

## Transition choice probabilities in logit

This is the author's accepted manuscript of the paper: Delle Site P., Salucci M.V. (2015)  
Transition choice probabilities in logit. *Economics Letters* 126, 135-139.

# Transition choice probabilities in logit

Paolo Delle Site\* and Marco Valerio Salucci

November 2, 2014

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

## 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.

**Keywords:** discrete choice, random utility, multinomial logit, transition probability, bi-extremal distribution

**JEL code:** D11

---

\*Corresponding author at: 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. Email: paolo.dellesite@uniroma1.it

# 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).

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  $k = 1, \dots, J$  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  $[\epsilon_1(s), \epsilon_1(t), \dots, \epsilon_J(s), \epsilon_J(t)]$ . We construct the joint distribution based on assumptions on marginal distributions. Multivariate statistics theory (Bilodeau and Brenner, 1999) provide definitions of multivariate and univariate marginal distributions.

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

Assumption 1 has the following implications. The first-choice random term  $\epsilon_k(s)$  is Gumbel distributed with scale parameter  $\lambda$ . 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:

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

where  $\epsilon_k(s)$  and  $\eta_k$  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  $\phi$  has a correlation meaning because the Pearson's product-moment correlation coefficient  $\rho$  increases monotonically with  $\phi$  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:

$$F_{\epsilon_k(s), \epsilon_k(t)}(x, y) \equiv$$

$$\mathbb{P}(\epsilon_k(s) \leq x, \epsilon_k(t) \leq y) = \mathbb{P}(\epsilon_k(s) \leq \min[x, y - \lambda \cdot \ln \phi], \eta_k \leq y - \lambda \cdot \ln(1 - \phi)) =$$

$$e^{-e^{-\frac{1}{\lambda} \cdot \min[x, y - \lambda \cdot \ln \phi]}} \cdot e^{-e^{-\frac{1}{\lambda} \cdot (y - \lambda \cdot \ln(1 - \phi))}} \quad (3)$$

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

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  $\tau_k$ . In the first choice the perception of unobservables is  $\epsilon_k(s) > \tau_k$ , because the natural logarithm of  $\phi$  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  $j$ ,  $i, j = 1, \dots, J, i \neq j$ .*

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

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

### 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 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  $[u_k(s), u_k(t)]$ ,  $k = 1, \dots, J$ , of utilities admit a p.d.f.  $f_{u_k(s), u_k(t)}(X_k, Y_k)$ ,  $k = 1, \dots, J$ , 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:*

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

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

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

Thus, the transition choice probability is given by:

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

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

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

By the rules of reduction of multiple integrals to iterated integrals (Zorich, 2004), we have:

$$\begin{aligned}
P_{ij}(s, t) &= \int_{x_i=-\infty}^{+\infty} \int_{y_j=-\infty}^{+\infty} \left[ \int_{X_1=-\infty}^{x_i} \cdots \int_{X_{k \neq i}=-\infty}^{x_i} \cdots \int_{X_J=-\infty}^{x_i} \int_{Y_1=-\infty}^{y_j} \cdots \int_{Y_{k \neq j}=-\infty}^{y_j} \cdots \int_{Y_J=-\infty}^{y_j} \right. \\
&\quad \left. \left[ \prod_{k=1; k \neq i, j}^J f_{u_k(s), u_k(t)}(X_k, Y_k) \right] \cdot f_{u_i(s), u_i(t)}(x_i, Y_i) \cdot f_{u_j(s), u_j(t)}(X_j, y_j) \right. \\
&\quad \left. dX_1 \dots dX_{k \neq i} \dots dX_J dY_1 \dots dY_{k \neq j} \dots dY_J \right] dx_i dy_j = \\
&= \int_{x_i=-\infty}^{+\infty} \int_{y_j=-\infty}^{+\infty} \left[ \prod_{k=1; k \neq i, j}^J \left( \int_{X_k=-\infty}^{x_i} \int_{Y_k=-\infty}^{y_j} f_{u_k(s), u_k(t)}(X_k, Y_k) dX_k dY_k \right) \right. \\
&\quad \left. \times \int_{Y_i=-\infty}^{y_j} f_{u_i(s), u_i(t)}(x_i, Y_i) dY_i \cdot \int_{X_j=-\infty}^{x_i} f_{u_j(s), u_j(t)}(X_j, y_j) dX_j \right] dx_i dy_j = \\
&= \int_{x_i=-\infty}^{+\infty} \int_{y_j=-\infty}^{+\infty} \prod_{k=1; k \neq i, j}^J F_{u_k(s), u_k(t)}(x_i, y_j) \\
&\quad \times \frac{\partial F_{u_i(s), u_i(t)}(x_i, y_j)}{\partial x_i} \cdot \frac{\partial F_{u_j(s), u_j(t)}(x_i, y_j)}{\partial y_j} dx_i dy_j \tag{9}
\end{aligned}$$

