Economic applications of optimal transport

Alfred Galichon (New York University)

Summer school "Optimal transport: numerical methods and applications," Lake Como school of advanced studies, May 7-11 2018
Lecture 1. Multinomial choice

- Introduction to multinomial and matching models
- Applications to consumer demand, labor markets, marriage will be given
- We'll study the mathematical structures, computation, econometrics
- Code in R is provided on the course directory (`https://github.com/alfredgalichon/ucla-2018`). Other language work but the LP solver used will be Gurobi, so the language of choice should have a convenient interface to these.

- Mathematical foundations:
  - [OTON] C. Villani, *Optimal Transport: Old and New*, AMS, 2008.
  - [OTAM] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Birkhäuser, 2015.
- Introduction with a fluid mechanics point of view:
  - [TOT] C. Villani, *Topics in Optimal Transportation*, AMS, 2003.
- Computational focus:
  - [NOT] G. Peyré, M. Cuturi (2018). *Numerical optimal transport*, Arxiv.
- Economics focus:
  - [OTME] A. Galichon. *Optimal Transport Methods in Economics*, Princeton, 2016.
  - [MWT] P.-A. Chiappori. *Matching with Transfers: The Economics of Love and Marriage*, Princeton, 2017.

Outline

1. Multinomial choice models and their inversion
2. Separable matching models with heterogeneity
3. Affinity estimation: a framework for statistical inference in matching models
4. Equilibrium transport: incorporating taxes in matching models

- Emax operator and generalized entropy of choice
- The Daly-Zachary-Williams theorem
- Models: GEV, the pure characteristics models, the random coefficient logit model, the probit model
- The inversion theorem

- [OTME], Ch. 9.2 and App. E
- McFadden (1981). "Econometric Models of Probabilistic Choice," in C.F. Manski and D. McFadden (eds.), *Structural analysis of discrete data with econometric applications*, MIT Press.
- McFadden (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration". *Ectra*.
- Berry, Levinsohn, and Pakes (1995). "Automobile Prices in Market Equilibrium," *Econometrica*.
- Berry and Pakes (2007). The pure characteristics demand model". *IER*.
- Train. (2009). *Discrete Choice Methods with Simulation*. 2nd Edition. Cambridge University Press.
- G and Salanié (2017). "Cupid's invisible hands". Preprint.
- Chiong, G and Shum, "Duality in Discrete Choice Models". *Quantitative Economics*, 2016.
- Bonnet, G, O'Hara and Shum (2017). "Yogurts choose consumers? Identification of Random Utility Models via Two-Sided Matching". Working paper.
- Greene and Hensher (1997) Multinomial logit and discrete choice models.

- Assume a consumer is facing a number of options $y \in \mathcal{Y}_0 = \mathcal{Y} \cup \{0\}$, where $y = 0$ is a default option. The consumer is drawing a utility shock which is a vector $\varepsilon = (\varepsilon_0, \ldots, \varepsilon_{|\mathcal{Y}|}) \sim \mathbf{P}$ such that the utility of option y is $U_y + \varepsilon_y$, while the outside option yields utility ε_0.
- U is called vector of *systematic utilities*; ε is called vector of *utility shocks*.
- We assume thoughout that $\mathbf{P}$ has a density with respect to the Lebesgue measure, and has full support.
- The preferred option is the one which attains the maximum in

$$\max_{y \in \mathcal{Y}} \left\{ U_y + \varepsilon_y, \varepsilon_0 \right\}.$$

Section 1

EMAX OPERATOR AND DEMAND MAP

- Let $s_y = \sigma_y(U)$ be the probability of choosing option y, where σ is given by

$$\sigma_y(U) = \Pr(U_y + \varepsilon_y \geq U_z + \varepsilon_z \text{ for all } z \in \mathcal{Y}_0).$$

The map σ is called *demand map*, and the vector s is called vector of market shares, or vector of choice probabilities.

