entropy

Article

A Microeconomic Interpretation of the Maximum Entropy Estimator of Multinomial Logit Models and Its Equivalence to the Maximum Likelihood Estimator

Pedro Donoso [1] and Louis de Grange [2],*

[1] Laboratorio de Modelamiento del Transporte y Uso del Suelo (LABTUS), Departamento de Ingeniería Civil, Universidad de Chile, Santiago, Chile; E-Mail: pdonoso@labtus.cl

[2] Escuela de Ingeniería Civil Industrial, Universidad Diego Portales, Santiago, Chile

* Author to whom correspondence should be addressed; E-Mail: louis.degrange@udp.cl; Tel.: +56-2-6768118; Fax: +56-2-6768130.

Abstract: Maximum entropy models are often used to describe supply and demand behavior in urban transportation and land use systems. However, they have been criticized for not representing behavioral rules of system agents and because their parameters seems to adjust only to modeler-imposed constraints. In response, it is demonstrated that the solution to the entropy maximization problem with linear constraints is a multinomial logit model whose parameters solve the likelihood maximization problem of this probabilistic model. But this result neither provides a microeconomic interpretation of the entropy maximization problem nor explains the equivalence of these two optimization problems. This work demonstrates that an analysis of the dual of the entropy maximization problem yields two useful alternative explanations of its solution. The first shows that the maximum entropy estimators of the multinomial logit model parameters reproduce rational user behavior, while the second shows that the likelihood maximization problem for multinomial logit models is the dual of the entropy maximization problem.

1. Introduction

The maximum entropy approach is widely used for formulating demand models, primarily aggregate, for urban transportation and land use systems. The first application to transportation demand was Wilson's aggregate doubly constrained gravity model of spatial trip distribution [1]. Since then, various writers have published similar versions with various improvements (see [2-4]).

A relevant class of travel demand models are the combined maximum entropy ones. These are aggregate designs that integrate multiple transportation user decisions such as trip generation, destination choice, and mode and/or route choice. Examples may be found in [5-17]. In every case the modeling method involves formulating an optimization problem with entropy components whose optimality conditions define the required demand models.

One of the main reasons for employing maximum entropy models is that their probabilistic formulation exhibits a multinomial logit structure which has an obvious microeconomic interpretation. The discrete choice multinomial logit model has been widely utilized for modeling urban transportation and land use systems due to its ability to represent, by means of a closed formula, the paradigm of the rational consumer (or producer) who maximizes utility (or benefit) within his or her possibility space. Both the model's simplicity and its limitations stem from the fact that the utility functions' stochastic errors are Gumbel-distributed i.i.d. random variables (see [18]). The most commonly used method for estimating the parameters of the model is maximum likelihood due to its good asymptotic statistical properties.

It was demonstrated in [5] that for the multinomial logit model, the entropy maximization problem with linear constraints has the same solution as the likelihood maximization problem because the Kuhn-Tucker conditions of the two are identical. As a consequence of this result, the Lagrange multipliers of the entropy maximization problem are the maximum likelihood estimators of the model. It did not, however, provide either an interpretation of entropy maximization in a microeconomic modeling context or any explanation of the origin of the above-mentioned equivalence. The intended contribution of the present paper is to fill these two gaps.

In what follows, Part 2 formulates the entropy maximization problem of a general discrete choice probability distribution with linear constraints, identifies its solution and specifies the dual problem. Part 3 gives a microeconomic interpretation of the dual problem based on the rational consumer behavior paradigm, while Part 4 provides a statistical interpretation of the dual in terms of the estimation criteria of the multinomial logit model parameters. Finally, Part 5 sets out our conclusions.

