Discrete Choice Models: Mathematical Methods, Econometrics, and Data Science

Alfred Galichon
New York University and Sciences Po

Draft version, comments welcomed.
Last updated June 24, 2025.

Contents

To Marius, André and Jacqueline Galichon.

La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi agli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intendere la lingua, e a conoscere i caratteri nei quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi e altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto.

— Galileo Galilei (1564-1642), *Il Saggiatore*, cap.6

Preface

This book has emerged from several courses and lecture series I gave over the last few years, usually in front of a diverse audience of graduate students and researchers with a background in economics or in mathematics. The goal of this work is to provide a concise treatment of discrete choice models and their applications, leveraging the powerful elegance of mathematics. The mathematical foundations of these models is half optimization theory and half probability and statistics. As we shall develop, there are intricate connections with the theory of optimal transport, which was the subject of my previous book, *Optimal Transport Methods in Economics*. However, in these lectures the framework is broader. The focus is on choice, demand, price formation and equilibrium, and the statistical estimation of the related models. The exposition is influenced by my own research, and beyond that, to my own tastes, so there is no attempt at an exhaustive treatment of these models. It is not a survey or a guide to the literature.

Gabriele Buontempo, Enzo Di Pasquale, Alexandre d'Epinay, Alessandro Facchini, Clément Montes and Jiawei Qian have provided helpful research assistance at various stages of this project. I am grateful to the PhD students of my Applied Microeconometrics class at New York University who have spotted typos and made suggestions on earlier versions of this text, in particular Diego Cussen, Nilima Pisharody, Caleb Maresca, Vasco Villas-Boas and Bill Wang. I am hugely indebted to several people who did an exhaustive reading of earlier drafts: Giuseppe Cognata, Facundo Danza, Amine Fahli, Octavia Ghelfi, Leon Guzman, Antoine Jacquet, Isabelle Perrigne, Maxime Sylvestre. In addition, the editorial team at Princeton University Press, in particular Hannah Paul, and three anonymous reviewers, provided useful feedback on earlier versions of the manuscript. This project was nourished by insightful conversations with Pierre-André Chiappori, Pauline Corblet, Jeremy Fox, Jean-Marc Robin, Bernard Salanié, Larry Samuelson, Matt Shum and Jonathan Vogel. I am particularly grateful to Antoine Jacquet, who did not only read and comment, but suggested improvements in many places. Writing this book while serving as director of New York University's academic center in Paris required some multitasking; but I could not have completed this project in a reasonable amount of time without relying on a great team of dedicated administrators of the center, among whom Beth Epstein, Martina Faltova, Chioma Iwuegbu, Valérie Michelin, Marcus Neeld and Xavier Séguy. Much of my research agenda over the years has been funded by two grants of the European Research Council of European Union (starting grant No. 313699, and consolidator grant No. 866274). These grants have been transformative, and this book would probably not have come to exist without them.

Introduction

This book offers a mathematical perspective of discrete choice models and related topics. Discrete choice models, also called multinomial choice models, are used to model situations where the consumer needs to choose one option among several. One of the first area where it has been applied has been the choice of transportation mode, but the methodology has proved useful to understand many decisions whose nature is economic (choice of education, employment, retirement, consumption), demographic (fertility decisions, migration), sociological (marriage, friendship formation), geographic (trade, urban dwelling) and political (vote). While discrete choice models originated in transportation studies and in psychology in the middle of the 20th century, they have made their way to other disciplines in the last 70 years, up to a point that they are now at the core of practically all of the social sciences.

In this book (and in much of the literature), the discrete choice approach is based on the idea of individual rationality: it postulates that individual decision-makers associate a numerical valuation called "utility" to each option that they are faced with. The rational choice of decision-makers consists of picking the option with the largest associated amount of utility; the choice of each agent is thus obtained by the outcome of a simple optimization problem, maximizing utility over a finite ("discrete") set of options. These choices are then aggregated across individuals in order to account for the share of the population choosing each option, which allows to express the demand (or "market share") for every option.

