Statistical Models and Applications

统计模型和应用研究

宋连莲　刘　帆／著

科　学　出　版　社

Statistical Models and Applications

统计模型和应用研究

宋连莲　刘　帆　著

科　学　出　版　社

北　京

内 容 简 介

本书针对常用的 probit 模型、MNL 模型、survival 等离散选择模型和贝叶斯、多选择等动态模型进行了系统的描述和分析，并在此基础上，构建出新型的拓展模型，解决新的实际问题。本书融合统计学、经济学和行为学三门学科的基础知识，论述统计模型在消费者购买研究、客户关系管理、电视节目观看行为研究、广告收视率预测研究、网页优化研究、客户信用评估等领域的应用价值。

本书适用于统计学、经济学和管理科学等学科教学科研人员及学生，也适合大量参与创新、研发活动的企业决策者、实施者及咨询机构人员等阅读。本书的诸多实证研究与结论，以及所提出的拓展研究模型，可以在企业高层管理者面对复杂多变的市场需求时，提出较为精准的预测方案，帮助其理解和分析并制定相应的策略。

图书在版编目（CIP）数据

统计模型和应用研究 = Statistical Models and Applications/宋连莲，刘帆著.
—北京：科学出版社，2017.8

ISBN 978-7-03-052328-0

Ⅰ. ①统… Ⅱ. ①宋… ②刘… Ⅲ. ①统计模型－研究－英文
Ⅳ. ①C815

中国版本图书馆 CIP 数据核字（2017）第 052824 号

责任编辑：方小丽 李 莉/责任校对：彭珍珍
责任印制：吴兆东/封面设计：无极书装

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
http：//www.sciencep.com

北京京华虎彩印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2017 年 8 月第 一 版 开本：B5（720×1000）
2017 年 8 月第一次印刷 印张：17
字数：340 000

定价：102.00 元

（如有印装质量问题，我社负责调换）

This book is financially supported by NSFC (No. 71602089), Natural and Social Science Foundation of Jiangsu Province (No. BK20160785, 16GLC001).

Contents

1 General Statistical Models

1.1 Motivation

In the mid-1980s, Train wrote the book *Qualitative Choice Analysis*. In this book, he introduced the basic models—logit and nested logit models in detail, including the model properties and applications in many areas. However, the first-generation choice models suffered many limitations which inhabited them in resolving real problems. More than 20 years have passed, in 2003, Train wrote the second book, *Discrete Choice Models with Simulation*, for stating the tremendous progress in the approaches and methods of choice analysis. Simulation had been the center of the progress, since it supplied a numerical approximation to integrals. Till today, more than 10 years have passed since 2003, the famous discrete choice models have been used in a wide range of our life, including marketing, housing, transportation and energy. However, a new generation of problems has emerged, which brings forth a new burden on the researchers.

The problems or situations that need modeling and explaining have become more complex and full of high technology. For example, marketing problems are not restricted to the physical shops, the purchase behavior has transferred onto the Internet. Original discrete choice models may not satisfy the requirements of the new issues. We summarize the new generation of problems as two categories. One is that the modeling methods themselves have become more complicated than the traditional discrete choice models. To fit the real situations, new modeling methods are developed when none of the existing models seem to be correct. People's behavior is changeable in a short period of time. Sometimes, we combine two different models into a new one to fit the disparate purchasing behavior and apply a new theory to a particular problem even though that theory has never been used in similar situations. The other is for the estimation method to implement the new models. When developing new models and employing a new theory, the relevant validations are necessary. Researchers need to be able to develop very complex probabilities and the corresponding likelihood function, and program the procedures into computer softwares. That is a very difficult process, since researchers often need to change their traditional computational perspectives, capture the up-to-date technics of coding and finally achieve the aims of parameter estimation and model managerial implications.

The purpose of this book is to extend the model construction and application of the traditional discrete choice models using the empirical examples. From the improved model development to the innovative parameter estimation procedure, we illustrate all steps in very detail. In particular, the major contribution of this book is to deal with the new problems emerged in today's complex situation through proposing innovative models. Not only do the proposed models have advantages in describing and explaining the choice behavior, but they also have advantages in conducting new managerial implications for managers.

