Transition choice probabilities in logit

Paolo Delle Site *, Marco Valerio Salucci

DICEA - Department of Civil, Architectural and Environmental Engineering, University "La Sapienza", Rome, Italy

HIGHLIGHTS

- Random terms are i.i.d. Gumbel across alternatives for each choice.
- Random terms are bi-extremal correlated across choices for each alternative.
- First- and second-choice probabilities are multinomial logit.
- Transition probabilities are in analytic form.
- Correlation coefficient varies in the full positive range between 0 and 1.

ARTICLE INFO

ABSTRACT

We consider the widely used multinomial logit model with i.i.d. Gumbel random terms. Transition choice probabilities, i.e. probabilities of choosing alternative i in the first choice and alternative j in the second, are available in analytic form in the two extreme cases where the random terms of each alternative are independent or perfectly correlated across choices. We extend these results and provide the transition probabilities in analytic form in the case where the random terms follow a bi-extremal distribution with correlation coefficient varying in the full positive range between zero and one.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Logit is the most popular of the discrete choice models derived from random utility maximisation. In logit, where random terms are i.i.d. according to a type I extreme value distribution, a.k.a. Gumbel distribution, choice probabilities take a particularly simple analytic form.

In this paper, we consider the sequence of two choices and the related transition choice probabilities, i.e. the probabilities of choosing alternative i in the first choice and alternative j in the second. De Palma and Kilani (2005, 2011) prove that, in logit, the transition probabilities also take an analytic form. However, their results apply to the special case where the random terms remain the same over the two choices, i.e. are perfectly correlated.

The assumption of perfectly correlated random terms is restrictive because unobservables, and related tastes, may change from one choice to another. Delle Site and Salucci (2013) relax this assumption, and provide the transition probabilities in terms of the joint cumulative distribution function of the random terms over the two choices. They show that the multiple integrals that express the transition probabilities can be solved by simulation if it is possible to draw from the joint distribution of the random terms. In the special case where the vector of the random terms in the first choice is independent of the vector of the random terms in the second choice, they prove that the transition probabilities are simply the product of the probability of choosing i in the first choice times the probability of choosing j in the second choice.

With logit, different bivariate distributions having Gumbel as univariate marginals are available to treat cases where the random terms are imperfectly correlated over the two choices. Bivariate extreme value distributions are reviewed in Kotz et al. (2000). Additionally, it is possible to construct a bivariate distribution with Gumbel marginals using copulas, on the basis of Sklar's theorem (Mari and Kotz, 2001; Nelsen, 1999).

* Correspondence to: DICEA - Department of Civil, Architectural and Environmental Engineering, University "La Sapienza". Via Eudossiana 18, 00184 Rome, Italy. Tel.: +39 06 445858923; fax: +39 06 445858774.

E-mail address: paolo.dellesite@uniroma1.it (P. Delle Site).

The paper shows that, if a bi-extremal distribution is assumed, the transition probabilities take an analytic form. The bi-extremal distribution has been investigated extensively in a number of papers by Tiago de Oliveira (a review is in Tiago de Oliveira, 1970). This distribution is related to sequences with Gumbel marginals generated by maximisation procedures in an autoregressive way. It includes as special cases independence (Pearson's product-moment correlation coefficient equal to zero) and perfect correlation (correlation coefficient equal to one). Therefore, the paper generalises existing results, since the correlation coefficient takes any value between zero and one.

It is possible to see that bivariate distributions based on Farlie–Gumbel–Morgenstern (FGM) copulas also yield transition probabilities in analytic form. However, FGM copulas have the limitation that the correlation coefficient has a low maximal value, significantly below one. Extensions of the FGM copulas have been designed to increase the maximal value of the correlation, but the full positive range of correlation is not modelled.

The paper is organised as follows. Section 2 provides notation and assumptions on the distribution of the random terms. Section 3 derives transition probabilities. Section 4 provides a numerical illustration. Section 5 concludes.