Of course, the utilities that individuals associate with each option are most often not observed by the analyst, but they can be inferred by the latter based on the data of the choices made. Discrete choice models are thus nonlinear econometric models predicting categorical outcomes – which option has been picked. To train these models, one generally takes a parametric form of the utilities, and estimates the value of the parameter vector which lead to fitting as best as possible the observed choices according to some criterion. Many criteria exist, such as maximum likelihood, methods of moments, minimax regret, maximum score, revealed preferences. The nature of this criterion very much depends on how exactly the utilities are modeled, computational convenience, desired robustness to misspecification, etc.

An attractive feature of discrete choice models is the ability to conduct counterfactual analysis: once the utilities have been estimated, it is then possible to predict the change in aggregate demand if some policy intervention were to be implemented, like the implementation of a tax affecting a transportation mode. Thus, in contrast with other types of methods, such as regression analysis or factor analysis, multinomial choice methods allow us to address in-depth causal questions as they attempt to capture the core of the decision-making process. Almost built-in into discrete choice models is the ability to offer welfare considerations; indeed, in discrete choice models the individual utilities are the primary object of focus, so computing the social welfare is simply a matter of summation. The models will therefore allow us to predict in a very straightforward manner the welfare impact of a policy change.

An important class of discrete choice models rely on the *random utility* paradigm. Take the example of an invidivual needing to choose a transportation mode. In the random utility framework, one assumes that the utilities that agents associate with each mode have a "systematic" component, which depends deterministically on the agents and the option's observable characteristics, but have in addition a "idiosyncratic" component, which captures the part of the decision that cannot be captured by observable characteristics alone. Even if two agents have the exact same observable characteristics, one may choose the train and one may choose the plane, simply because their idiosyncratic utility terms differ. One therefore speaks of random utility, or stochastic choice, which is a bit of a misnomer because it does not necessarily mean that the idiosyncratic terms are randomly drawn by the agents, or that there is any randomness in the decision-making process. However, as the idiosyncratic term is unknown to the analyst, it is random from their point of view for all practical purposes. And although the analyst does not know the idiosyncratic terms, they typically postulate that the distribution is known, or at least, belongs to a parametric family of distributions. From the point of view of the analyst, the idiosyncratic term is a random utility term whose distribution is known.

The random utility paradigm does not allow to predict the choice that a particular individual will make, but, as chapter 1 shall show, it allows the analyst to compute the probability that a particular individual will choose one option or the other. This is closely connected to the classification problem in machine learning – in fact, many of the tools are the same. The random utility framework allows to compute aggregate quantities, such as the aggregate welfare (the sum of the welfare of individuals in the population), the predicted market shares, and the part of the aggregate welfare that is due to systematic utilities, and the part that is due the idiosyncratic terms – the latter quantity being defined (up to a sign) as the *entropy of choice*, a concept introduced in chapter 1. The analyst also needs to solve the inverse problem of recovering the systematic utilities based on the observation of the market shares, a fundamental problem called the "market share inversion problem". As we shall see in chapter 1, convex analysis is the appropriate mathematical framework to perform these calculations without making any significant assumptions on the random utilities. Thanks to convex analysis, we will be able to formulate the basic calculations as a convex optimization problem, which is both practically useful and will have important consequences for the analysis of the problem. For instance it will allow one to deduce results about the existence and uniqueness of a solution to the market share inversion problem.

As show in chapter 2, some distributions of random utilities lead to simple formulas for the expression of the welfare, the predicted market shares, and in some cases, the entropy of the choice and the demand inversion. The most famous case is the logit (or "logistic") framework, which assumes that the random utilities follow an extreme value distribution, more precisely independent and identically distributed Gumbel variables. The Gumbel distribution is one of the three stable distributions arising in extreme value theory; it is the limiting distribution (after renormalization) of the maximum of a large class of independently distributed random variables. As is probably already familiar to most readers, the logit model allows for simple formulas for the market shares and allows to solve the market share inversion problem in closed form. It also leads to an entropy of choice that coincides (up to a sign) with the Gibbs entropy.