where the last equality is consequence of the following:

$$\begin{aligned}
\frac{\partial F_{u_i(s), u_i(t)}(x_i, y_j)}{\partial x_i} &= \frac{\partial}{\partial x_i} \int_{X_i=-\infty}^{x_i} \int_{Y_i=-\infty}^{y_j} f_{u_i(s), u_i(t)}(X_i, Y_i) dX_i dY_i = \\
&= \int_{Y_i=-\infty}^{y_j} f_{u_i(s), u_i(t)}(x_i, Y_i) dY_i \tag{10}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial F_{u_j(s), u_j(t)}(x_i, y_j)}{\partial y_j} &= \frac{\partial}{\partial y_j} \int_{X_j=-\infty}^{x_i} \int_{Y_j=-\infty}^{y_j} f_{u_j(s), u_j(t)}(X_j, Y_j) dX_j dY_j = \\
&= \int_{X_j=-\infty}^{x_i} f_{u_j(s), u_j(t)}(X_j, y_j) dX_j \tag{11}
\end{aligned}$$

Notice that we have used capital letters to denote arguments of the p.d.f., lower case letters to denote the specific values that are arguments of the c.d.f.  $\square$

The following proposition provides the transition probabilities in the model with the assumptions of Section 2. Let  $I(\cdot)$  be the indicator function.

**Proposition 1.** *In a model where the distribution of the random terms satisfies assumptions 1 and 2, and with the ranking of the alternatives  $a_1 \leq \dots \leq a_J$ ,  $a_k = v_k(s) - v_k(t) - \ln \phi$ ,  $k = 1, \dots, J$ , the transition choice probabilities are given by:*

$$P_{ij}(s, t) = J_0 + \sum_{r=1}^{J-1} [I(i > r, j \leq r) \cdot J_{r1} + I(i > r, j \geq r + 1) \cdot J_{r2}] \quad (12)$$

$$P_{ii}(s, t) = P_i(s) - \sum_{j=1; j \neq i}^J P_{ij}(s, t) \quad (13)$$

$$i, j = 1, \dots, J; \quad i \neq j; \quad \phi \in [0, 1]$$

where

$$J_0 = \frac{(1 - \phi) \cdot e^{v_i(s) + v_j(t)}}{\left( \sum_{k=1}^J e^{v_k(s)} \right) \cdot \left[ \phi \cdot e^{-v_1(s) + v_1(t)} \cdot \sum_{k=1}^J e^{v_k(s)} + (1 - \phi) \cdot \sum_{k=1}^J e^{v_k(t)} \right]} \quad (14)$$

$$J_{r1} = \frac{e^{v_i(s) + v_j(t)}}{\sum_{k=r+1; k \neq i, j}^J e^{v_k(s)} + e^{v_i(s)}} \cdot \left[ \frac{1}{e^{-v_{r+1}(s) + v_{r+1}(t)} \cdot \xi_1 + \xi_2} - \frac{1}{e^{-v_r(s) + v_r(t)} \cdot \xi_1 + \xi_2} \right] \quad (15)$$

$$\xi_1 = \phi \cdot \left( \sum_{k=r+1; k \neq i, j}^J e^{v_k(s)} + e^{v_i(s)} \right) \quad (16)$$