- Note that if $s = \sigma(U)$, then $s_y > 0$ for all $y \in \mathcal{Y}_0$ and $\sum_{y \in \mathcal{Y}_0} s_y = 1$.
- Note that because the distribution $\mathbf{P}$ of ε is continuous, the probability of being indifferent between two options is zero, and hence we could have indifferently replaced weak preference $\geq$ by strict preference $>$. Without this, choice probabilities may not have been well defined.

- $\sigma_y(U)$ is increasing in U_y.
- $\sigma_y(U)$ is weakly decreasing in $U_{y'}$ for $y' \neq y$.
- If one replaces (U_y) by $(U_y + c)$, for a constant c, one has $\sigma(U + c) = \sigma(U)$.

- Because of the last property, we can normalize the utility of one of the alternatives. We will normalize the utility of the utility associated to $y = 0$, and hence take
$$U_0 = 0.$$
- Thus in the sequel, σ will be seen as a mapping from $\mathbb{R}^{\mathcal{Y}}$ to the set of $(s_y)_{y\in\mathcal{Y}}$ such that $s_y > 0$ and $\sum_{y\in\mathcal{Y}} s_y < 1$, and the choice probability of alternative $y = 0$ is recovered by
$$s_0 = 1 - \sum_{y\in\mathcal{Y}} s_y.$$

- Define the expected indirect utility of consumers by

$$G(U) = \mathbb{E}\left[\max_{y\in\mathcal{Y}}(U_y + \varepsilon_y, \varepsilon_0)\right]$$

  This is called *Emax operator*, a.k.a. *McFadden's surplus function*.

- As the expectation of the maximum of terms which are linear in U, G is convex function in U (strictly convex in fact), and

$$\frac{\partial G}{\partial U_y}(U) = \Pr(U_y + \varepsilon_y \geq U_z + \varepsilon_z \text{ for all } z \in \mathcal{Y}_0).$$

  But the right-hand side is simply the probability s_y of chosing option y; therefore, we get:
  Theorem (Daly-Zachary-Williams). *The demand map σ is the gradient of the Emax operator G, that is*

$$\sigma(U) = \nabla G(U). \tag{1}$$

- Assume that $\mathbf{P}$ is the distribution of i.i.d. *centered type I extreme value* a.k.a. *centered Gumbel* terms, which has c.d.f.

$$F(z) = \exp\left(-\exp\left(-x + \gamma\right)\right)$$

where $\gamma = 0.5772...$ (Euler's constant). The mean of this distribution is zero.

- Basic fact from extreme value theory: if ε_1,...,ε_n are i.i.d. Gumbel distributions, then $\max\left\{U_y + \varepsilon_y\right\}$ has the same distribution as $\log\left(\sum_{y=1}^n \exp U_y\right) + \epsilon$, where ϵ is also a Gumbel. (Proof of this fact later).
- Notes:
  - This distribution is sometimes called the "Gumbel max" distribution, to contrast it with the distribution of its opposite, which is then called "Gumbel min".
  - The literature usually calls "standard Gumbel" the distribution with c.d.f. $\exp\left(-\exp\left(-x\right)\right)$; but that distribution has mean γ, which is why we slightly depart from the convention.

- The Emax operator associated with the logit model can be given in closed form as

$$G(U) = \log\left(1 + \sum_{y\in\mathcal{Y}} \exp(U_y)\right)$$

where $s_0 = 1 - \sum_{y\in\mathcal{Y}} s_y$. This is called a *log-partition function*.

- As a result, the choice probability of alternative y is proportional to the exponential of the systematic utility associated with U, that is

$$\sigma_y(U) = \frac{\exp U_y}{1 + \sum_{y'\in\mathcal{Y}} \exp(U_{y'})}$$

which is called a *Gibbs distribution*.