2. Formulation of Entropy Maximization Problem and its Dual

Consider the following entropy maximization problem (ME) of a discrete choice probability distribution with linear constraints:

$$\textbf{ME}: \quad \underset{p}{Max} \; E(p) \;=\; \underset{p}{Max} \; -\sum_{i,a} N_i \, p_{a/i} \, ln \, p_{a/i} \tag{1}$$

$$\overset{s.t.}{\underset{i,a}{\sum}} N_i \, p_{a/i} \, x_{aik} = \sum_{i,a} N_{ai} \, x_{aik} \qquad \text{for all } k \tag{2}$$

$$\sum_a p_{a/i} \;=\; 1 \quad \text{for all } i \tag{3}$$

where variable $p_{a/i}$ is the probability that an individual of type i chooses alternative a, the quantities (x_{aik}) are the attributes of alternative a perceived by individual i, N_{ai} is the observed number of type i individuals who chose alternative a and N_i is the number of type i individuals. The problem is disaggregate in individuals if $N_i = 1$ for all i; if not, we get an aggregated formulation. It is assumed that each individual chooses a single alternative, as expressed in constraint (3). Note also that:

$$N_i = \sum_a N_{ai} \qquad \text{for all } i \tag{4}$$

Now let (β_k) be the Lagrange multipliers of constraint (2). Applying the Kuhn-Tucker conditions to ME we obtain as the solution to the problem the following multinomial logit model:

$$p_{a/i}(\beta) = \frac{exp\left(\sum_k \beta_k x_{aik}\right)}{\sum_{a'} exp\left(\sum_k \beta_k x_{a'ik}\right)} \qquad \text{for all } a,I \tag{5}$$

This same model is obtained if we apply the random utility theory approach (see [18,19]), specifying that:

$$\lambda V_{ai} = \sum_k \beta_k x_{aik} \tag{6}$$

where V_{ai} is the deterministic component of the conditional indirect utility perceived by individual i from alternative a and λ is the scale factor for a Gumbel probability distribution of the random utility error $(\lambda > 0)$. As demonstrated by [5], the Lagrangian multipliers (β_k) are the maximum likelihood estimators of the model (5). The Lagrangian function of the ME problem is:

$$L = -\sum_{i,a} N_i p_{a/i} \ln p_{a/i} + \sum_k \beta_k \left(\sum_{i,a} N_i p_{a/i} x_{aik} - \sum_{i,a} N_{ai} x_{aik}\right) + \sum_i \theta_i \left(\sum_a p_{a/i} - 1\right) \tag{7}$$

where θ_i is the Lagrange multiplier of constraint (3). The equivalent problem of ME is therefore:

ME:

$$\underset{p,\beta,\theta}{Max}\, L(p,\beta,\theta) = \underset{p,\beta,\theta}{Max}\left\{-\sum_{i,a} N_i p_{a/i} \ln p_{a/i} + \sum_k \beta_k \left(\sum_{i,a} N_i p_{a/i} x_{aik} - \sum_{i,a} N_{ai} x_{aik}\right) + \sum_i \theta_i \left(\sum_a p_{a/i} - 1\right)\right\} \tag{8}$$

If the value of $\ln p_{a/i}$ obtained from (5) is substituted into the first summation term of the Lagrangian (7) and the value of $p_{a/i}$ is substituted into the third summation term, the resulting expression reduces to:

$$L = \sum_i N_i \ln\left(\sum_{a'} exp\left(\sum_k \beta_k x_{a'ik}\right)\right) - \sum_{i,a} N_{ai} \sum_k \beta_k x_{aik} \tag{9}$$

From this expression we can formulate the known dual of the ME problem (see [20]) as:

$$\textbf{DME: } \underset{\beta}{Min}\; D(\beta) = \underset{\beta}{Min}\left\{\sum_i N_i \ln\left(\sum_{a'} exp\left(\sum_k \beta_k x_{a'ik}\right)\right) - \sum_{i,a} N_{ai} \sum_k \beta_k x_{aik}\right\} \tag{10}$$

In the following two sections we present two interpretations of the DME problem, one microeconomic and the other statistical. By extension, they are applied to the ME problem.