Yet for all its appeal, the logit paradigm is a very rigid framework which has significant shortcomings. In the transportation mode example, it would specify that the random utility associated with taking bus, train and plane are independent, which does not capture the fact that some travelers may dislike air travel in a manner that cannot be predicted by their observable characteristics, thereby introducing a correlation between the random utilities associated with the "bus" and "train" options. Consequently, chapter 2 moves on to exploring distributions of random utility that retain tractability with more flexibility, in particular allowing for this type of correlation. An important class of such distributions presented there is the class of *multivariate extreme value distributions*, discussed in section 2.2. These random variables can be obtained by an ingenious combination of i.i.d. Gumbel variables used as factors, in a somewhat similar way any Gaussian vector can be obtained by a linear combination of i.i.d. standard normal factors. Multivariate extreme value distributions, which following Daniel McFadden are often called "Generalized Extreme Value" in econometrics, lead to a closed-form expression for the welfare function, the market shares, and sometimes also for the entropy of choice, as in the case of the nested logit model, one of the most important representatives of the class.

The logit framework plays a central role in structural estimation, as we begin to see in the subsequent chapter 3 on the logistic regression. Consider a stochastic choice problem where the systematic utilities belong to a parametric family and the random utilities are i.i.d. Gumbel. Given a parameter vector, the model predicts the probabilities that each agent will pick the various options, which leads to the specification of a tractable parametric family of choice probability. Assuming that the observations are independently sampled, one can then form the log-likelihood of the sample. In this setting, logistic regression is simply maximum likelihood estimation. It is one of the most important topics of this book, so the entire chapter 3 is dedicated to its study. In addition to being a maximum likelihood estimation problem, the logistic regression can be interpreted as a method of moments, and also as a minimax-regret procedure. There is a useful link with generalized linear models, an important part of the statistical toolbox which generalizes the Poisson regression: in section 3.4 we recall that connection, known as the "Poisson trick" among the machine learning community, which asserts that the logistic regression amounts to a Poisson regression with the addition of a fixed effect associated to each option. From the computational point of view, the logistic regression has many attractive features. It is a convex optimization problem, which leads to easy computational methods as discussed in section 3.3. One can state simple conditions to characterize the existence and the uniqueness of a parameter estimator, as carried out in section 3.5 and 3.6. When the dimensionality of the parameter is large, the model needs to be regularized by adding a penalty term to prevent overfitting; this can be done for various types of regularizations such as LASSO, with dedicated algorithms such as the proximal gradient descent methods also recalled in chapter 3.7.

All the models seen until this point have been based on the logistic framework or its multivariate extreme value generalization (in chapter 2). In contrast, the *characteristics-based approach*, a popular alternative approach to demand estimation, is introduced in chapter 4. It relinquishes the hope of getting closed-form expression for the welfare and other quantities, but it provides a simple and easy-to-interpret geometric description of the interactions between the characteristics of the decision-maker and the characteristics associated with each option. In the most popular version, explored in section 4.1, one assumes that this interaction is a scalar product, or more generally, a bilinear form. This allows one to make use of Euclidian geometry to describe the choice problem, as one may then locate the characteristics of the consumers who choose a particular option on a polytope in the characteristics space. As we shall see, this problem is closely related to the theory of optimal transport, which was the focus of my previous book [87], and one can leverage the power of computational geometry to efficiently compute the predicted market shares and perform demand inversion. Mixing the characteristics approach and the logistic framework leads to the *random coefficient logit* specification explored in section 4.2, and proposed by Berry, Levinsohn and Pakes. In this model, the random utility terms is a sum of two independent components, one term with a Gumbel distribution and one that is characteristics-based. Here again, the theory of optimal transport offers insights as discrete choice problems with random coefficient logit random utility can be reformulated in terms of an entropic optimal transport problem, for which many computational tools exist. The random coefficient logit specification is the foundation for Berry, Levinsohn and Pakes' method for structural estimation which is then presented in section 4.4. This framework deals with endogeneity and allows to model not only the demand side, but possibly the supply side as well, with imperfect competition among sellers.