$$\xi_2 = \sum_{k=1; k \neq i, j}^r e^{v_k(t)} + e^{v_j(t)} + (1 - \phi) \cdot \left( \sum_{k=r+1; k \neq i, j}^J e^{v_k(t)} + e^{v_i(t)} \right) \quad (17)$$

$$J_{r2} = \frac{(1 - \phi) \cdot e^{v_i(s) + v_j(t)}}{\sum_{k=r+1; k \neq i, j}^J e^{v_k(s)} + e^{v_i(s)} + e^{v_j(s)}} \cdot \left[ \frac{1}{e^{-v_{r+1}(s) + v_{r+1}(t)} \cdot \xi_3 + \xi_4} - \frac{1}{e^{-v_r(s) + v_r(t)} \cdot \xi_3 + \xi_4} \right] \quad (18)$$



$$\xi_3 = \phi \cdot \left( \sum_{k=r+1; k \neq i, j}^J e^{v_k(s)} + e^{v_i(s)} + e^{v_j(s)} \right) \quad (19)$$

$$\xi_4 = \sum_{k=1; k \neq i, j}^r e^{v_k(t)} + (1 - \phi) \cdot \left( \sum_{k=r+1; k \neq i, j}^J e^{v_k(t)} + e^{v_i(t)} + e^{v_j(t)} \right) \quad (20)$$

*Proof.* We only need to prove Eq. (12), because Eq. (13) is an application of proposition 1 in Delle Site and Salucci (2013).

First, we provide the c.d.f. of the bivariate vector  $[u_k(s), u_k(t)]$  of the random utilities of alternative  $k$  under assumption 1. We have:

$$\begin{aligned} F_{u_k(s), u_k(t)}(x, y) &\equiv \mathbb{P}(u_k(s) \leq x, u_k(t) \leq y) = \\ &\mathbb{P}(v_k(s) + \epsilon_k(s) \leq x, v_k(t) + \epsilon_k(t) \leq y) = \\ &e^{-e^{-\min[x-v_k(s), y-v_k(t)] - \ln \phi}} \cdot e^{-e^{-y+v_k(t)+\ln(1-\phi)}} \end{aligned} \quad (21)$$

where we have used the c.d.f. of the vector of the random terms  $[\epsilon_k(s), \epsilon_k(t)]$  given by Eq. (3).

The c.d.f.  $F_{u_k(s), u_k(t)}(x, y)$  is non differentiable, and, therefore, does not admit a p.d.f. To use lemma 1 and obtain the transition probabilities, we need to partition the 2-dimensional Euclidean space  $\mathbb{R}^2$  into regions  $\Omega_r$  where the c.d.f. of the bivariate vector of random terms of each alternative is, at least twice, differentiable. We denote by  $F_{u_k(s), u_k(t)}^r(x_i, y_j)$  the (differentiable) c.d.f. of the bivariate vector  $[u_k(s), u_k(t)]$  over the region  $\Omega_r \subset \mathbb{R}^2$ .

For ease of notation let:

$$F_k^r \equiv F_{u_k(s), u_k(t)}^r(x_i, y_j) \quad (22)$$

$$F_{i,x}^r \equiv \frac{\partial F_{u_i(s), u_i(t)}^r(x_i, y_j)}{\partial x_i} \quad (23)$$

$$F_{j,y}^r \equiv \frac{\partial F_{u_j(s), u_j(t)}^r(x_i, y_j)}{\partial y_j} \quad (24)$$

Based on Eq. (21) we have:

$$F_k^r = \begin{cases} F_k^x & \text{if } x_i - v_k(s) \leq y_j - v_k(t) - \ln \phi \\ F_k^y & \text{otherwise} \end{cases} \quad (25)$$

with

$$F_k^x = e^{-e^{-x_i+v_k(s)}} \cdot e^{-e^{-y_j+v_k(t)+\ln(1-\phi)}} \quad (26)$$

$$F_k^y = e^{-e^{-y_j+v_k(t)+\ln \phi}} \cdot e^{-e^{-y_j+v_k(t)+\ln(1-\phi)}} = e^{-e^{-y_j+v_k(t)}} \quad (27)$$