Specifically, we follow the traditional steps of applying statistical models in marketing. Firstly, we discuss the customer-base analysis and answer such questions as which customers are more likely to stop purchasing from the firm? What level of transactions should be expected from the customer base? How do the customers respond to the marketing activities? Secondly, we extend the traditional methods of applying classification models into the area of credit scoring: 1) relaxing the assumption of the independence between the probability and the time of default; 2) treating the missing defaulting labels as latent variables and applying an augmentation technique; 3) introducing a discrete truncated exponential distribution to model the time of default. An empirical analysis is also conducted to show the performance of our proposed model. Thirdly, we extend the model application in the area of television audience viewing behavior, including television program and advertisement (ad) choice behavior: 1) modeling the dynamic multiple choice behavior of television audience to predict and improve program audience ratings; 2) modeling commercial viewing patterns to convert minute information into second information to evaluate advertisement effectiveness. We use three chapters to discuss the topics suggested in those two areas. The common aim of those two topics is to construct statistical models that examine television audience viewing behavior in order to increase audience ratings of target programs or advertisements (ads). The differences between them can be described from two aspects. On the one hand, topic 1 targets audience viewing behavior to programs. Whereas, topic 2 develops a model for the viewing behavior towards television advertisements. On the other hand, topic 2 develops a methodology to measure audience ratings. Whereas, topic 1 predicts audience ratings using a more sophisticated model. Fourthly, the fast development of web technologies has revolutionized the way we spend our time. The number of web users is increasing every year. The statistical models are absolutely being used in modeling the web browsing behavior. Those topics are described by two chapters in this book. Fifthly, the entrepreneurship area is also discussed in this book by Chapter 10. We develop a new model which combines the statistical theory into the stochastic multicriteria acceptability analysis to build an improved global entrepreneurship index (GEI), which we term it the holistic acceptability index. That method differs from the conventional wisdom that assigns exact values to corresponding weights, but explores the weight space to allow each country to specify a preference. Sixthly, our method is confirmed using an empirical study measuring the entrepreneurship of the top 10 countries in terms of 2016 GEI.

1.2 Properties of Discrete Choice Models

1.2.1 Background

Discrete choice model (DCM) is a modeling method of choice from a set of alternatives. That set of alternatives is called the choice set. The set must be mutually exclusive and collectively exhaustive and the number of alternatives is finite. The number of alternatives is too large for the respondent to consider, McFadden has suggested a sampling method to select a sample of alternative for respondents to make choices. An assumption is suggested that a utility function can be found to describe a decision maker's preference when he is confronted with a set of alternatives. Utility, as a measure of well-being, has no natural level or scale. While, neo-classical economists comment that a utility function can allow the decision maker to rank the alternatives in a consistent and unambiguous manner. Then another assumption of utility maximizing behavior by the decision maker is proposed: the choice is a deterministic process.

However, psychologists present that it is impossible to specify and construct a discrete choice model that will always succeed in predicting the chosen alternative by the decision maker. Accordingly, Thurstone (1927) proposed a utility function that allows the decision maker to rank the alternatives, but there are fluctuations inherent in the process of evaluating alternatives. As noted by Tversky (1972), "when faced with a choice among several alternatives, people often experience uncertainty and inconsistency. That is, people are often not sure which alternative they should select, nor do they always take the same choice under seemingly identical conditions." Thus, the choice is an outcome of a probabilistic process.

1.2.2 Random Utility

Two sources are for the randomness of the utility function. One is the inability of the analysts to formulate individual behavior, since they lack information of individuals' characteristics and information of alternatives' attributes. The other is the stochastic behavior of the decision maker, the circumstances under the decisions can "perturb" the perception and the desirability of alternatives. A random utility function is necessary to model the random utility.