3. Microeconomic Interpretation of the Entropy Maximization Dual Problem

If (6) is substituted into (10) then, given (4), the dual problem can be written as:

$$\textbf{DME}: \underset{\beta}{Min} \ \frac{1}{\lambda} D(\beta) = \underset{\beta}{Min} \sum_{i,a} N_{ai}\left(EMU_i(\beta) - V_{ai}(\beta)\right) \tag{11}$$

where:

$$EMU_i(\beta) = \frac{1}{\lambda} ln\left(\sum_{a'} exp\left(\lambda V_{a'i}(\beta)\right)\right) \qquad \text{for all } i \tag{12}$$

is the maximum expected utility of individual i [21].

Given that $EMU_i(\beta) \geq V_{ai}(\beta)$ for all a,i, the DME problem consists in finding the parameters (β_k) such that for each individual the expected utility for the chosen alternative approaches as closely as possible to the maximum expected utility of the various available alternatives. This allows us to interpret the Lagrange multipliers of restrictions (2) micro economically in that they adjust to the fullest extent possible so that the model reflects rational behavior by the individuals.

An interesting variant of the ME problem occurs when the parameters (β_k) are known. If this condition is applied to (8) then, given (6), this variant is:

$$\textbf{VME}: \underset{p,\theta}{Max} \ \tilde{L}(p,\theta) = \underset{p,\theta}{Max}\left\{-\sum_{i,a} N_i p_{a/i} \ln p_{a/i} + \lambda \sum_{i,a} N_i p_{a/i} V_{ai} + \sum_i \theta_i\left(\sum_a p_{a/i} -1\right)\right\} \tag{13}$$

or equivalently:

$$\textbf{VME}: \underset{p}{Max} \ M(p) = \underset{p}{Max}\left\{\sum_{i,a} N_i p_{a/i} V_{ai} - \frac{1}{\lambda}\sum_{i,a} N_i p_{a/i} \ln p_{a/i}\right\} \tag{14}$$

$$\overset{s.t.}{\underset{a}{\sum}} p_{a/i} = 1 \quad \text{for all } i \tag{15}$$

The entropy term of this problem can be interpreted simply as the penalty imposed on the deterministic problem for finding a discrete choice probability law that maximizes the sum, over all individuals, of the expected value of the deterministic utilities of the chosen alternatives.

There is, however, another interesting microeconomic interpretation. Since the solution of the VME problem is (5), we deduce that the maximum expected utility of individual i, defined by (12), can be written as:

$$EMU_i = \frac{1}{\lambda} ln\left(\frac{exp(\lambda V_{ai})}{p_{a/i}}\right) = V_{ai} - \frac{1}{\lambda} \ln p_{a/i} \qquad \text{for all } a,i \tag{16}$$

In (16) the function is linear in the logarithm of the probability. It can be proved that the functions that are linear the logarithm are the only ones that gives what is called proper local score functions.

If we multiply (16) by $p_{a/i}$ and sum over a, then, given (15), we have:

$$EMU_i = \sum_a p_{a/i} V_{ai} - \frac{1}{\lambda} \sum_a p_{a/i} \ln p_{a/i} \qquad \text{for all } i \qquad (17)$$

Thus, the value of the objective function of the VME problem at the optimum is the sum, over all individuals, of their maximum expected utilities among the available alternatives. Observe also that $M(p) \leq \sum_i N_i EMU_i$ for all p. The discrete choice probability law that solves the VME problem therefore conforms with the rational consumer paradigm for each individual as regards expected value.

4. Statistical Interpretation of the Entropy Maximization Dual Problem

The statistical interpretation of the DME problem (10) is obtained from the following reformulation of its objective function:

$$
\begin{aligned}
D(\beta) &= -\left(\sum_{i,a} N_{ai} \sum_k \beta_k x_{aik} - \sum_i N_i \ln\left(\sum_{a'} exp\left(\sum_k \beta_k x_{a'ik} \right) \right) \right) \\
&= -\left(\sum_{i,a} N_{ai} \ln\left(exp\left(\sum_k \beta_k x_{aik} \right) \right) - \sum_{i,a} N_{ai} \ln\left(\sum_{a'} exp\left(\sum_k \beta_k x_{a'ik} \right) \right) \right) \qquad (18) \\
&= -\sum_{i,a} N_{ai} \ln\left(\frac{exp\left(\sum_k \beta_k x_{aik} \right)}{\sum_{a'} exp\left(\sum_k \beta_k x_{a'ik} \right)} \right)
\end{aligned}
$$

Thus, the negative of the DME problem objective function is just the log likelihood function of the multinomial logit model. It follows, then, that the log likelihood maximization problem is equivalent to the entropy maximization problem because its dual is just the DME problem.

