Random utility model (RUM)

The random utility model (RUM) employs a revealed preference argument to describe the conditional likelihood associated with discrete choices. We'll focus on the binary choice case, $D = 0$ or 1. Consider a decision framed by reference to a decision maker's preferences or utility.

$$D_i^* = \gamma^T X_i - \nu_i$$

such that observed choice $D = 1$ when $D_i^* > 0$ and $D = 0$ otherwise. We say D_i^* (utility) is a latent (unobservable to the analyst) variable, $\gamma^T X_i$ is the observable component (even if γ is unknown), and ν_i is the unobservable component of utility for individual i.

For known γ

$$\begin{aligned}
\Pr(D = 1 \mid \gamma, X_i) &= \Pr(D_i^* > 0 \mid \gamma, X_i) \\
&= \Pr\left(\gamma^T X_i - \nu_i > 0 \mid \gamma, X_i\right) \\
&= \Pr\left(\nu_i < \gamma^T X_i \mid \gamma, X_i\right)
\end{aligned}$$

and

$$\Pr(D = 0 \mid \gamma, X_i) = 1 - \Pr\left(\nu_i < \gamma^T X_i \mid \gamma, X_i\right)$$

This leaves the distribution for ν to be assigned. Suppose the ν_i are exchangeable (alternatively, we don't know they are not) then we can assign a homogeneous distribution for ν, or in other words, treat draws for the various individuals as if their unobservable (components of) utilities are from the same population.

What is the population assignment? We offer two observations. First, discrete states combined with normally distributed information yield a logistic posterior distribution. Second, the maximum entropy assignment for a discrete choice is a logistic distribution. Hence, we assign a logistic distribution to ν. This probability assignment for ν is often referred to as the link function.

$$p_i \equiv \Pr\left(\nu_i < \gamma^T X_i \mid \gamma, X_i\right) = \frac{1}{1 + \exp\left[-\gamma^T X_i\right]}$$

A Bernoulli distribution or binomial distribution involving a single draw frames the probability story.

$$\Pr(D_i \mid \gamma, X_i) = p_i^{D_i} (1 - p_i)^{1 - D_i}, \quad D_i = 0, 1$$

A sample of n exchangeable draws is described by

$$\prod_{i=1}^{n} \Pr(D_i \mid \gamma, X_i) = \prod_{i=1}^{n} p_i^{D_i} (1 - p_i)^{1 - D_i}$$

The above describes the sampling distribution.

If γ is unknown, we wish to assess the informativeness of a specific sample for the unknown parameters, γ. This is referred to as the likelihood function.

$$\ell(\gamma \mid D, X) = \prod_{i=1}^{n} p_i^{D_i} (1 - p_i)^{1 - D_i}$$

The likelihood function has the same form as the sampling distribution but a different interpretation. A sensible way to guess γ is to find the most likely $\gamma' s$ conditional on the evidence. This is called maximum likelihood estimation (MLE).

$$\max_{\gamma} \ell(\gamma \mid D, X) \equiv \max_{\gamma} \prod_{i=1}^{n} p_i^{D_i} (1 - p_i)^{1 - D_i}$$

It is usually simpler to work with the log-likelihood (a monotone transformation of the objective function).

$$\max_{\gamma} \sum_{i=1}^{n} D_i \log p_i + (1 - D_i) \log(1 - p_i)$$

Substituting the logistic link function gives

$$\max_{\gamma} \sum_{i=1}^{n} D_i \log \frac{1}{1 + \exp[-\gamma^T X_i]} + (1 - D_i) \log \frac{1}{1 + \exp[\gamma^T X_i]}$$

First order conditions (*foc*) for a maximum involve setting the gradient (vector of partial derivatives with respect to the parameters) equal to zero.

$$\nabla \equiv \begin{bmatrix} \frac{\partial \log \ell}{\partial \gamma_1} \\ \frac{\partial \log \ell}{\partial \gamma_2} \\ \vdots \\ \frac{\partial \log \ell}{\partial \gamma_k} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where a typical element is

$$\frac{\partial \log \ell}{\partial \gamma_j} = \sum_{i=1}^{n} \frac{\left[D_i - F\left(\gamma^T X_i\right)\right] f\left(\gamma^T X_i\right) X_{ji}}{F\left(\gamma^T X_i\right)\left[1 - F\left(\gamma^T X_i\right)\right]} = 0$$

$$F\left(\gamma^T X_i\right) = \frac{1}{1 + \exp[-\gamma^T X_i]}$$

and density function

$$f\left(\gamma^T X_i\right) = \frac{\exp\left[-\gamma^T X_i\right]}{\left(1 + \exp[-\gamma^T X_i]\right)^2} = \exp\left[-\gamma^T X_i\right] F\left(\gamma^T X_i\right)^2$$

For the logistic distribution, the log-likelihood is smooth and globally concave. Therefore, solutions to the *foc* are straightforward as long as there are no regressors, X_j, that are perfect explanators for choice. Solving the *foc* yields the k parameter estimates $\widehat{\gamma}$.

Variation in the parameter estimates is reflected in the negative inverse of the Hessian (matrix of second partial derivatives with respect to the parameters) evaluated at MLE, $\widehat{\gamma}$.

$$\begin{aligned} Var\left[\widehat{\gamma} \mid D, X\right] &= -H\left(\widehat{\gamma}\right)^{-1} \\ &= -\begin{bmatrix} \frac{\partial^2 \log \ell}{\partial\gamma_1\partial\gamma_1} & \frac{\partial^2 \log \ell}{\partial\gamma_1\partial\gamma_2} & \cdots & \frac{\partial^2 \log \ell}{\partial\gamma_1\partial\gamma_k} \\ \frac{\partial^2 \log \ell}{\partial\gamma_2\partial\gamma_1} & \frac{\partial^2 \log \ell}{\partial\gamma_2\partial\gamma_2} & \cdots & \frac{\partial^2 \log \ell}{\partial\gamma_2\partial\gamma_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \log \ell}{\partial\gamma_k\partial\gamma_1} & \frac{\partial^2 \log \ell}{\partial\gamma_k\partial\gamma_2} & \cdots & \frac{\partial^2 \log \ell}{\partial\gamma_k\partial\gamma_k} \end{bmatrix}_{\widehat{\gamma}}^{-1} \end{aligned}$$

This matrix is always square ($k \times k$), symmetric and is invertible if there is no redundant information in $H\left(\widehat{\gamma}\right)$. In other words, if $-H\left(\widehat{\gamma}\right)$ is positive definite (has only strictly positive eigenvalues or $x^T\left(-H\left(\widehat{\gamma}\right)\right)x > 0$ for all $x \neq 0$). The estimated standard errors associated with $\widehat{\gamma}$ are the square roots of the main diagonal elements of $-H\left(\widehat{\gamma}\right)^{-1}$.