- Assume that the random utility shock is scaled by a factor T. Then

$$\sigma_y(U) = \frac{\exp(U_y/T)}{1 + \sum_{y'\in\mathcal{Y}} \exp(U_{y'}/T)}$$

which is sometimes called the *soft-max operator*, and converges as $T \to 0$ toward

$$\max_{y\in\mathcal{Y}} \{U_y, 0\}.$$

Section 2

DEMAND INVERSION AND THE ENTROPY OF CHOICE

- In many settings, the econometrician observes the market shares s_y and wants to deduce the corresponding vector of systematic utilities. That is, we would like to solve:
  Problem. *Given a vector s with positive entries satisfying $\sum_{y \in \mathcal{Y}} s_y < 1$, characterize and compute the set*

$$\sigma^{-1}(s) = \left\{ U \in \mathbb{R}^{\mathcal{Y}} : \sigma(U) = s \right\}.$$

- This problem is called "demand inversion," or "conditional choice probability inversion," or "identification problem." It is a central issue in econometrics/industrial organization and will be a key building block for matching models.

- We saw in Lecture 3 how to invert gradient of convex functions: if G is strictly convex and C^1, then

$$\sigma^{-1}(s) = \nabla G^{-1}(s) = \nabla G^*(s).$$

- G^* is the Legendre-Fenchel transform of G; we call it the *entropy of choice*, defined by

$$G^*(s) = \max_U \left\{ \sum_{y \in \mathcal{Y}} s_y U_y - G(U) \right\}. \tag{2}$$

- Hence, $\sigma^{-1}(s)$ is the vector U such that

$$U \in \arg\max_U \left\{ \sum_{y \in \mathcal{Y}} s_y U_y - G(U) \right\}.$$

- Convex duality implies that if s and U are related by $s \in \partial G\left(U\right)$, then

$$G(U) = \sum_{y \in \mathcal{Y}} s_y U_y - G_x^*(s). \tag{3}$$

- But letting $Y = \arg\max_y \left\{U_y + \varepsilon_y\right\}$, $G\left(U\right) = \mathbb{E}\left[U_Y + \varepsilon_Y\right]$ implies

$$G(U) = \sum_{y \in \mathcal{Y}} s_y U_y + \mathbb{E}\left[\varepsilon_Y\right],$$

thus one has

$$G^*(s) = -\mathbb{E}\left[\varepsilon_Y\right]. \tag{4}$$

Hence, the entropy of choice $G^*\left(s\right)$ is interpreted as minus the expected amount of heterogeneity needed to rationalize the choice probabilities s.

- Then
$$G^{*}(s) = s_0 \log(s_0) + \sum_{y \in \mathcal{Y}} s_y \log s_y$$
where $s_0 = 1 - \sum_{y \in \mathcal{Y}} s_y$. Hence, G^{*} is a bona fide entropy function when **P** is Gumbel–hence the name of *entropy of choice* in general.
- As a result,
$$\sigma_y^{-1}(s) = \log \frac{s_y}{s_0}$$
which is the celebrated *log-odds ratio formula*: the log of the odds of alternatives y and 0 identify the difference between the systematic utilities of these alternatives.

Section 3

Models of demand

Let $\mathbf{F}$ be a cumulative distribution such that function g defined by

$$g(x_1, ..., x_n) = -\log \mathbf{F}(-\log x_1, ..., -\log x_n) \tag{5}$$

is positive homogeneous of degree 1. (This inverts into $\mathbf{F}(u_1, ..., u_n) = \exp\left(-g\left(e^{-u_1}, ..., e^{-u_n}\right)\right)$). We have by a theorem of McFadden (1978):

THEOREM

Let $(\varepsilon_y)_{1 \leq y \leq n}$ be a random vector with c.d.f. $\mathbf{F}$, and define

$$Z = \max_{y=1,\ldots,n} \{U_y + \varepsilon_y\}.$$

Then Z has the same distribution as $\log g\left(e^{U_1}, ..., e^{U_n}\right) + \gamma + \epsilon$, where ϵ is a standard Gumbel. In particular,

$$\mathbb{E}\left[\max_{y=1,\ldots,n} \{U_y + \varepsilon_y\}\right] = \log g\left(e^{U_1}, ..., e^{U_n}\right) + \gamma$$

where γ is the Euler constant $\gamma \simeq 0.5772$.

