Pair Comparison

Zhi Li, Netflix

April 25, 2022

1 Thurstone Model

Let ψ_1, ..., ψ_J be the quality values of J stimuli (e.g. PVS). Consider a generative model:

$$\begin{aligned} u_i &\sim N\left(\psi_i, \upsilon^2\right), \\ u_j &\sim N\left(\psi_j, \upsilon^2\right), \end{aligned}$$

where u_i is the "observed" quality of stimuli i, and ψ_i is the hidden "true" quality to be estimated. υ is the standard deviation associated with each stimuli, and in the Thurstone Case V model, υ is assumed constant for all stimuli, and without loss of generality, let us use $\upsilon = \frac{1}{\sqrt{2}}$. Assume that stimuli i is preferred over j if $u_i - u_j > 0$, or $\psi_i - \psi_j + N(0,1) > 0$. We can express the probability of i preferred over j as:

$$\Pr(i \text{ is preferred over } j) = p_{ij} = \Phi(\psi_i - \psi_j),$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ is the cumulative function of Gaussian distribution with $N(0,1)$. Let a number of subjects do the paired comparisons (PCs), and define

$$\begin{aligned} \alpha_{ij} &: \#\{i \text{ is preferred over } j\}, \\ n_{ij} &: \#\{\text{paired comparisons between } i \text{ and } j\}. \end{aligned}$$

It is easy to see that $n_{ij} = \alpha_{ij} + \alpha_{ji}$. Assume independence between the PCs, the probability that the counts of i is preferred over j is α_{ij} is

$$\Pr(A_{ij} = \alpha_{ij}) = \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} p_{ij}^{\alpha_{ij}} p_{ji}^{\alpha ji}.$$

The likelihood function of parameters $\{\psi_i\}$ given observations $\{\alpha_{ij}\}$ is

$$\begin{aligned} L(\{\alpha_{ij}\}|\{\psi_i\}) &= \prod_{ij} \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} p_{ij}^{\alpha_{ij}} p_{ji}^{\alpha ji} \\ &= \prod_{ij} \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} \Phi(\psi_i - \psi_j)^{\alpha_{ij}} \Phi(\psi_j - \psi_i)^{\alpha ji}. \end{aligned}$$

The log-likelihood function is then:

$$\ell(\{\alpha_{ij}\}|\{\psi_i\}) = \sum_{ij} \log \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} + \alpha_{ij} \log \Phi(\psi_i - \psi_j) + \alpha_{ji} \log \Phi(\psi_j - \psi_i).$$

The maximum likelihood estimation (MLE) of the parameters is

$$\{\hat{\psi}_i\} = \underset{\{\psi_i\}}{\arg\max}\ \ell(\{\alpha_{ij}\}|\{\psi_i\}),$$

which can be solved numerically.

To obtain the confidence interval of the MLE solution, we first calculate the first- and second-order partial derivatives. Let $f(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ be the probability density of Gaussian distribution with $N(0,1)$, and $d(x) = f'(x) = -xf(x)$ be the derivative of $f(x)$. The first-order derivative is

$$\frac{\partial \ell}{\partial \psi_i} = \sum_j \alpha_{ij} \frac{f(\psi_i - \psi_j)}{\Phi(\psi_i - \psi_j)} - \alpha_{ji} \frac{f(\psi_j - \psi_i)}{\Phi(\psi_j - \psi_i)}.$$

The second-order derivatives are

$$\begin{aligned}
\lambda_{ii} = \frac{\partial^2 \ell}{\partial \psi_i^2} &= \sum_j \alpha_{ij} \frac{\Phi(\psi_i - \psi_j) d(\psi_i - \psi_j) - f(\psi_i - \psi_j)^2}{\Phi(\psi_i - \psi_j)^2} \\
&\quad + \alpha_{ji} \frac{\Phi(\psi_j - \psi_i) d(\psi_j - \psi_i) - f(\psi_j - \psi_i)^2}{\Phi(\psi_j - \psi_i)^2} \\
&= \sum_j \frac{(\alpha_{ij} + \alpha_{ji}) \left(\Phi(\psi_i - \psi_j) d(\psi_i - \psi_j) - f(\psi_i - \psi_j)^2\right)}{\Phi(\psi_i - \psi_j)^2}, \\
\lambda_{ij} = \frac{\partial^2 \ell}{\partial \psi_i \partial \psi_j} &= -\left(\alpha_{ij} \frac{\Phi(\psi_i - \psi_j) d(\psi_i - \psi_j) - f(\psi_i - \psi_j)^2}{\Phi(\psi_i - \psi_j)^2}\right. \\
&\quad \left. + \alpha_{ji} \frac{\Phi(\psi_j - \psi_i) d(\psi_j - \psi_i) - f(\psi_j - \psi_i)^2}{\Phi(\psi_j - \psi_i)^2}\right) \\
&= -\left(\frac{(\alpha_{ij} + \alpha_{ji}) \left(\Phi(\psi_i - \psi_j) d(\psi_i - \psi_j) - f(\psi_i - \psi_j)^2\right)}{\Phi(\psi_i - \psi_j)^2}\right).
\end{aligned}$$

Construct the Hessian matrix $H = [\lambda_{ij}]$. Since their is one degree of freedom dependency in $\{\psi_i\}$ since it is scaling-invariant, we construct

$$C = \begin{bmatrix} -H & 1 \\ 1' & 0 \end{bmatrix}^{-1}$$

of dimension $(J+1) \times (J+1)$. The estimated variances of ψ_i is

$$Var(\psi_i) = diag(C)[i]$$

for $i = 1, ..., J$.

