

Econometric Models of Probabilistic Choice: Beyond McFadden's Formulas

Olga Kosheleva, Vladik Kreinovich and Songsak Sriboonchitta

Abstract Traditional decision theory assumes that for every two alternatives, people always make the same (deterministic) choice. In practice, people's choices are often probabilistic, especially for similar alternatives: the same decision maker can sometimes select one of them and sometimes the other one. In many practical situations, an adequate description of this probabilistic choice can be provided by a logit model proposed by 2001 Nobelist D. McFadden. In this model, the probability of selecting an alternative a is proportional to $\exp(\beta \cdot u(a))$, where $u(a)$ is the alternative's utility. Recently, however, empirical evidence appeared that shows that in some situations, we need to go beyond McFadden's formulas. In this paper, we use natural symmetries to come up with an appropriate generalization of McFadden's formulas.

1 Need to Go Beyond McFadden's Probabilistic Choice Models: Formulation of the Problem

Traditional (deterministic choice) approach to decision making. In the traditional (deterministic) approach to decision making (see, e.g., [2, 4, 5, 8, 9]), we assume that for every two alternatives a and b :

- either the decision maker always prefers the alternative a ,
- or the decision maker always prefers the alternative b ,

O. Kosheleva
University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA
e-mail: olgak@utep.edu

V. Kreinovich (✉)
Department of Computer Science, University of Texas at El Paso,
500 W. University, El Paso, TX 79968, USA
e-mail: vladik@utep.edu

S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: songsakecon@gmail.com

- or the decision maker always states that the alternatives a and b are absolutely equivalent to him/her.

Under this assumption, preferences of a decision maker can be described by a utility function which can be defined as follows. We select two alternatives which are not present in the original choices:

- a very bad alternative a_0 , and
- a very good alternative a_1 .

Then, each actual alternative a is better than the very bad alternative a_0 and worse than the very good alternative a_1 : $a_0 < a < a_1$. To gauge the quality of the alternative a to the decision maker, we can consider lotteries $L(p)$ in which we get a_1 with probability p and a_0 with the remaining probability $1 - p$.

In accordance with our assumption, for every p , we either have $L(p) < a$ or $a < L(p)$, or we have equivalence $L(p) \sim a$.

When $p = 1$, the lottery $L(1)$ coincides with the very good alternative a_1 and is, thus, better than a : $a < L(1)$. When $p = 0$, the lottery $L(0)$ coincides with the very bad alternative a_0 and is, thus, worse than a : $L(0) < a$. Clearly, the larger the probability p of the very good outcome, the better the lottery; thus, if $p < p'$, then:

- $a < L(p)$ implies $a < L(p')$, and
- $L(p') < a$ implies $L(p) < a$.

Therefore, we can conclude that $\sup\{p : L(p) < a\} = \inf\{p : a < L(p)\}$. This joint value

$$u(a) \stackrel{\text{def}}{=} \sup\{p : L(p) < a\} = \inf\{p : a < L(p)\}$$

has the following properties:

- if $p < u(a)$, then $L(p) < a$; and
- if $p > u(a)$, then $a < L(p)$.

In particular, for every small $\varepsilon > 0$, we have $L(u(a) - \varepsilon) < a < L(u(a) + \varepsilon)$. In other words, modulo arbitrary small changes in probabilities, the alternative a is equivalent to the lottery $L(p)$ in which a_1 is selected with the probability $p = u(a)$:

$$a \equiv L(u(a)).$$

This probability $u(a)$ is what is known as *utility*.

Once we know all the utility values, we can decide which alternative the decision maker will choose: the one with the largest utility. Indeed, as we have mentioned, $p < p'$ implies that $L(p) < L(p')$, so when $u(a) < u(b)$, we have

$$a \equiv L(u(a)) < L(u(b)) \equiv b$$

and thus, $a < b$.

The above definition of utility depends on the choice of two alternatives a_0 and a_1 . If we select two different benchmarks a'_0 and a'_1 , then, as one can show, the new values

of the utility are linearly related to the previous ones: $u'(a) = k \cdot u(a) + \ell$, for some real numbers $k > 0$ and ℓ . Thus, *utility is defined modulo linear transformation*.

Actual choices are often probabilistic. In practice, people sometimes make different choices when repeatedly presented with the same pair of alternatives a and b . This is especially true when the compared alternatives a and b are close in value. In such situations, we cannot predict which of the alternatives will be chosen.

The best we can do is try to predict the frequency (probability) $P(a, b)$ with which the decision maker will select a over b . More generally, we would like to predict the probability $P(a, A)$ of selecting an alternative a from a given set of alternatives A that contains a .