2. Notation and assumptions

Let $k = 1, \ldots, J$ denote the alternatives. Let s and t denote, respectively, the first and the second choice. For alternative k, the first-choice utility is expressed by $u_k(s) \equiv v_k(s) + \epsilon_k(s)$, where $v_k(s)$ is the systematic utility, $\epsilon_k(s)$ is the random term. Similarly, the second-choice utility is expressed by $u_k(t) \equiv v_k(t) + \epsilon_k(t)$.

Choice probabilities depend on the joint distribution of the $2 \times J$-dimensional random vector $[\epsilon_1(s), \epsilon_1(t), \ldots, \epsilon_J(s), \epsilon_J(t)]$. We construct the joint distribution based on assumptions on marginal distributions. Multivariate statistics theory (Bilodeau and Brenner, 1999) provides definitions of multivariate and univariate marginal distributions.

Assumption 1. The vector $[\epsilon_k(s), \epsilon_k(t)]$, $k = 1, \ldots, J$, follows a bi-extremal distribution with scale parameter λ and correlation parameter ϕ.

Assumption 1 has the following implications. The first-choice random term $\epsilon_k(s)$ is Gumbel distributed with scale parameter λ. The marginal c.d.f. of $\epsilon_k(s)$ is $F_{\epsilon_k(s)}(x) = e^{-e^{-x/\lambda}}$, the mean is $\gamma \cdot \lambda$, the standard deviation is $\pi \cdot \lambda / \sqrt{6}$, where $\gamma = 0.5772$ is the Euler constant. A standard Gumbel is obtained when $\lambda = 1$.

The second-choice random term $\epsilon_k(t)$ is generated according to the rule:

$$\epsilon_k(t) = \max\left[\epsilon_k(s) + \lambda \cdot \ln \phi, \eta_k + \lambda \cdot \ln(1 - \phi)\right] \quad (1)$$

where $\epsilon_k(s)$ and η_k are i.i.d., and $\phi \in [0, 1]$.

When $\phi = 0$, $\epsilon_k(s)$ and $\epsilon_k(t)$ are i.i.d. since $\epsilon_k(t) = \eta_k$. When $\phi = 1$, $\epsilon_k(t) = \epsilon_k(s)$, i.e. the random terms are perfectly correlated. The parameter ϕ has a correlation meaning because Pearson's product-moment correlation coefficient ρ increases monotonically with ϕ according to the expression (proof in Tiago de Oliveira, 1970):

$$\rho(\phi) = -\frac{6}{\pi^2} \cdot \int_0^\phi \frac{\ln z}{1 - z} dz \quad (2)$$

which yields $\rho = 0$ when $\phi = 0$, and $\rho = 1$ when $\phi = 1$.

Based on Eq. (1), the marginal c.d.f. of the vector $[\epsilon_k(s), \epsilon_k(t)]$ is:

$$
\begin{aligned}
F_{\epsilon_k(s),\epsilon_k(t)}(x, y) &\equiv \mathbb{P}(\epsilon_k(s) \leq x, \epsilon_k(t) \leq y) \\
&= \mathbb{P}(\epsilon_k(s) \leq \min[x, y - \lambda \cdot \ln \phi], \\
&\qquad \eta_k \leq y - \lambda \cdot \ln(1 - \phi)) \\
&= e^{-e^{-\frac{1}{\lambda} \cdot \min[x, y - \lambda \cdot \ln \phi]}} \cdot e^{-e^{-\frac{1}{\lambda} \cdot (y - \lambda \cdot \ln(1-\phi))}}.
\end{aligned} \quad (3)
$$

As a consequence of Eq. (3), the marginal c.d.f. of $\epsilon_k(t)$ is also Gumbel with scale parameter λ: $F_{\epsilon_k(t)} = \mathbb{P}(\epsilon_k(s) \leq +\infty, \epsilon_k(t) \leq y) = e^{-e^{-y/\lambda}}$.