Then, the DME problem can therefore be reformulated as:

$$\textbf{DME}: \underset{\beta}{Max} \quad -D(\beta) = \underset{\beta}{Max} \sum_{i,a} N_{ai} \ln p_{a/i}(\beta) \qquad (19)$$

Note also that this equivalence is independent of the number of individuals or alternatives or whether the probabilistic model is aggregate in individuals or not.

The variance-covariance matrix of the parameters (β_k) that solve the ME problem is derived as the inverse of the expected value of the matrix of second derivatives of the dual function D of the DME problem, known as the information matrix (see [22-28]).

The second derivatives of the dual problem measure the variations in the optimum of the ME or DME objective function caused by changes in the parameters of the resulting model. Thus, if a parameter undergoes a large change but the dual problem optimum varies very little (that is, the second derivative of the dual problem is small), that parameter provides little information and intuitively its variance should be very high. If, on the other hand, the second derivative of the dual with respect to a given parameter is high, the parameter is significant and its variance should be small. Notice that this point is closely related to the fact that the Fisher information matrix in an exponential family is inversely proportional to the variance-covariance matrix.

By a similar process we can estimate the variance-covariance matrix of the parameters of the multinomial logit model using the log likelihood function. This implies that by demonstrating that this

function is the negative of the dual function we will have reconciled the two approaches to estimating the variance-covariance matrix of the model parameters.

It also follows from the equivalence of the DME problem (10) and problem (19) that the latter is equivalent to problem (11). From this we deduce that the maximum likelihood estimators of the multinomial logit model represent rational behavior by the users in terms of expected value. This method of estimating the multinomial logit model parameters is therefore congruent with the rational consumer paradigm defined by the model itself.

5. Conclusions

Maximum entropy models have been widely applied to various economic systems, and are particularly common in representations of supply and demand behavior in urban transportation and land use modeling. A major criticism of these models, however, is that they do not embody actual behavioral rules of the agents in the system and their parameters adjust only to the constraints imposed by the modeler. This paper demonstrated that two useful alternative interpretations of the models' optimality conditions can be derived from an analysis of the dual of the classical entropy maximization problem.

The first of these interpretations was derived from the fact that the dual problem consists in finding parameters of the utility (benefit) functions of users (suppliers) such that each user behaves rationally, that is, maximizes utility (benefit) in terms of the expected value of the consumption (production) alternatives available. The maximum entropy model parameters thus take on a clear microeconomic interpretation.

The second interpretation is that from a statistical standpoint: the dual maximum entropy problem is equivalent to the likelihood maximization problem for multinomial logit models. This in turn explains the equivalence between the problems of likelihood maximization and entropy maximization. Furthermore, with this result the methods for estimating the variance-covariance matrices of maximum entropy and maximum likelihood estimators are completely reconciled. Finally, we note that both interpretations are valid for aggregate or disaggregate models independently of the number of individuals or alternatives considered.

A possible extension of this study would be to analyze logit models with non-linear constraints and hierarchical structures. Given the characteristics that distinguish them from multinomial logit models, the results may be different from those presented here.

References and Notes

1. Wilson, A.G. *Entropy in Urban and Regional Modeling*; Pion: London, UK, 1970.
2. Fang, S.; Tsao, J. Linearly-constrained entropy maximization problem with quadratic cost and its applications to transportation planning problems. *Transp. Sci.* **1995**, *29*, 353-365.
3. Thorsen, I.; Gitlesen, J.P. Empirical evaluation of alternative model specifications to predict commuting flows. *J. Reg. Sci.* **1998**, 38, 273-292.
4. De Grange, L.; Fernandez, J.E.; De Cea, J. A consolidated model of trip distribution. *Transp. Res.* **2010**, *46,* 61-75.