In the probabilistic situation, we can still talk about utilities. In the probabilistic case, we can still have a deterministic distinction:

- for some pairs (a, b) , the decision maker selects a more frequently than b : $P(a, b) > 0.5$; in such situations, we can say that a is preferable to b ($b < a$);
- for some other pairs (a, b) , the decision maker selects b more frequently than a : $P(a, b) < 0.5$; in such situations, we can say that b is preferable to a ($a < b$);
- finally, for some pairs (a, b) , the decision maker selects a exactly as many times as b ($a \sim b$): $P(a, b) = 0.5$; in such situations, we can say that to this decision maker, a and b are equivalent.

Usually, the corresponding preference relations are transitive. For example, if $a < b$ and $b < c$, i.e., if in most situations, the decision maker selects b rather than a and c rather than b , then we should expect $a < c$, i.e., we should expect that in most cases, the decision maker will prefer c to a .

Because of this, we can still perform the comparison with lotteries, and thus, come up with the utility $u(a)$ of each alternative—just like we did in the deterministic case. The main difference from the deterministic case is that:

- in the deterministic case, once we know all the utilities, we can uniquely predict which decision the decision maker will make;
- in contrast, in the probabilistic case, after we know the utility values $u(a)$ and $u(b)$, we can predict which of the two alternatives will be selected more frequently, but we still need to find out the probability $P(a, b)$.

Natural assumption. Since the alternatives can be described by their utility values, it is reasonable to assume that the desired probability $P(a, A)$ of selecting an alternative a from the set $A = \{a, \dots, b\}$ of alternatives depends only on the utilities $u(a), \dots, u(b)$ of these alternatives.

McFadden's formulas for probabilistic selection. The 2001 Nobelist D. McFadden proposed the following formula for the desired probability $P(a, A)$:

$$P(a, A) = \frac{\exp(\beta \cdot u(a))}{\sum_{b \in A} \exp(\beta \cdot u(b))};$$

see, e.g., [6, 7, 10]. In many practical situations, this formula indeed describes people's choices really well.

Need to go beyond McFadden's formulas. While McFadden's formula works in many practical situations, in some case, alternative formulas provide a better explanation of the empirical choices; see, e.g., [3] and references therein.

In this paper, we use natural symmetries to come up with an appropriate generalization of McFadden's formulas.

2 Analysis of the Problem

A usual important assumption and its consequences (see, e.g., [4]). In principle, we may have many different alternatives a, b, \dots . In some cases, we prefer a , in other cases, we prefer b . It is reasonable to require that once we have decided on selecting either a or b , then the relative frequency of selecting a should be the same as when we simply select between a and b , with no other alternatives present:

$$\frac{P(a, A)}{P(b, A)} = \frac{P(a, b)}{P(b, a)} = \frac{P(a, b)}{1 - P(a, b)}.$$

Once we make this assumption, we can then describe the general probabilities $P(a, A)$ in terms of function of one variable. Indeed, let us add a new alternative a_n to our list of alternatives. Then, because of our assumption, we have:

$$\frac{P(a, A)}{P(a_n, A)} = \frac{P(a, a_n)}{1 - P(a, a_n)},$$

hence

$$P(a, A) = P(a_n, A) \cdot f(a),$$

where we denoted $f(a) \stackrel{\text{def}}{=} \frac{P(a, a_n)}{1 - P(a, a_n)}$. In other words, for every alternative a ,

we have $P(a, A) = c \cdot f(a)$, where $c \stackrel{\text{def}}{=} P(a_n, A)$ does not depend on a . This constant c can then be found from the condition that one of the alternatives $b \in A$ will be selected, i.e., that $\sum_{b \in A} P(b, A) = 1$. Substituting $p(b, A) = c \cdot f(b)$ into this formula, we conclude that $c \cdot \sum_{b \in A} f(b) = 1$, hence

$$c = \frac{1}{\sum_{b \in A} f(b)},$$

and thus,

$$P(a, A) = \frac{f(a)}{\sum_{b \in A} f(b)}.$$

So, the probabilities are uniquely determined by some values $f(a)$ corresponding to different alternatives. Since we assumed that the probabilities depend only on the utilities $u(a)$, we thus conclude that $f(a)$ must depend only on the utilities, i.e., that we have $f(a) = F(u(a))$ for some function $F(u)$. In terms of this function, the above formula for the probabilistic choice takes the form

$$P(a, A) = \frac{F(u(a))}{\sum_{b \in A} F(u(b))}. \tag{1}$$

From this viewpoint, all we need to do is to find an appropriate function $F(u)$.

The function $F(u)$ must be monotonic. The better the alternative a , i.e., the larger its utility $u(a)$, the higher should be the probability that we select this alternative. Thus, it is reasonable to require that the function $F(u)$ is an increasing function of the utility u .