We also have:

$$F_{i,x}^r = \begin{cases} F_{i,x}^x & \text{if } x_i - v_i(s) \leq y_j - v_i(t) - \ln \phi \\ F_{i,x}^y & \text{otherwise} \end{cases} \quad (28)$$

with

$$F_{i,x}^x = -e^{-x_i+v_i(s)} \cdot e^{-e^{-x_i+v_i(s)}} \cdot e^{-e^{-y_j+v_i(t)+\ln(1-\phi)}} \quad (29)$$

$$F_{i,x}^y = 0 \quad (30)$$

and:

$$F_{j,y}^r = \begin{cases} F_{j,y}^x & \text{if } x_i - v_j(s) \leq y_j - v_j(t) - \ln \phi \\ F_{j,y}^y & \text{otherwise} \end{cases} \quad (31)$$

with

$$F_{j,y}^x = -e^{-e^{-x_i+v_j(s)}} \cdot e^{-y_j+v_j(t)+\ln(1-\phi)} \cdot e^{-e^{-y_j+v_j(t)+\ln(1-\phi)}} \quad (32)$$

$$F_{j,y}^y = -e^{-y_j+v_j(t)} \cdot e^{-e^{-y_j+v_j(t)}} \quad (33)$$

We assume w.l.o.g. the ranking  $a_1 \leq \dots \leq a_J$ , with  $a_k = v_k(s) - v_k(t) - \ln \phi$ ,  $k = 1, \dots, J$ . We partition the 2-dimensional Euclidean space  $\mathbb{R}^2$  into the following sets:

$$\Omega_0 \equiv \{x_i \leq y_j + a_1\}$$

$$\Omega_r \equiv \{y_j + a_r \leq x_i \leq y_j + a_{r+1}\} \quad r = 1, \dots, J-1$$

$$\Omega_J \equiv \{x_i \geq y_j + a_J\} \quad (34)$$

We have:

$$\text{in } \Omega_0 : \begin{cases} F_k^0 = F_k^x & k \neq i, j \\ F_{i,x}^0 = F_{i,x}^x \\ F_{j,y}^0 = F_{j,y}^x \end{cases} \quad (35)$$

$$\text{in } \Omega_r, \quad r = 1, \dots, J-1 : \begin{cases} F_k^r = F_k^y & k = 1, \dots, r; \quad k \neq i, j \\ F_k^r = F_k^x & k = r+1, \dots, J; \quad k \neq i, j \\ F_{i,x}^r = F_{i,x}^y = 0 & \text{if } (i \leq r) \\ F_{i,x}^r = F_{i,x}^x \text{ and } F_{j,y}^r = F_{j,y}^y & \text{if } (i > r, j \leq r) \\ F_{i,x}^r = F_{i,x}^x \text{ and } F_{j,y}^r = F_{j,y}^x & \text{if } (i > r, j \geq r+1) \end{cases} \quad (36)$$

$$\text{in } \Omega_J : \begin{cases} F_k^J = F_k^y & k \neq i, j \\ F_{i,x}^J = F_{i,x}^y = 0 \\ F_{j,y}^J = F_{j,y}^y \end{cases} \quad (37)$$

Using lemma 1 we can write the transition probabilities as a sum of double integrals:

$$P_{ij}(s, t) = \sum_{r=0}^J \int \int_{\Omega_r} \prod_{k=1; k \neq i, j}^J F_k^r \cdot F_{i,x}^r \cdot F_{j,y}^r dx_i dy_j \quad (38)$$

Notice that if  $a_r = a_{r+1}$  the set  $\Omega_r$  reduces to a line in  $\mathbb{R}^2$  and, therefore, the integral on  $\Omega_r$  vanishes.