The random utility function is constructed as this:

$$U = \text{deterministic component } (V) + \text{random component } (\varepsilon)$$

where V presents observable attributes of alternatives and characteristics of individuals; ε captures the effects of stochastic behavior of individuals and unobserved attributes and characteristics.

$$U_{ij} = V_{ij}(x_{ij1}, x_{ij2}, \cdots, x_{ijk}) + \varepsilon_{ij}$$

That is the utility function for individual i and alternative j, where x_{ijk} denotes the observed characteristics of individual i and the observed kth attributes of alternative j. Let X_{ij} be a vector of x_{ijk}, that is, $X'_{ij} = (1, x_{ij1}, x_{ij2}, \cdots, x_{ijk})$.

That X-vector might include simple attributes (e.g., income, price per trip, etc.) or explicit interactions of the attributes of the alternatives and the individuals (e.g., price, income).

If we can assume a linear function for the effects of the factors, the random utility becomes:

$$U_{ij} = X'_{ij}\beta + \varepsilon_{ij} \quad \text{where } \beta' = (\beta_0, \beta_1, \cdots, \beta_k) \tag{1.1}$$

β_0 is the intercept term representing the intrinsic preference of individual i towards alternative j; β_k is the response coefficient of the kth attribute; U_{ij} is the utility of the jth alternative to the ith individual.

Equation (1.1) is under these two assumptions: 1) the vector β reflects the response to the attribute of the alternative, which is assumed to be identical for all the individuals in the population; 2) random variation in the U_{ij} is introduced through the additive disturbance term ε_{ij} which is assumed to be independently and identically distributed across individuals and alternatives for multinomial logit (MNL), but the error terms could be dependent on each other for other models.

1.3 Multinomial Logit Model

1.3.1 General Multinomial Logit Model

The random components of the utilities of different alternatives are IID (independent and identically distributed) according to Gumbel distribution (extreme value distribution).

Derivation of multinomial logit is shown as below:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad \text{where } \varepsilon_{ij} \sim \text{Gumbel}(\eta,\ \mu)$$

Since ε is Gumbel distributed with parameters (η, μ), then:

$F(\varepsilon) = \exp[-e^{-\mu(\varepsilon-\eta)}]$ and $f(\varepsilon) = \mu e^{-\mu(\varepsilon-\eta)} \exp[-e^{-\mu(\varepsilon-\eta)}]$, $\mu > 0$, $-\infty < \varepsilon < \infty$

Without loss of generality, set $\eta = 0$. Then:

$$U_1 = V_1 + \varepsilon_1 \sim \text{Gumbel}(V_1, \mu)$$
$$U_2 = V_2 + \varepsilon_2 \sim \text{Gumbel}(V_2, \mu)$$
$$\cdots$$
$$U_J = V_J + \varepsilon_J \sim \text{Gumbel}(V_J, \mu)$$

The probability that alternative 1 is chosen as below:

$$P_i(1) = \Pr[U_1 \geqslant U_2,\ U_1 \geqslant U_3,\ \cdots,\ U_1 \geqslant U_J]$$
$$= \int_{-\infty}^{\infty} f(U_1) \int_{-\infty}^{U_1} f(U_2) \int_{-\infty}^{U_1} f(U_3) \cdots \int_{-\infty}^{U_1} f(U_j) \mathrm{d}U_1 \mathrm{d}U_2 \mathrm{d}U_3 \cdots \mathrm{d}U_J = \frac{e^{\mu V1}}{e^{\mu V1} + e^{\mu V2} + \cdots + e^{\mu Vj}}$$

In the general multinomial logit model, the coefficients reflect the responses (tastes) of individuals in the population to the attributes of alternatives.

1.3.2 Random-coefficients Multinomial Logit Model

In the coefficients in the general multinomial logit model, β reflects the responses (tastes) of individuals in the population to the attributes of alternatives. A constant β means that all individuals' tastes are identical with respect to the observed attributes of alternatives embodied in the X_{ij} vectors. Consequently, the logit model formulation implies that all individuals of identical observed characteristics have identical tastes with respect to the observed attributes of alternative. The problems are emerging: a constant β may not be the case in reality; the variation in response to different attributes (unobserved heterogeneity) will influence individual's choice behavior; if such heterogeneity is ignored, the parameter estimates of the logit model will have a downward bias. In order to resolve the problems, a random-coefficient specification is used to deal with the unobserved heterogeneity. The parameters of the logit model are treated as realisation of random variables representing the individuals' preferences and their responses to the marketing variables.