5. Anas, A. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transp. Res.* **1983**, *17*, 13-23.

6. Boyce, D.; LeBlanc L.; Chon, K.; Lee, Y; Lin, K. Implementation and computational issues for combined models of location, destination, mode and route choice. *Environ. Plan.* **1983**, *15*, 1219-1230

7. Fotheringham, A. Modeling hierarchical destination choice. *Environ. Plan.* **1986**, *18*, 401-418.

8. Boyce, D.; LeBlanc, L.; Chon, K. Network equilibrium models of urban location and travel choices: A retrospective survey. *J. Reg. Sci.* **1988**, *28*, 159-183.

9. Safwat, K.; Magnanti, T. A combined trip generation, trip distribution, modal split and traffic assignment model. *Transp. Sci.* **1988**, *22*, 14-30.

10. Brice, S. Derivation of nested transport models within a mathematical programming framework. *Transp. Res.* **1989**, *23*, 19-28.

11. Fernandez, J.E.; De Cea, J.; Florian, M.; Cabrera, E. Network equilibrium models with combined modes. *Transp. Sci.* **1994**, *28*, 182-192.

12. Oppenheim, N. *Urban Travel Demand Modeling*; John Wiley & Sons: New York, NY, USA, 1995.

13. Abrahamsson, T.; Lundqvist, L. Formulation and estimation of combined network equilibrium models with applications to stockholm. *Transp. Sci.* **1999**, *33*, 80-100.

14. Boyce, D.; Bar-Gera, H. Validation of multiclass urban travel forecasting models combining origin-destination, mode, and route choices. *J. Reg. Sci.* **2003**, *43*, 517-540.

15. Ham, H.; Tschangho, J.; Boyce, D. Implementation and estimation of a combined model of interregional, multimodal commodity shipments and transportation network flows. *Transp. Res.* **2005**, *39*, 65-79.

16. Garcia, R.; Marin, A. Network equilibrium with combined modes: models and solution algorithms. *Transp. Res.* **2005**, *39*, 223-254.

17. De Cea, J.; Fernandez, J.E.; De Grange, L. Combined models with hierarchical demand choices: A multi-objective entropy optimization approach. *Transp. Rev.* **2008**, *28*, 415-438.

18. McFadden, D. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, Zarembka P., Ed.; Academic Press: New York, NY, USA, 1974.

19. Ortuzar, J. de D.; Willumsen, L.G. *Modeling Transport*; John Wiley & Sons: Chichester, UK, 2001.

20. Fang, S.; Rajasekera, J.; Tsao, J. *Entropy Optimization and Mathematical Programming*; Kluwer Academic Publisher: Norwell, MA, USA, 1997.

21. Williams, H.C.W.L. On the formation of travel demand models and economic evaluation measures of user benefit. *Environ. Plan.* **1977**, *9*, 285-344.

22. Imbens, G.W.; Johnson, P.; Spady, R.H. Information theoretic approaches to inference in moment condition models. *Econometrica* **1998**, *66*, 333-357.

23. Golan, A. Information and entropy econometrics — Editor's view. *J. Econom.* **2002**, *107*, 1-15.

24. Golan, A. *Information and Entropy Econometrics: A Review and Synthesis*; Now Publishers Inc.: Hanover; MA, USA, 2008.

25. Bercher, J.F.; Besnerais, G.L.; Demoment, G. The maximum entropy on the mean method, noise and sensitivity. In *Maximum Entropy and Bayesian Studies*, Skilling, J., Sibisi, S., Eds.; Kluwer Academic Publishers: Cambridge, UK, 1996.

26. Csiszar, I.; Shields, P. Information theory and statistics: A tutorial. *Found. Tr. Commun. Inform. Theory* **2004**, *1*, 417-528.

27. Soofi, E.S.; Retzer, J.J. Information indices: Unifications and applications. *J. Econom.* **2002**, *107*, 17-40.

28. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 2006.