We have:

$$P_{ij}(s, t) = J_0 + \sum_{r=1}^{J-1} [I(i > r, j \leq r) \cdot J_{r1} + I(i > r, j \geq r + 1) \cdot J_{r2}] \quad (39)$$

with:

$$J_0 = \int_{y_j=-\infty}^{+\infty} \int_{x_i=-\infty}^{y_j+a_1} \prod_{k=1; k \neq i, j}^J F_k^x \cdot F_{i,x}^x \cdot F_{j,y}^x dx_i dy_j \quad (40)$$

$$J_{r1} = \int_{y_j=-\infty}^{+\infty} \int_{x_i=y_j+a_r}^{y_j+a_{r+1}} \prod_{k=1; k \neq i, j}^r F_k^y \cdot \prod_{k=r+1; k \neq i, j}^J F_k^x \cdot F_{i,x}^x \cdot F_{j,y}^y dx_i dy_j \quad (41)$$

$$J_{r2} = \int_{y_j=-\infty}^{+\infty} \int_{x_i=y_j+a_r}^{y_j+a_{r+1}} \prod_{k=1; k \neq i, j}^r F_k^y \cdot \prod_{k=r+1; k \neq i, j}^J F_k^x \cdot F_{i,x}^x \cdot F_{j,y}^x dx_i dy_j \quad (42)$$

The analytic solution of the double integrals of Eq. (40), (41) and (42) is a simple calculus exercise. Each integral has the following form where  $A, B, C, D$  and  $E$  are constants:

$$\begin{aligned} & \int_{y=-\infty}^{+\infty} \int_{x=y+A}^{y+B} C \cdot e^{-x} \cdot e^{-D \cdot e^{-x}} \cdot e^{-y} \cdot e^{-E \cdot e^{-y}} dx dy = \\ & = \frac{C}{D} \cdot \left( \frac{1}{D \cdot e^{-B} + E} - \frac{1}{D \cdot e^{-A} + E} \right) \end{aligned} \quad (43)$$

The rule of reduction of double integrals to iterated integrals is used. Therefore, we obtain, respectively, Eq. (14), (15) and (18)<sup>1</sup>.  $\square$

## 4 Numerical illustration

We provide a numerical example in the binomial case. Transition choice probabilities reduce, with any ranking of the two alternatives, to:

$$\begin{aligned} P_{ij}(s, t) &= \frac{e^{v_i(s)}}{e^{v_i(s)} + e^{v_j(s)}} \\ &\times \frac{(1 - \phi) \cdot e^{v_j(t)}}{\phi \cdot (e^{v_i(s)} + e^{v_j(s)}) \cdot e^{-\min[v_i(s)-v_i(t), v_j(s)-v_j(t)]} + (1 - \phi) \cdot (e^{v_i(t)} + e^{v_j(t)})} \end{aligned}$$

---

<sup>1</sup>Detailed solution of the integrals is available from the authors upon request.

$$+ \max \left[ 0, \frac{e^{v_j(t)}}{e^{v_i(t)} + e^{v_j(t)}} - \frac{e^{v_j(t)}}{e^{v_i(t)} + e^{v_j(t)} + \phi \cdot (e^{v_i(s)-v_j(s)+v_j(t)} - e^{v_i(t)})} \right] \quad (44)$$

$$P_{ii}(s, t) = P_i(s) - P_{ij}(s, t) \quad (45)$$

$$i, j = 1, 2; \quad i \neq j; \quad \phi \in [0, 1]$$

The following special cases are derived from Eq. (44). When the random terms of each alternative are independent across choices ( $\phi = 0$ ), we have:

$$P_{ij}(s, t) = \frac{e^{v_i(s)}}{e^{v_i(s)} + e^{v_j(s)}} \cdot \frac{e^{v_j(t)}}{e^{v_i(t)} + e^{v_j(t)}} = P_i(s) \cdot P_j(t) \quad (46)$$

which is the result obtained by Delle Site and Salucci (2013). When the random terms of each alternative are perfectly correlated across choices ( $\phi = 1$ ), we have:

$$P_{ij}(s, t) = \max \left[ 0, \frac{e^{v_j(t)}}{e^{v_i(t)} + e^{v_j(t)}} - \frac{e^{v_j(s)}}{e^{v_i(s)} + e^{v_j(s)}} \right] = \max [0, P_j(t) - P_j(s)] \quad (47)$$

which is the specialisation to two alternatives of the result obtained by De Palma and Kilani (2011; proposition 1).