The random utility can be written as below:

$U_{ij} = X'_{ij}\beta + \varepsilon_{ij}$ where β is a random vector independent of ε_{ij}

For a particular individual i, the vector $\Theta_i = \{\beta_{0i1}, \beta_{0i2}, \cdots, \beta_{0iJ}, \beta_{i1}, \beta_{i2}, \cdots, \beta_{iK}\}$ is assumed to be a realisation from a continuous multivariate distribution $G(\Theta)$. Conditional on the value of the vector Θ from a continuous distribution, the probability of a randomly drawn individual choosing alternative j on occasion t will be given by:

$$P_t(j|\Theta) = \frac{\exp\left\{\beta_{0j} + \sum_{k=1}^{K}\beta_k x_{jkt}\right\}}{\sum_{l=1}^{J}\exp\left\{\beta_{0j} + \sum_{k=1}^{K}\beta_k x_{jkt}\right\}}$$

Therefore, $P_t(j) = \int P_t(j \mid \Theta)\mathrm{d}G(\Theta)$.

It is difficult to specify the underlying distribution of heterogeneity, parameters usually involve evaluating difficult multiple integrals when a continuous multivariate distribution is used. Thus, maximum likelihood estimation (MLE) is selected to estimate the model parameters. The likelihood function for individual i conditional on the value of the vector Θ_s is given by:

$$L_i(\cdot \mid \Theta_s) = \prod_{t=1}^{T_i}\left\{\prod_{j=1}^{J}[P_{it}(j \mid \Theta_s)]^{\delta_{ijt}}\right\}$$

where δ_{ijt} is 1 if the ith individual chooses alternative j on occasion t and it is 0 otherwise. J is the total number of alternatives and T_i denotes the number of occasions made by the ith individual. Therefore, the overall sample likelihood function is given by:

$$L = \prod_{i=1}^{I}\left[\sum_{s=1}^{S}\left[\prod_{t=1}^{T_i}\left[\prod_{j=1}^{J}[P_{it}(j \mid \Theta_s)]^{\delta_{ijt}}\right]\right]\alpha(\Theta_s)\right]$$

The number of support point S can be determined by using a stopping rule procedure based on the Bayesian information criterion (BIC). $\mathrm{BIC} = -\mathrm{LL} + \frac{1}{2} * R * \log(N)$, where LL is log-likelihood; R represents the number of parameters estimates; N denotes the total number of observation in the sample.

1.4 Nested Logit Model

The nested logit model is distinct from the multinomial logit model in the way that the alternatives faced by a decision maker can be partitioned into subsets, named as "nests". The IIA (independence from irrelevant alternatives) property exists within the nests but not across the nests. In other words, for any two alternatives that are in the same nest, the ratio of probabilities is independent of the attributes of all other alternatives inside or outside the nest. It is clear that "Auto alone" and "Carpool" can form a nest and that "Bus" and "Rail" can form another from Figure 1.1.

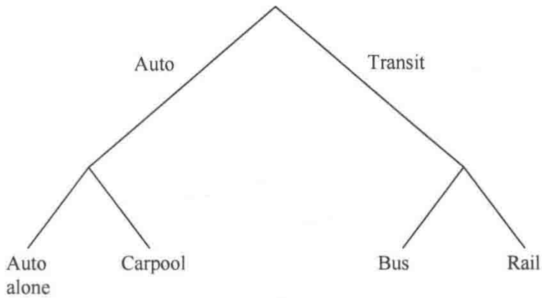

Figure 1.1 Tree diagram for mode choice

The nested framework can be presented in a hierarchical or tree structure such that each branch denotes a subnet of alternatives or named nest, every leaf on each branch denotes an alternative.

The nested logit model is derived by assuming the vector of the random components of the J alternatives, $\varepsilon_n = \langle \varepsilon_{n1}, \cdots, \varepsilon_{nJ} \rangle$ follows a type of generalized extreme value distribution (GEV) with the cumulative distribution:

$$F(\varepsilon_n) = \exp\left(-\sum_{k=1}^{K} \left(\sum_{j \in B_k} e^{-\varepsilon_{nj}/\lambda_k} \right)^{\lambda_k} \right)$$

where B_1 to B_k are K non-overlapping subsets or nests in which the set of J alternatives are partitioned into; λ_k measures the degree of independence of the error components in the utility among the alternatives in nest k.