Up to this point, the book has regarded the systematic utilities as an exogenous primitive of the model. However, one may want to incorporate random utility specifications into equilibrium models where the systematic utilities depend on the price or other quantities that are adjusted at equilibrium. Chapter 5 offers several examples of such type of models of allocation and equilibrium pricing. The first example is international trade, the primary model of which is the *gravity equation*. In the gravity equation, trade flows are adjusted by supply and demand according to the propensity of pairs of countries to trade with each other, which depends on factor such as geographic distance, trade agreements, cultural proximity, and many other regressors; but they also depend on the sizes of the countries, as measured by their volumes of exports and imports. The trade flows are therefore adjusted at equilibrium by prices, which are materialized by exporter- and importer-fixed effects.

Section 5.3 recalls an important reformulation of the gravity equation as a Poisson equation with two-way fixed effects, thus extending the "Poisson trick" to bipartite models.

The incorporation of logistic random utility in various microeconomic frameworks yields variants of the gravity equation. This is the case of the models of matching with transfers pioneered by Gary Becker, and section 5.4 introduces the class of *empirical models of matching*, which are standard models of matching with the addition of a random utility term. The introduction of the random utility term has multiple benefits: accounting for the heterogeneity that is not observed by the analyst, imposing uniqueness of the equilibrium matching in a large population, and providing smoothness which is desirable for estimation and inference purposes. A pioneering example of an empirical matching framework is the model of Choo and Siow, which has been successfully applied to the analysis of the marriage market, and, to a lesser extent, to the labor market. The Choo and Siow framework is a model of bipartite matching with transferable utility, meaning that prices (wages or other forms of utility transfers via bargaining) adjust at equilibrium in order to clear supply and demand. It incorporates logit heterogeneity, meaning that agents' utilities are the sum of a systematic utility term, which is the outcome of a bargaining process, plus a random utility term following an i.i.d. Gumbel distribution. This structure allows the reformulation the problem of equilibrium matching as a pair of interdependent discrete choice models on each side of the market, and its solution using convex optimization or as a generalized linear model with two-way fixed effects.

The one-to-one, bipartite framework of the Choo and Siow model can be viewed as restrictive in some situations. For example if one would like to study the same-sex marriage market (which is not bipartite), or the employer-employee matching market (which is not one-to-one). Fortunately, as seen in section 5.5, empirical models of matching à-la Choo and Siow can be extended quite directly to more general models of coalition formation. While models of matchings do not necessarily put explicit emphasis on prices (although they assume the existence of prices to clear the market), *hedonic models*, covered in section 5.6, directly model their formation. In that paradigm, a good is produced and consumed in different varieties, or qualities by heterogeneous producers and consumers. For instance, cars are differentiated on the quality space, and are imperfect substitutes for one another, both from the consumers' and the producers' perspectives. The prices enter the systematic utilities of both consumers and producers, and both side of the market face a discrete choice problem. At equilibrium, prices adjust so that the demand for each quality from the consumers' side equates the corresponding supply on the producers' side, and the structural parameters on both sides can be learned in that way.

Our discussion until now has focused on static models, in the sense that it has not taken into account the effect of present decisions on future outcomes. Take maintenance decisions for example: deciding on the preventive maintenance of a vehicle generates a present-period cost which may exceed the present-period benefit; but performing the maintenance procedure may be justified because it will decrease the costs of operations in the future. Of course, one could incorporate a discounted value in the systematic utilities associated with the maintenance or no-maintenance decision, but this value depends on the various decisions that will be taken in the future, as the decision-maker will be faced with other choices (maintaining, how to operate, selling the vehicle, etc.) at each future period. This is the core of the issue that *dynamic discrete choice models*, covered in chapter 6, are tackling. The chapter starts with a discussion in section 6.1 of these models from a linear optimization point of view which is not typical in most treatments of the topics, but which highlights the connections with the models seen thus far. Logistic random utility is then explicitly introduced in section 6.2, where it is shown that a dynamic discrete choice model is made of a series of static discrete choice problems dynamically linked together by a *Bellman equation*. A distinction must be made between models where a finite number of decisions are made sequentially, which is the finite-horizon case, and models where there is no end to the sequence of decision problems, the infinite-horizon case, where one must discount the values of future period utilities. In the infinite horizon case, the absence of a time horizon induces some stationarity in the value associated to each option in a specific state. Inference is worked out, in section 6.3 for the finite-horizon case, and in section 6.6 for the infinite-horizon case. The chapter concludes with an investigation in section 6.9 of dynamic matching models, which are two-sided discrete choice models.