Figure 1 shows the variation of the transition probabilities  $P_{12}(s, t)$  and  $P_{11}(s, t)$  with the parameter  $\phi$  when the systematic utilities take the values:  $v_1(s) = 0.6$ ,  $v_1(t) = 0.3$ ,  $v_2(s) = 0.3$ ,  $v_2(t) = 0.5$ . According to intuition, the probability of shifting from alternative 1 to alternative 2 decreases with correlation.

Figure 2 shows the variation, for five distinct values of the parameter  $\phi$ , of the transition probability  $P_{12}(s, t)$  with the systematic utility of alternative 1 in the first choice  $v_1(s)$ , the other systematic utilities being constant at the values above. The curves show a non-differentiability point at  $v_1(s) = 0.1$ , which is the point where the maximum and the minimum functions in Eq. (44) change argument. The only exception is the curve for  $\phi = 0$  which is everywhere differentiable.

## 5 Conclusions

The paper has provided the transition probabilities in a model where first- and second-choice probabilities are multinomial logit. The use of the bi-extremal distribution yields transition probabilities in analytic form. The correlation coefficient takes values in the full positive range between 0 and 1. As shown clearly by the numerical example, transition probabilities are non differentiable with respect to the estimation parameters in the systematic utilities.

Extension of the results to state dependence is straightforward. With second-choice systematic utilities depending on the alternative chosen in the first choice, transition probabilities can be computed using proposition 1 by appropriately evaluating the second-choice systematic utilities according to the transition, as in De Palma and Kilani (2005). Extension to logit with random coefficients also is straightforward, since proposition 1 provides kernels of probabilities.

Extension to sequences of multiple choices is of relevance to estimation, and is an area of future research. With repeated observations of the same individual and under an independence assumption across individuals, the likelihood is the product of sequence probabilities. Multiple choices can be modelled assuming a Markov process at the level of the random terms according to the drawing rule of Eq. (1). Simulations carried out by the authors have confirmed that the resulting choice process is not Markov.

In the absence of the Markov property at the choice level, factoring of sequence probabilities into transition probabilities, which is a convenient property of Markov chains (Norris, 1998), is not possible. Nevertheless, the use of the bi-extremal distribution yields sequence probabilities in analytic form, and the full range between 0 and 1 of the correlation coefficient is modelled. The limitation is that sequence probabilities, and the resulting likelihood, are non differentiable in the estimation parameters. This is problematic, in particular, for inference, since theorems that are commonly used to derive properties of maximum likelihood estimators assume that the likelihood is at least twice differentiable.

## Acknowledgements

The authors wish to thank Stefano D'Addona, Valerio Gatta, Anders Karlström, Edoardo Marcucci and the anonymous referee for comments. The usual disclaimer applies.

## References

- [1] Bilodeau M., Brenner D. (1999) Theory of Multivariate Statistics. Springer-Verlag, New York.
- [2] Delle Site P., Salucci M.V. (2013) Transition choice probabilities and welfare analysis in random utility models with imperfect before-after correlation. *Transportation Research Part B - Methodological* 58, 215-242.
- [3] De Palma A., Kilani K. (2005) Switching in the logit. *Economics Letters* 88(2), 196-202.
- [4] De Palma A., Kilani K. (2011) Transition choice probabilities and welfare analysis in additive random utility models. *Economic Theory* 46(3), 426-454.
- [5] Kotz S., Balakrishnan N., Johnson N.L. (2000) Continuous Multivariate Distributions. Volume 1: Models and Applications. John Wiley & Sons, New York.
- [6] Mari D.D., Kotz S. (2001) Correlation and Dependence. Imperial College Press, London.
- [7] Nelsen R.B. (1999) An Introduction to Copulas. Springer, New York.
- [8] Norris J.R. (1998) Markov Chains. Cambridge University Press, Cambridge UK.
- [9] Tiago de Oliveira J. (1970) Extremal distributions: statistical decisions. *Trabajos de Estadística* 21, 107-117.
- [10] Zorich V.A. (2004) Mathematical Analysis II. Springer, Berlin.