The bi-extremal distribution is characterised by a behavioural interpretation when the random terms are regarded as individual specific. Consider a population of individuals with identical systematic part of the utilities. Each individual changes her random terms, i.e. her perception of unobservables, from one choice to another. Based on the drawing rule of Eq. (1), there is an individual-specific minimum perception threshold for unobservables across repeated choices for each alternative equal to $\tau_k = \epsilon_k(s) + \lambda \cdot \ln \phi$. In fact, in the second choice the perception of unobservables is never lower than τ_k. In the first choice the perception of unobservables is $\epsilon_k(s) > \tau_k$, because the natural logarithm of ϕ is negative when $\phi \in [0, 1]$.

Assumption 2. The vector $[\epsilon_i(s), \epsilon_i(t)]$ of the random terms of alternative i and the vector $[\epsilon_j(s), \epsilon_j(t)]$ of the random terms of alternative j are independent, for any pair of alternatives i and j, $i, j = 1, \ldots, J, i \neq j$.

Assumption 2 implies that the joint c.d.f. of the $2 \times J$-dimensional vector $[\epsilon_1(s), \epsilon_1(t), \ldots, \epsilon_J(s), \epsilon_J(t)]$ is:

$$F_{\epsilon_1(s),\epsilon_1(t),\ldots,\epsilon_J(s),\epsilon_J(t)}(x_1, y_1, \ldots, x_J, y_J) = \prod_{k=1}^{J} F_{\epsilon_k(s),\epsilon_k(t)}(x_k, y_k). \quad (4)$$

3. Choice probabilities

Since both the first- and second-choice marginal distributions of the random terms are i.i.d. Gumbel, the first- and second-choice probabilities $P_k(s), P_k(t), k = 1, \ldots J$, are multinomial logit. In the following, we assume the scale parameter $\lambda = 1$. This is justified since this parameter is not identified with maximum likelihood. As it is easily seen, the scale parameter only rescales utilities in each choice.

To derive the transition choice probabilities $P_{ij}(s, t) \equiv \mathbb{P}(u_i(s) \geq u_k(s), k \neq i; u_j(t) \geq u_k(t), k \neq j)$, the following lemma is needed.

Lemma 1. *If the vectors $[u_k(s), u_k(t)]$, $k = 1, ..J$, of utilities admit a p.d.f. $f_{u_k(s),u_k(t)}(X_k, Y_k)$, $k = 1, \ldots, J$, and are independent across alternatives, then the transition choice probabilities are expressed in terms of the c.d.f. $F_{u_k(s),u_k(t)}(x_k, y_k)$ of the vectors of utilities by:*

$$
\begin{aligned}
P_{ij}(s, t) = \int_{x_i=-\infty}^{+\infty} \int_{y_j=-\infty}^{+\infty} \prod_{k=1; k \neq i,j}^{J} F_{u_k(s),u_k(t)}(x_i, y_j) \\
\cdot \frac{\partial F_{u_i(s),u_i(t)}(x_i, y_j)}{\partial x_i} \cdot \frac{\partial F_{u_j(s),u_j(t)}(x_i, y_j)}{\partial y_j} dx_i dy_j.
\end{aligned} \quad (5)
$$

Proof. First, consider that, under the independence across alternatives assumption, the p.d.f. of the $2 \times J$-dimensional vector $[u_1(s), u_1(t), \ldots, u_J(s), u_J(t)]$ of utilities is:

$$f_{u_1(s),u_1(t),\ldots,u_J(s),u_J(t)}(X_1, Y_1, \ldots, X_J, Y_J) = \prod_{k=1}^{J} f_{u_k(s),u_k(t)}(X_k, Y_k). \quad (6)$$

Thus, the transition choice probability is given by:

$$P_{ij}(s, t) = \int \ldots \int_S \prod_{k=1}^{J} f_{u_k(s),u_k(t)}(X_k, Y_k) dX_k dY_k \quad (7)$$

where the set $S \subset \mathbb{R}^{2 \times J}$ is defined by:

$$
\begin{aligned}
S &\equiv \left\{ u_k(s) \leq u_i(s), k \neq i; u_k(t) \leq u_j(t), k \neq j \right\} \\
&= \left\{ X_k \leq X_i = x_i, k \neq i; Y_k \leq Y_j = y_j, k \neq j \right\}. \quad (8)
\end{aligned}
$$