While prices, understood as adjustable monetary transfers, have played a prevalent role in our analysis, one should note that there are many situations in the economy when demand is not regulated by monetary prices, but by other adjustment mechanisms. Taxis are a good example: as the price of taxi rides is generally regulated and fixed, the demand for taxis in times of short supply is not regulated by prices, but by waiting lines. Though waiting lines or other manifestation of congestion, the utility associated with the options for which the capacity in insufficient will be decreased, up to a point at which the demand exactly meets capacity. In that case, waiting times therefore play the role of numéraire, instead of money. While waiting times share some similarities with monetary prices, they have a big difference: they cannot be transferred to the other side of the market. Picking up a passenger who has waited a long time does not make a driver better off. As a result, chapter 7 studies *equilibirum models with non-transferable utility* and needs to build a distinctive mathematical machinery on top of the exisiting one. A first objective in the chapter is to understand the effect of the capacity constraints on the systematic utilities through shadow prices. The consequence of rationing on welfare is studied in section 7.2 and monotone comparative statics, which is the study of how utilities respond to changes in capacity, is studied in section 7.3. The machinery developed in this chapter leads to the introduction in section 7.5 of an empirical matching model without prices, which is a nontransferable utility matching model with the addition of random utility terms. The celebrated deferred acceptance algorithm of Gale and Shapley is revisited and adapted to fit into the framework of discrete choice in section 7.6, and its insightful reinterpretation by Hatfield and Milgrom is in turn adapted to the context of discrete choice models and presented in section 7.7.

While the first chapter showed that most of the welfare and demand inversion analysis could be performed without assuming the logit distributional framework, and chapters 2 and 4 covered examples of distributions of random utilities that could be used as alternatives to logit, the subsequent chapters (from chapter 3 to chapter 7) used almost exclusively the logit framework in the interest of simplicity and because of the link with logistic regression. However, it is interesting to note that these chapters could have been written with very general distributions. Chapter 8 goes back to the models of these four chapters and shows how they can be naturally generalized beyond logit random utilities. Some attention needs to be paid however, to how the adaptation is done. For instance, the natural estimation paradigm is no longer maximum likelihood, as in this context the maximum likelihood estimation problem is not convex anymore, but minimax regret estimation, which coincides with maximum likelihood in the logistic case, but retains convexity and the moment matching interpretation outside of that case. With this empirical strategy in mind, we are able to revisit one-sided discrete choice models (section 8.1), empirical models of two-sided matching (section 8.2), models of coalition formation (section 8.3), and dynamic models (section 8.4).

Positioning. This book touches upon many disciplines from different horizons. As it is hopefully clear from the various examples alluded to above, discrete choice methods span over many disciplines in the social sciences. We have mentioned **economics**, **statistics**, **political science**, **marketing**, **psychology**, **operations research**, **geography**, **sociology**, **transportation studies**, yet the list is not exhaustive. But from a mathematical standpoint, discrete choice models are connected with an exciting mix of tools, whose diversity is evidenced by the range of the topics spanned in the mathematical appendix, appendix A. The general theory and the welfare analysis seen in chapter 1 makes a heavy use of **optimization theory**, more specifically **linear programming**, **convex analysis** and **optimal transport**. The use of special distributions in chapter 2 borrows from **probability theory** and **mathematical statistics**, more specifically **extreme value distributions**. The characteristics approach seen there used some **Euclidian geometry**, more specifically **polyhedral geometry**. The Poisson regression formulation in chapter 3 uses standard **statistical inference theory**. The gravity equation and the empirical models of matching seen in chapter 5 are closely connected with **entropic optimal transport**. The dynamic discrete choice models seen in chapter 6 connect with **dynamic programming**, more specifically **reinforcement learning** and **Markov decision processes**. And the chapter on discrete choice models with limited availability builds on the important theory of **M-functions** and **submodular optimization**. The text makes frequent appeals to **tensor algebra** with **vectorization** and **Kronecker products**, and to **numerical optimization algorithms**. A particular emphasis has been placed on coding. Finally, python code demos in appendix B should make a strong point that far from being abstract constructions, all the concepts introduced in this book are practical and implementable.

