{n} = \frac{u_{i_1+} \cdot u_{+i_2}}{n^2}.$$

NB: the MLEs for these marginals are their empirical frequencies.

2.4.5 Verification of Constraints

Check that $\hat{p}_{i_1 i_2}$ satisfies the probability constraints:

1. **Non-negativity**: $\hat{p}_{i_1 i_2} \geq 0$ since $u_{i_1+} \geq 0$, $u_{+i_2} \geq 0$, and $n > 0$.
2. **Sum to 1**:

$$\begin{aligned} \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \hat{p}_{i_1 i_2} &= \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \frac{u_{i_1+} \cdot u_{+i_2}}{n^2} \\ &= \frac{1}{n^2} \left(\sum_{i_1=1}^{r_1} u_{i_1+} \right) \left(\sum_{i_2=1}^{r_2} u_{+i_2} \right) \\ &= \frac{1}{n^2} \cdot n \cdot n = 1. \end{aligned}$$

Hence, $\hat{p}_{i_1 i_2}$ is a valid probability distribution.

3 Example

Consider a 2×2 contingency table:

	Success	Failure	Total
Treatment	45	15	60
Control	30	30	60
Total	75	45	120

Here:

- $u_{1+} = 60$, $u_{2+} = 60$ (row totals)
- $u_{+1} = 75$, $u_{+2} = 45$ (column totals)
- $n = 120$

Under the independence model:

$$\begin{aligned} \hat{p}_{11} &= \frac{60 \times 75}{120^2} = \frac{4500}{14400} = 0.3125 \\ \hat{p}_{12} &= \frac{60 \times 45}{120^2} = \frac{2700}{14400} = 0.1875 \\ \hat{p}_{21} &= \frac{60 \times 75}{120^2} = 0.3125 \\ \hat{p}_{22} &= \frac{60 \times 45}{120^2} = 0.1875 \end{aligned}$$

Check: $0.3125 + 0.1875 + 0.3125 + 0.1875 = 1.0000$.