Compare the cumulative distributions of logit and nested logit for the vector: $\varepsilon_n = \langle \varepsilon_{n1}, \cdots, \varepsilon_{nJ} \rangle$.

$$\text{For logit}: F(\varepsilon_n) = \prod_{j=1}^{J} \exp(-e^{-\varepsilon_{nj}}) = \exp\left(-\sum_{j=1}^{J} e^{-\varepsilon_{nj}}\right).$$

$$\text{For nested logit}: F(\varepsilon_n) = \exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_k} e^{-\varepsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right).$$

The value λ_k measures the degree of independence and $(1-\lambda_k)$ represents the degree of correlation of the error terms among the alternatives in nest k. The larger the value of λ_k, the smaller the correlation will be. When λ_k equals 1 for all k, the correlation of the error term will become 0, meaning that there is no correlation of the error terms among the alternatives for all the nests. It is clear that if all λ_k equal 1, the nest logit becomes a simple standard logit model.

In order to obtain a valid nested logit model, the value of λ_k should lie between 0 and 1 (Kling and Herriges 1995; Herriges and Kling 1996; Train et al. 1987; Lee 1999). If λ_k is bigger than 1, then the model will become inconsistent with utility maximizing behavior for some range of the explanatory variables. For λ_k less than 0, the model will become inconsistent with utility maximizing behavior for all value of explanatory variables.

By assuming the unobserved components following GEV, the probability of an individual choosing alternative i in subset B_k can be derived as follows:

$$P_{niB_k} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k-1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_l}}$$

The probability above can be used to demonstrate that the IIA property holds within a nest but not across nests. For instance, given that alternative i is in nest k and alternative m in nest l, the ratio of probabilities between choosing alternatives i and m becomes:

$$\frac{P_{niB_k}}{P_{nmB_l}} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k-1} \Big/ \sum_{l=1}^{K}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_l}\right)^{\lambda_l}}{e^{V_{nm}/\lambda_l}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l-1} \Big/ \sum_{l=1}^{K}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_l}\right)^{l}} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k-1}}{e^{V_{nm}/\lambda_l}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_l}\right)^{\lambda_l-1}}$$

If alternatives i and m are in different nests (i.e., $k \neq l$), the parentheses in the ratio can not cancel out, but that ratio depends on the characteristics of all alternatives in the nest k and l that contain alternatives i and m. Note that this ratio does not depend on the attributes of alternatives in nests other than k and l. Hence, a form of IIA holds, even for alternatives in different nests. That form of IIA can be described as "independence from irrelevant nests (IIN)".

If alternatives i and m are in the same nest, that is, k equals l, the terms in the parentheses are canceled out and the ratio then becomes: $\frac{P_{niB_k}}{P_{nmB_k}} = \frac{e^{V_{ni}/\lambda_k}}{e^{V_{nm}/\lambda_k}}$.

From that equation, one can see that the ratio of probabilities is independent of the characteristics of all alternatives other than i and m and IIA holds for alternatives within a nest.

One can use the likelihood ratio test to test whether all λ_k in the nested logit equal 1.

The deterministic component of the utility in nested logit model can also be divided into two parts. The first part varies over nest k and is denoted by W_{nk}, that part is constant for all the alternatives within a nest. The second part varies over the alternatives within a nest and is denoted by Y_{ni}. The utility is therefore written as follows:

$$U_{ni} = W_{nk} + Y_{ni} + \varepsilon_{ni}$$

Hence, the probability of an alternative in a nested logit can be formulated as a product of a marginal and conditional probabilities.

1.5 Probit Models

Probit models can handle random taste variation, they allow any patterns of substitutions and they can be applied to panel data with temporally correlated errors. The probit model is derived under the assumption of multivariate normal (MVN) distribution for the error component of the random utility function. The assumption may be appropriate as many random variables can be assumed to have a normal distribution. However, there are situations where the use of normal distribution is not appropriate.