Scope. In writing this book, some choices were made, and the book does not address some topics that are related with the book's subject. The book says little, for example, about welfare analysis. Topics such as maximum score estimation, partial identification, counterfactual analysis, assortment optimization are not covered. The random utility paradigm is prevalent. Other points of view, such as the Bayesian approach, are not represented. Also, some valid critiques of discrete choice models (such as the centrality of the assumption of individual rationality, the reliance on distributional assumptions on unobserved heterogeneity, the individualistic approach) are given little or no discussion in this text. The reader should not expect an encyclopedic treatment in this text. The selection of topics and points of view offered here is obviously biased towards the author's own tastes and inclinations.

Audience. Given the versatility of the topics, it is hard to predict if the book will appeal to a very narrow or to a very wide audience, but while writing this book I have bet on the latter. Graduate students and researchers in one of the social sciences disciplines listed above with a good command of college-level mathematics should find it useful to gain a deeper understanding of the mathematical tools on which these models rest. Mathematicians and data scientists may also find the book useful to understand what type of applications and models economists are working on. My previous book, *Optimal Transport Methods in Economics* [87], has been fortunate to attract this dual audience, and therefore create a bridge between two communities.

My hope is that the same should happen for the present project.

Prerequisites. This book features a number of mathematical appendices to make it as self-contained as possible, but it does assume a minimal amount of mathematical knowledge. The reader is expected to possess college-level mathematical notions, like the basis of linear algebra, of real analysis, and of probability and statistics. For example, the book assumes prior knowledge of the matrix product, but not of the Kronecker product; of an exponential distribution, but not of an exponential family; of maximum likelihood estimation but not of M-estimation; of ordinary least squares, but not of generalized linear models. Notions such as "almost surely", "absolutely continuous distribution", "converging subsequence", "full rank matrix"... will be assumed part of common knowledge. Some knowledge of optimization theory will be helpful but not strictly required as all the notions are introduced in several appendices. The book also assumes a basic knowledge of Python; without it, the reader won't be able to appreciate the code demos which are a significant part of the learning experience. The appendices do not feature a Python tutorial as there are many such tutorials on the web. They do cover more advanced topics, such as automatic differentiation.

Organization of this book

This book is organized in seven chapters and two appendices. Chapter 1 covers the basics of choice models with random utility, and chapter 4 gives examples of such models. Chapter 3 connects random utility models with the logistic regression and generalized linear models. Chapter 5 studies applications to trade, matching and equilibrium pricing models. Chapter 6 covers dynamic discrete choice models. Chapter 7 deals with discrete choice models under capacity constraints. Chapter 8 extends all the models introduced in the book to general distributions of random utility. Appendix A provides the mathematical toolbox needed for this book. Appendix B provides code demos that illustrate the chapters of this book.

Notations and conventions

Terminology. Our use of mathematical terms will try to find a good balance between precision and convenience. By *probability measure* we shall mean a Borel probability measure; by a *set*, a measurable set. A *continuous probability* or *continuous distribution* will mean a probability measure which is absolutely continuous with respect to the Lebesgue measure; a *convex* function will mean a convex function in the usual sense which can take any real value or $+\infty$, is lower semi-continuous, and which is not identically equal to $+\infty$. A *rectangle* of $\mathbb{R}^d$ is a Cartesian product of d intervals of the real line (which may be either closed or open, either bounded or unbounded).

Abbreviations. We will use a limited number of classical abbreviations: *c.d.f.* for "cumulative distribution function"; *p.d.f.* for "probability density function"; *a.s.* for "almost surely"; *s.t.* for "such that" or "subject to"; *l.s.c.* for "lower semicontinuous", *u.s.c.* for "upper semicontinuous"; and *w.l.o.g.* for "without loss of generality". A *poset* will mean a "partially ordered set".