2 Bradley-Terry Model

Similar as the Thurstone model, let ψ_1, ..., ψ_J be the quality values of J stimuli (e.g. PVS). Assume that the probability of stimuli i is preferred over j follows the form:

$$\Pr(i \text{ is preferred over } j) = p_{ij} = H(\psi_i - \psi_j),$$

where

$$H(x) = \frac{1}{1 + e^{-x}}$$

is the sigmoid function. Similarly, let

$$\begin{aligned} \alpha_{ij} &: \#\{i \text{ is preferred over } j\}, \\ n_{ij} &: \#\{\text{paired comparisons between } i \text{ and } j\}. \end{aligned}$$

It is easy to see that $n_{ij} = \alpha_{ij} + \alpha_{ji}$. We can also write:

$$p_{ij} = H(\psi_i - \psi_j) = \frac{1}{1 + e^{-\psi_i + \psi_j}} = \frac{e^{\psi_i}}{e^{\psi_i} + e^{\psi_j}}.$$

Assuming independence between the PCs, the probability that the counts of i is preferred over j is α_{ij} is

$$\Pr(A_{ij} = \alpha_{ij}) = \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} p_{ij}^{\alpha_{ij}} p_{ji}^{\alpha_{ji}}.$$

The likelihood function of parameters $\{\psi_i\}$ given observations $\{\alpha_{ij}\}$ is

$$\begin{aligned} L(\{\alpha_{ij}\}|\{\psi_i\}) &= \prod_{ij} \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} p_{ij}^{\alpha_{ij}} p_{ji}^{\alpha_{ji}} \\ &= \prod_{ij} \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} \left(\frac{e^{\psi_i}}{e^{\psi_i} + e^{\psi_j}} \right)^{\alpha_{ij}} \left(\frac{e^{\psi_j}}{e^{\psi_i} + e^{\psi_j}} \right)^{\alpha_{ji}}. \end{aligned}$$

The log-likelihood function is then:

$$\ell(\{\alpha_{ij}\}|\{\psi_i\}) = \sum_{ij} \log \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} + \alpha_{ij} \log e^{\psi_i} + \alpha_{ji} \log e^{\psi_j} - n_{ij} \log \left(e^{\psi_i} + e^{\psi_j} \right).$$

Define $p_i := e^{\psi_i}$, $i = 1, ..., J$. We can define the log-likelihood function in terms of $\{p_i\}$ instead:

$$\ell(\{\alpha_{ij}\}|\{p_i\}) = \sum_{ij} \log \begin{pmatrix} n_{ij} \\ \alpha_{ij} \end{pmatrix} + \alpha_{ij} \log p_i + \alpha_{ji} \log p_j - n_{ij} \log \left(p_i + p_j \right).$$

Take the first-order derivative of p_i:

$$\frac{\partial \ell}{\partial p_i} = \sum_j \frac{\alpha_{ij}}{p_i} - \frac{n_{ij}}{p_i + p_j}.$$

Solving p_i for $\frac{\partial \ell}{\partial p_i} = 0$, we get:

$$p_i = \frac{\sum_j \alpha_{ij}}{\sum_j \frac{n_{ij}}{p_i + p_j}}.$$

We can solve for p_j (or equivalently ψ_i) iteratively.

To obtain the confidence interval of the MLE solution, we first calculate the second-order partial derivatives and construct the hessian matrix. The second-order derivatives are

$$\begin{aligned} \lambda_{ii} = \frac{\partial^2 \ell}{\partial p_i^2} &= \sum_j -\frac{\alpha_{ij}}{p_i^2} + \frac{n_{ij}}{(p_i + p_j)^2}, \\ \lambda_{ij} = \frac{\partial^2 \ell}{\partial p_i \partial p_j} &= \frac{n_{ij}}{(p_i + p_j)^2}. \end{aligned}$$

Construct the Hessian matrix $H = [\lambda_{ij}]$. Since their is one degree of freedom dependency in $\{\psi_i\}$ since it is scaling-invariant, we construct

$$C = \left[\begin{array}{cc} -H & 1 \\ 1' & 0 \end{array} \right]^{-1}$$

of dimension $(J+1) \times (J+1)$, and the estimated variances of p_i, is

$$Var(p_i) = diag(C)[i],$$

for $i = 1, ..., J$. Since $\psi_i = \log p_i$, we have $d\psi_i/dp_i = 1/p_i$. Then

$$Var(\psi_i) = Var(p_i)/p_i^2.$$