PROOF.

Let F_Z be the c.d.f. of $Z = \max_{y=1,...,n} \{U_y + \varepsilon_y\}$. One has

$$\begin{aligned} F_Z(z) &= \Pr\left(\max_{y=1,...,n} \{U_y + \varepsilon_y\} \leq z\right) = \Pr\left(\forall y:\ \varepsilon_y \leq z - U_y\right) \\ &= \mathbf{F}\left(z - U_1, ..., z - U_n\right) = \exp\left(-g\left(e^{U_1 - z}, ..., e^{U_n - z}\right)\right) \\ &= \exp\left(-e^{-z} g\left(e^{U_1}, ..., e^{U_n}\right)\right) = \varphi\left(z - \log g\left(e^{U_1}, ..., e^{U_n}\right) - \gamma\right) \end{aligned}$$

where $\varphi(z) := \exp\left(-e^{-(z-\gamma)}\right)$ is the cdf of the standard Gumbel distribution. Hence Z has the distribution of $\log g\left(e^{U_1}, ..., e^{U_n}\right) + \gamma + \epsilon$, where ϵ is a standard Gumbel. $\square$

- As a result, the choice probability of alternative y is

$$\sigma_y(U) = \frac{\frac{\partial g}{\partial x_y}\left(e^{U_1}, ..., e^{U_n}\right)}{g\left(e^{U_1}, ..., e^{U_n}\right)} e^{U_y}.$$

- The GEV framework has several commonly used examples: logit, nested logit, mixture of logit...
- We just saw the logit model, in which $g(x_1, ..., x_n) = e^{-\gamma} \sum_{y=1}^{n} x_y$. In this case, the distribution of

$$Z = \max_{y=1,...,n} \left\{U_y + \varepsilon_y\right\}$$

is $\log \sum_{y=1}^{n} e^{U_y} + \epsilon$, where ϵ is a standard Gumbel.

- The nested logit model is an instance of GEV model where alternatives can be grouped in nests. Eg, people choose their means of transportation (nest), and within this nest, a particular operator.
- Let $\mathcal{X}$ be the set of nests and assume that for each nest x, there is a set $\mathcal{Y}_x$ alternatives. Let U_{xy} be utility from alternative y in nest x, and $\lambda_x \in [0,1]$ and

$$g\left(U_{xy}\right) = e^{-\gamma} \sum_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}_x} U_{xy}^{1/\lambda_x} \right)^{\lambda_x}.$$

- In this case,

$$G\left(U\right) = \mathbb{E}\left[\max_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}_x} \left\{U_{xy} + \varepsilon_{xy}\right\}\right] = \log \sum_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}_x} e^{U_{xy}/\lambda_x} \right)^{\lambda_x}$$

$$\sigma_{xy}\left(U\right) = \frac{\left(\sum_{y \in \mathcal{Y}_x} e^{U_{xy}/\lambda_x}\right)^{\lambda_x}}{\sum_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}_x} e^{U_{xy}/\lambda_x}\right)^{\lambda_x}} \frac{e^{U_{xy}/\lambda_x}}{\left(\sum_{y \in \mathcal{Y}_x} e^{U_{xy}/\lambda_x}\right)}$$

so the demand map has an interesting interpretation as "choice of nest then choice of alternative".