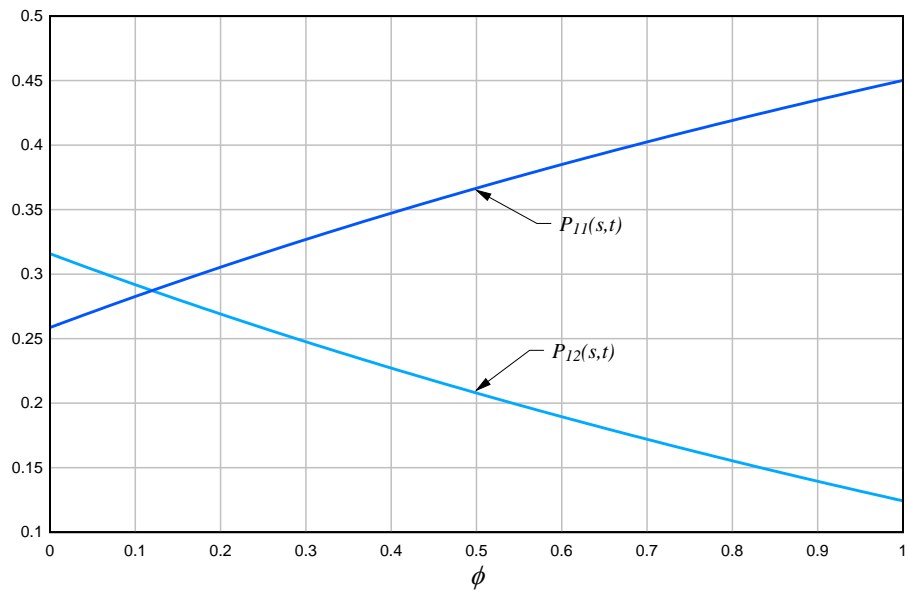


Figure 1: Transition probabilities  $P_{12}(s, t)$  and  $P_{11}(s, t)$  as function of  $\phi$



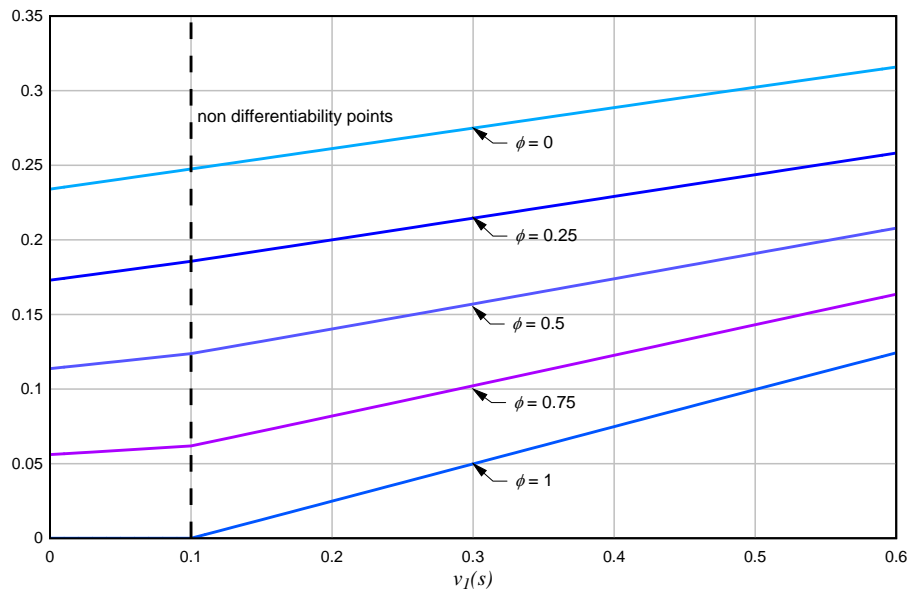


Figure 2: Transition probability  $P_{12}(s, t)$  as function of  $v_1(s)$