The function $F(u)$ is defined modulo a constant factor. The above formula does not uniquely define the function $F(u)$: indeed, if we multiply all the values of $F(u)$ by a constant, i.e., consider the new function $F'(u) = C \cdot F(u)$, then in the Formula (1), constants C in the numerator and the denominator will cancel each other, and thus, we will get the exact same probabilities.

Vice versa, if two functions $F(u)$ and $F'(u)$ always lead to the same probabilities, this means, in particular, that for every two utility values u_1 and u_2 , we have

$$\frac{F(u_2)}{F(u_1) + F(u_2)} = \frac{F'(u_2)}{F'(u_1) + F'(u_2)}.$$

Reversing both sides of this equality and subtracting 1 from both sides, we conclude that

$$\frac{F(u_1)}{F(u_2)} = \frac{F'(u_1)}{F'(u_2)},$$

i.e., equivalently, that

$$\frac{F'(u_1)}{F(u_1)} = \frac{F'(u_2)}{F(u_2)}.$$

In other words, the ratio $\frac{F'(u)}{F(u)}$ is the same for all utility values u , and is, therefore, a constant C . Thus, in this case, $F'(u) = C \cdot F(u)$.

So, two functions $F(u)$ and $F'(u)$ always lead to the same probabilities if and only if their differ by a constant factor.

In these terms, how can we explain the original McFadden's formulas. As we have mentioned earlier, utilities are defined modulo a general linear transformation. In particular, it is possible to add a constant to all the utility values

$$u(a) \rightarrow u'(a) = u(a) + c$$

and still get the description of exactly the same preferences. Since this shift does not change the preferences, it is therefore reasonable to require that after such a shift, we get the exact same probabilities.

Using new utility values $u'(a) = u(a) + c$ means that we replace the values $F(u(a))$ with the values $F(u'(a)) = F(u(a) + c)$. This is equivalent to using the *original* utility values but with a *new function* $F'(u) \stackrel{\text{def}}{=} F(u + c)$.

As we have mentioned earlier, the requirement that the two functions $F(u)$ and $F'(u)$ describe the same probabilities is equivalent to requiring that $F'(u) = C \cdot F(u)$ for some constant C , so $F(u + c) = C \cdot F(u)$. The factor C is, in general, different for different shifts c : $C = C(c)$. Thus, we conclude that

$$F(u + c) = C(c) \cdot F(u). \quad (2)$$

It is known (see, e.g., [1]) that every monotonic solution to this function equation has the form $F(u) = C_0 \cdot \exp(\beta \cdot u)$. This is exactly McFadden's formula.

Discussion. For a general monotonic function, the proof of the function-equation result may be somewhat complicated. However, under a natural assumption that the function $F(u)$ is differentiable, this result can be proven rather easily.

First, we take into account that $C(c)$ is a ratio of two differentiable functions $F(u + c)$ and $F(u)$, and is, thus, differentiable itself. Since both functions $F(u)$ and $C(c)$ are differentiable, we can differentiate both sides of the equality (2) by c and take $c = 0$. As a result, we get the following equality:

$$\frac{dF}{du} = \beta \cdot F,$$

where we denoted $\beta \stackrel{\text{def}}{=} \frac{dC}{dc} \Big|_{c=0}$. By moving all the terms containing the unknown F to one side and all other terms to the other side, we conclude that:

$$\frac{dF}{F} = \beta \cdot du.$$

Integrating both sides of this equality, we get $\ln(F) = \beta \cdot u + C_1$, where C_1 is the integration constant. Thus, $F(u) = \exp(\ln(F)) = C_0 \cdot \exp(\beta \cdot u)$, where we denoted $C_0 \stackrel{\text{def}}{=} \exp(C_1)$.

Our main idea and the resulting formulas. Please note that while adding a constant to all the utility values does not change the probabilities computed by using

McFadden's formula, multiplying all the utility values by a constant—which is also a legitimate transformation for utilities—*does* change the probabilities.

Therefore, we cannot require that the probability Formula (1) not change for all possible linear transformations of utility: once we require shift-invariance, we get McFadden's formula which is not scale-invariant.

Since we cannot require invariance with respect to *all* possible re-scalings of utility, we should require invariance with respect to *some* family of re-scalings.

If a formula does not change when we apply each transformation, it will also not change if we apply them one after another, i.e., if we consider a composition of transformations. Each shift can be represented as a superposition of many small (infinitesimal) shifts, i.e., shifts of the type $u \rightarrow u + B \cdot dt$ for some B . Similarly, each scaling can be represented as a superposition of many small (infinitesimal) scalings, i.e., scalings of the type $u \rightarrow (1 + A \cdot dt) \cdot u$. Thus, it is sufficient to consider invariance with respect to an infinitesimal transformation, i.e., a linear transformation of the type