The probit model:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad \text{for all } j$$

where $\varepsilon_n^{\mathrm{T}} = \{\varepsilon_{n1}, \varepsilon_{n2}, \cdots, \varepsilon_{nJ}\} \sim N(0, \Omega), \varnothing(\varepsilon_n) = \frac{1}{(2\pi)^{\frac{J}{2}} |\Omega|^{\frac{1}{2}}} e^{-\frac{1}{2}\varepsilon_n' \Omega^{-1} \varepsilon_n \frac{1}{2}}$.

The choice probability for alternative i is as below:

$$P_{ni} = \text{prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i)$$

There is no close form for the integral. It must be evaluated numerically by simulation.

Since that only differences in utility matter, one may work with the differences of utilities.

Let $\bar{U}_{nji}=U_{nj}-U_{ni}$, $\bar{V}_{nji}=V_{nj}-V_{ni}$ and $\bar{\varepsilon}_{nji}=\varepsilon_{nj}-\varepsilon_{ni}$ Then $P_{ni}=\Pr[\bar{U}_{nji}<0$ for all $j\neq i]$

Let $\bar{\varepsilon}_{ni}=\{\bar{\varepsilon}_{n1i},\bar{\varepsilon}_{n2i},\cdots,\bar{\varepsilon}_{nJi}\}$ (with $(J-1)$ components)

and $\varnothing(\tilde{\varepsilon}_{ni})=\dfrac{1}{(2\pi)^{\frac{J-1}{2}}\left|\bar{\Omega}_i\right|^{\frac{1}{2}}}e^{-\frac{1}{2}\varepsilon_{ni}'\bar{\Omega}_i\tilde{\varepsilon}_{ni}}$

where $\bar{\Omega}_i$ can be determined from Ω as follows.

Construct a matrix M_i by using a $(J-1)$ identity matrix. Add one extra column of -1s as the ith column. Then $\bar{\Omega}_i=M_i\Omega M_i^{\mathrm{T}}$.

For example:

For $J=3$ and $i=2$

$$M_i=\begin{pmatrix}1 & -1 & 0\\ 0 & -1 & 1\end{pmatrix},\quad \Omega=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{pmatrix}$$

Hence $\bar{\Omega}_2=\begin{pmatrix}1 & -1 & 0\\ 0 & -1 & 1\end{pmatrix}\begin{pmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{pmatrix}\begin{pmatrix}1 & 0\\ -1 & -1\\ 0 & 1\end{pmatrix}$

One of the two main reasons for using the probit model is for its ease of handling repeated choices from panel data (the other reason is for dealing with correlations among alternatives).

Consider a decision maker n who faces a choice among J alternatives in each of T time periods. The utility that he obtains from alternative j in period t is as below:

$$U_{njt}=V_{njt}+\varepsilon_{njt}$$

Since the T choices are made by the same decision maker, one would expect ε_{njt} to be correlated over time as well as over alternatives, since factors that are not observed by the researcher can persist over time (by the same decision maker). The vector of errors for all alternatives in all time periods is as below:

$$\varepsilon_n=\{\varepsilon_{n11},\cdots,\varepsilon_{nJ1},\varepsilon_{n12},\cdots,\varepsilon_{nJ2},\cdots,\varepsilon_{n1T},\cdots,\varepsilon_{nJT}\}^{\mathrm{T}}$$

The covariance matrix of this vector, Ω, is of dimension $JT\times JT$. Consider a sequence of choices of alternatives made by a decision maker: $\{i_1,\cdots,i_T\}$. The probability of that decision maker making such choices is as below:

$$P_{ni}=\Pr[U_{nit}>U_{njt},\forall j\neq i,\forall t]=\Pr[V_{nit}+\varepsilon_{nit}>V_{njt}+\varepsilon_{njt},\forall j\neq i,\forall t]$$

When the values of J and T are large, it is very difficult to find the choice probability for probit model. That is the main reason why the probit model is much less popular than the multinomial logit model. In Section 1.6, we will discuss the simulation