- The entropy of choice G^* in the nested logit model is given by

$$G^*(s) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_x} \lambda_x s_{xy} \ln s_{xy} + \sum_{x\in\mathcal{X}} (1-\lambda_x) \left(\sum_{z\in\mathcal{Y}_x} s_{xz} \right) \ln \left(\sum_{z\in\mathcal{Y}_x} s_{xz} \right) \quad (6)$$

  if $s_{xy} \geq 0$ and $\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_x} s_{xy} = 1$, $G^*(s) = +\infty$ otherwise.
- Identification in the nested logit model: with normalization $\sum_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}_x} e^{U_{xy}/\lambda_x} \right)^{\lambda_x} = 1$, one has $s_{xy} = \left(\sum_{y\in\mathcal{Y}_x} e^{U_{xy}/\lambda_x} \right)^{\lambda_x - 1} e^{U_{xy}/\lambda_x}$, thus $\sum_{y\in\mathcal{Y}_x} e^{U_{xy}/\lambda_x} = \left(\sum_{y\in\mathcal{Y}_x} s_{xy} \right)^{1/\lambda_x}$, therefore

$$U_{xy} = \lambda_x \log s_{xy} - (\lambda_x - 1) \log \sum_{y\in\mathcal{Y}_x} s_{xy}.$$

- The data (‘`1a-appli-travelmode`’) is taken from Greene and Hensher (1997). 210 individuals are surveyed about their choice of travel mode between Sydney, Canberra and Melbourne, and the various costs (time and money) associated with each alternative. Therefore there are $840 = 4 \times 210$ observations, which we can stack into ‘`travelmodedataset`’ a 3 dimensional array whose dimensions are mode,individual,dummy for choice+covariates.
- First, we compute the unconditional market shares:
  ```
  s = apply(X = travelmodedataset[,,3],FUN = mean, MARGIN = 1)
  ```
  which yields:
  ```
  air        train      bus        car
  0.2761905  0.3000000  0.1428571  0.2809524
  ```

- Define "car" as the default alternative. The utilities in the logit model are obtained by the log-odds ratio formula:
  ```
  Ulogit = log(s[1:4]/s[4])
  ```
  which yields
  ```
   air          train       bus          car
   -0.01709443  0.06559728  -0.67634006  0.00000000
  ```
- Now compute these utilities using a nested logit model with two nests, "noncar" and "car", and taking $\lambda = 0.5$ in both nests. Do:
  ```
  Unocar = lambda[1]*log(s[1:3])+(1-lambda[1]) *
  log(sum(s[1:3]))
  Ucar = lambda[2]*log(s[4])+(1-lambda[2]) * log(sum(s[4]))
  Unested = c(Unocar,Ucar ) - Ucar
  ```
  which yields
  ```
   air        train      bus        car
   0.4613240  0.5026698  0.1317012  0.00000000
  ```
- We see how correlation whithin nests impacts the estimation of the systematic utilities. Why?

- The GEV models are convenient analytically, but not very flexible.
  - The logit model imposes zero correlation across alternatives
  - The nested logit allows for nonzero correlation, but in a very rigid way (needs to define nests).
- A good example is the probit model, where ε is a Gaussian vector. For this model, there is no close-form solution neither for G nor for G^*.
- More recently, a number of modern models don't have closed-form either. These models require simulation methods in order to approximate them by discrete models.

- The pure characteristics model (Berry and Pakes, 2007) can be motivated as follows. Assume y stands for the number of bedrooms. The logit model would assume that the random utility associated with a 2-BR is uncorrelated with a 3-BR, which is not realistic.
- Let ξ_y is the typical size of a bedroom of size y, one may introduce ϵ as the valuation of size; in which case the utility shock associated with y should be $\varepsilon_y = \epsilon \xi_y$. More generally, the characteristics ξ_y is a d-dimensional (deterministic) vector, and $\epsilon \sim \mathbf{P}_\epsilon$ is a (random) vector of the same size standing for the valuations of the respective dimensions, so that
$$\varepsilon_y = \epsilon^\intercal \xi_y.$$
- For example, if each alternative y stands for a model of car, the first component of ξ_y may be the price of car y; the other components may be other characteristics such as number of seats, fuel efficiency, size, etc. In that case, for a given dimension $y \in \mathcal{Y}_0$, ϵ_y is the (random) valuation of this dimension by the consumer with taste vector ϵ.