Standard notations. The following notations, adopted throughout the book, are more or less standard in the literature. The scalar product of two vectors x and y of the same dimensions is denoted $x^\top y$. The Euclidean norm of x is denoted $\|x\|$. Given a function $f : \mathbb{R}^d \to \mathbb{R}$, the *gradient* of f at x, denoted $\nabla f(x)$, is the vector of partial derivatives $(\partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_d)^\top$, and $D^2 f(x)$ is the *Hessian matrix* at x, which is the matrix of second derivatives $(\partial^2 f(x)/\partial x^i \partial x^j)$, $1 \leq i, j \leq d$. The set of $y \in \mathbb{R}^d$ such that $f(z) \geq (z-x)^\top y + f(x)$ for all $z \in \mathbb{R}^d$ defines the *subdifferential* of f at x. The *Legendre-Fenchel* transform of f, denoted f^*, is defined as $f^*(y) = \max_{x \in \mathbb{R}^d} \left\{ x^\top y - f(x) \right\}$. Given a function $f : \mathbb{R}^k \to \mathbb{R}^l$, the *Jacobian matrix* of f at x, denoted $Df(x)$, is the matrix of partial derivatives $(\partial f^i(x)/\partial x^j)$, $1 \leq i \leq l$, $1 \leq j \leq k$. For X a subset of $\mathbb{R}^n$, the notation $conv\,(X)$ stands for the convex hull of X; these are the points of the form $\sum_{i=1}^I \lambda_i x_i$, $x_i \in X$, $\lambda_i \geq 0$, $\sum_{i=1}^I \lambda_i = 1$, for any family (x_i) of elements of X. The notation $cone\,(X)$ stands for the convex cone generated by X, which has the same definition without the requirement $\sum_{i=1}^I \lambda_i = 1$. The *Dirac mass* at x_0, denoted δ_{x_0}, is the probability distribution which gives unit mass to x_0. Given a compact set C of $\mathbb{R}^d$, $\mathcal{U}(C)$ is the uniform probability distribution on C. The set $L^1(\mathcal{P})$ is the set of functions which are integrable with respect to the probability $\mathcal{P}$. The set $\mathcal{M}(\mathcal{P}, \mathcal{Q})$ is the set of probability measures π such that if $(X, Y) \sim \pi$, then $X \sim \mathcal{P}$ and $Y \sim \mathcal{Q}$. If $\mathcal{P}$ is a probability distribution over $\mathbb{R}^d$ and $T : \mathbb{R}^d \to \mathbb{R}^{d'}$, then the *push-forward* of $\mathcal{P}$ by T, denoted $T\#\mathcal{P}$, is a probability distribution on $\mathbb{R}^{d'}$ which is the distribution of $T(X)$ where $X \sim \mathcal{P}$. The expectation of a random vector $X \sim \mathcal{P}$ is denoted $\mathbb{E}_\mathcal{P} X$, and the variance-covariance matrix of X is denoted $\mathbb{V}_\mathcal{P}(X) := \mathbb{E}_\mathcal{P}[XX^\top] - (\mathbb{E}_\mathcal{P}[X])(\mathbb{E}_\mathcal{P}[X])^\top$. The notation $X \perp\!\!\!\perp Y$ denotes that the random variables X and Y are independent. Also, the c.d.f. associated with $X \sim \mathcal{P}$ is denoted indifferently F_X or $F_\mathcal{P}$. The *quantile* of that distribution, denoted Q_X or $Q_\mathcal{P}$, is defined as the right-continuous pseudo-inverse of the c.d.f., namely $Q_X(t) = \inf \{x : F_X(x) > t\}$. Given a set E, the *power set* of E, denoted 2^E, is the set of subsets of E. If $(E, \geq)$ is a poset and $x \in E$, the notation $[x, \infty)$ will denote the set $\{y \in E : x \leq y\}$; similarly, the set $(-\infty, x]$ will denote the set $\{y \in E : x \geq y\}$. A *correspondence* Γ from E to F, denoted $\Gamma : E \rightrightarrows F$ is a map $\Gamma : E \to 2^F$. Given a lattice L, the set of sublattices of L is denoted $\mathcal{L}(L)$. Given x and y in $\mathbb{R}^d$, the notation $x \ll y$ means $x_i < y_i$ for all $1 \leq i \leq d$. Given $x \in \mathbb{R}^d$ and $B \subseteq \{1, ..., d\}$, we denote x_B the subvector of x whose indices are in B, and we identify x with (x_B, x_{B^c}), where B^c is the complement of B in $\{1, ..., d\}$. Given a subset E of $\mathbb{R}^d$, the *interior* of E, denoted E^{int}, is defined as the complement of the closure of the complement of E. The *extended real line*, denoted $\overline{\mathbb{R}}$, is defined as the set $\mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$. A *permutation* of size n is a bijection of $\{1, ..., n\}$ onto itself. If n is a natural number, $[n]$ denotes the set $\{1, ..., n\}$. The set of permutations of size n is denoted $\mathfrak{S}_n$, and called the *symmetric group*. Given a finite set E, the *cardinality* or number of elements of E is denoted $|E|$. Following standard conventions in economics, given a vector $x \in \mathbb{R}^d$ and an index $i \in \{1, ..., d\}$, the notation x_{-i} will represent the subvector of x in $\mathbb{R}^{d-1}$ obtained by removing the i-th entry. Given a set X, $cl(X)$ is the closure of X, $conv(X)$ is its convex hull, and $cch(X)$ is its convex closure. The *convex indicator function* of X, denoted ι_X, is equal to 0 on X, and to $+\infty$ on the complement of X. The *binary indicator function*, denoted $\mathbf{1}_X$, is equal to 1 on X and to 0 on the complement of X. Whenever there is no ambiguity, we will call either function an *indicator function*. For $X \in \mathbb{N}$, and $x \in [X]$, $\mathbf{e}_X^x$ denotes the x-th vector of the canonical basis of $\mathbb{R}^X$; it is the x-th column of the identity matrix of order X. Given two vectors z and z' in $\mathbb{R}^n$, we denote $z \wedge z'$ the vector $(\min(z_i, z_i'))_i$, and $z \vee z'$ the vector $(\max(z_i, z_i'))_i$. We denote z^+ the vector $z \vee 0$ and z^- the vector $(-z) \vee 0$. If v is a vector of $\mathbb{R}^n$, we denote $diag(v)$ or $\boldsymbol{\Delta}_v$ the square diagonal matrix of size $n \times n$ with v on the diagonal.