$$u \rightarrow u' = (1 + A \cdot dt) \cdot u + B \cdot dt.$$

Invariance means that the values $F(u')$ lead to the same probabilities as the original values $F(u)$, i.e., that $F(u')$ is obtained from $F(u)$ by an appropriate (infinitesimal) re-scaling $F(u) \rightarrow (1 + C \cdot dt) \cdot F(u)$. In other words, we require that

$$F((1 + A \cdot dt) \cdot u + B \cdot dt) = (1 + C \cdot dt) \cdot F(u),$$

i.e., that

$$F(u + (A \cdot u + B) \cdot dt) = F(u) + C \cdot F(u) \cdot dt. \quad (3)$$

Here, by definition of the derivative, $F(u + q \cdot dt) = F(u) + \frac{dF}{du} \cdot q \cdot dt$. Thus, from (3), we conclude that

$$F(u) + (A \cdot u + B) \cdot \frac{dF}{du} \cdot dt = F(u) + C \cdot F(u) \cdot dt.$$

Subtracting $F(u)$ from both sides and dividing the resulting equality by dt , we conclude that

$$(A \cdot u + B) \cdot \frac{dF}{du} = C \cdot F(u).$$

We can separate the variables by moving all the terms related to F to one side and all the terms related to u to another side. As a result, we get

$$\frac{dF}{F} = C \cdot \frac{du}{A \cdot u + B}.$$

We have already shown that the case $A = 0$ leads to McFadden's formulas. So, to get a full description of all possible probabilistic choice formulas, we need to

consider the cases when $A \neq 0$. In these cases, for $x \stackrel{\text{def}}{=} u + k$, where $k \stackrel{\text{def}}{=} \frac{B}{A}$, we have

$$\frac{dF}{F} = c \cdot \frac{dx}{x},$$

where $c \stackrel{\text{def}}{=} \frac{C}{A}$. Integration leads to $\ln(F) = c \cdot \ln(x) + C_0$ for some constant C_0 , thus $F = C_1 \cdot x^c$ for $C_1 \stackrel{\text{def}}{=} \exp(C_0)$, i.e., to

$$F(u) = C_1 \cdot (u + k)^c. \tag{4}$$

Conclusions and discussion. In addition to the original McFadden’s formula for the probabilistic choice, we can also have the case when $F(u)$ is described by the formula (4) and where, therefore, the probabilistic choice is described by the formula

$$P(a, A) = \frac{(u(a) + k)^c}{\sum_{b \in A} (u(b) + k)^c}.$$

This expression is in good accordance with the empirical dependencies described in [3] that also contain power-law terms.

It is worth mentioning that while we derived the new formula as an *alternative* to McFadden’s formula, this new formula can be viewed as a *generalization* of McFadden’s formula. Indeed, it is known that $\exp(u) = \lim_{n \rightarrow \infty} \left(1 + \frac{u}{n}\right)^n$ and thus, $\exp(\beta \cdot u) = \lim_{n \rightarrow \infty} \left(1 + \frac{\beta \cdot u}{n}\right)^n$. Thus, when n is large, the use of McFadden’s expression $F(u) = \exp(\beta \cdot u)$ is practically indistinguishable from the use of the power-law expression $F_{\approx}(u) = \left(1 + \frac{\beta \cdot u}{n}\right)^n$. This power-law expression, in its turn, can be represented in the form (4), with $c = n$, $k = \frac{n}{\beta}$, and $C_1 = \left(\frac{\beta}{n}\right)^n$.

So, instead of a 1-parametric McFadden’s formula, we now have a 2-parametric formula. We can use this additional parameter to get an even more accurate description of the actual probabilistic choice.

Acknowledgements This work was supported by Chiang Mai University, Thailand. This work was also supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award “UTEP and Prudential Actuarial Science Academy and Pipeline Initiative” from Prudential Foundation.

References

1. Aczél J, Dhombres J (2008) *Functional equations in several variables*. Cambridge University Press, Cambridge, UK
2. Fishburn PC (1969) *Utility theory for decision making*. Wiley Inc., New York
3. Jakubczyk M (2016) Estimating the membership function of the fuzzy willingness-to-pay/accept for health via Bayesian modeling. In: *Proceedings of the 6th world conference on soft computing*, Berkeley, California, 22–25 May 2016
4. Luce D (2005) *Individual choice behavior: a theoretical analysis*. Dover, New York
5. Luce RD, Raiffa R (1989) *Games and decisions: introduction and critical survey*. Dover, New York
6. McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) *Frontiers in econometrics*. Academic Press, New York, pp 105–142
7. McFadden D (2001) Economic choices. *Am Econ Rev* 91:351–378
8. Nguyen HT, Kosheleva O, Kreinovich V (2009) Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction. *Int J Intell Syst* 24(1):27–47
9. Raiffa H (1970) *Decision analysis*. Addison-Wesley, Reading
10. Train K (2003) *Discrete choice methods with simulation*. Cambridge University Press, Cambridge