Notations introduced in this book. This book will also introduce some original notations that are not standard in the literature. Given a probability distribution $\mathcal{P}$ over $\mathbb{R}^{\mathcal{Y}}$, the *welfare function* a.k.a. *Emax function* $G_{\mathcal{P}}$ is denoted as $G_{\mathcal{P}}(U) = \mathbb{E}_{\mathcal{P}}[\max_{y \in \mathcal{Y}} \{U_y + \varepsilon_y\}]$, where $\varepsilon \sim \mathcal{P}$. When there is no ambiguity the subscript $\mathcal{P}$ will be omitted. When $G_{\mathcal{P}}$ is differentiable, its gradient is denoted $\boldsymbol{\pi}_{\mathcal{P}}(U) = \nabla G_{\mathcal{P}}(U)$ and called the *market share map*. The Legendre-Fenchel transform of $G_{\mathcal{P}}$, denoted in a standard way $G_{\mathcal{P}}^*$, will be referred to as the *generalized entropy of choice* associated with $\mathcal{P}$. If A is a matrix of term A_{ij}, and $f : \mathbb{R} \to \mathbb{R}$, $f(A)$ denotes the matrix of term $f(A_{ij})$. The notation $[0{:}n]$ denotes the set of integers $\{0, 1, ..., n\}$. If I and J are integers, the notation $[I \times J]$ denotes the list of pairs ij for $1 \leq i \leq I$, and $1 \leq j \leq J$, ordered in the lexicographic ordering: $11, 12, ..., 1J, 21, 22, ..., 2J, ..., I1, I2, ..., IJ$. This notation extends to lists of triples

$[I \times J \times K]$ and to any tuples.