- Assume without loss of generality that $\varepsilon_y = 0$, that is $\xi_0 = 0$ as we can always reduce the setting to this case by replacing ξ_y by $\xi_y - \xi_0$.
- Letting Z be the $|\mathcal{Y}| \times d$ matrix of (y, k)-term ξ_y^k, this rewrites as

$$\varepsilon = Z\epsilon.$$

- Hence, we have

$$G(U) = \mathbb{E}\left[\max\left\{U + Z\epsilon, 0\right\}\right].$$

and

$$\sigma_y(U) = \Pr\left(U_y - U_z \geq (Z\epsilon)_y - (Z\epsilon)_z \ \ \forall z \in \mathcal{Y}_0 \setminus \{y\}\right).$$

- When $d = 1$ (scalar characteristics), one has $\sigma_y(U) = \Pr\left(U_y - U_z \geq (\xi_y - \xi_z)\epsilon \; \forall z \in \mathcal{Y}_0 \backslash \{y\}\right)$, and thus

$$\sigma_y(U) = \Pr\left(\max_{z:\xi_y > \xi_z} \left\{\frac{U_y - U_z}{\xi_y - \xi_z}\right\} \leq \epsilon \leq \min_{z:\xi_y < \xi_z} \left\{\frac{U_y - U_z}{\xi_y - \xi_z}\right\}\right)$$

  with the understanding that $\max_{z \in \emptyset} f_z = -\infty$ and $\min_{z \in \emptyset} f_z = +\infty$.
- Therefore, letting $\mathbf{F}_\epsilon$ be the c.d.f. associated with the distribution of ϵ, one has a closed-form expression for σ_y:

$$\sigma_y(U) = \mathbf{F}_\epsilon\left(\left[\max_{z:\xi_y > \xi_z} \left\{\frac{U_y - U_z}{\xi_y - \xi_z}\right\}, \min_{z:\xi_y < \xi_z} \left\{\frac{U_y - U_z}{\xi_y - \xi_z}\right\}\right]\right)$$

- When $\mathbf{P}_\epsilon$ is the $\mathcal{N}(0, S)$ distribution, then the pure characteristics model is called a Probit model; in this case,

$$\varepsilon \sim \mathcal{N}(0, \Sigma) \text{ where } \Sigma = ZSZ^\intercal.$$

- Note the distribution ε will not have full support unless $d \geq |\mathcal{Y}|$ and Z is of full rank.
- Computing σ in the Probit model thus implies computing the mass assigned by the Gaussian distribution to rectangles of the type

$$[l_y, u_y].$$

When Σ is diagonal (random utility terms are i.i.d. across alternatives), this is numerically easy. However, this is computationally difficult in general (more on this later).

- In a number of cases, one cannot compute the choice probabilities $\sigma(U)$ using a closed-form expression. In this case, we need to resort to simulation to compute G, G^*, σ and σ^{-1}.
- The idea is that:
  - one is able to compute G and G^* for discrete distributions (more on this later)
  - the sampled versions of G, G^*, σ and σ^{-1} converge to the populations objects when the sample size is large.

- McFadden's smoothed accept-reject simulator (SARS) consists in sampling $\varepsilon \sim P$: $\varepsilon^1, ..., \varepsilon^N$, and replacing the max by the smooth-max

$$\sigma_{N,T,y}(U) = \sum_{i=1}^{N} \frac{1}{N} \frac{\exp\left(\left(U_y + \varepsilon_y^i\right)/T\right)}{\sum_z \exp\left(\left(U_z + \varepsilon_z^i\right)/T\right)}$$

- One seeks U so that the induced choice probabilities are s, that is

$$s_y = \sum_{i=1}^{N} \frac{1}{N} \frac{\exp\left(\left(U_y + \varepsilon_y^i\right)/T\right)}{\sum_z \exp\left(\left(U_z + \varepsilon_z^i\right)/T\right)}.$$

- The associated Emax operator is

$$G_{N,T}(U) = \mathbb{E}_{\mathbf{P}_N}\left[G_{\text{logit}}\left(U + \varepsilon^i\right)\right]$$

so the underlying random utility structure is a random coefficient logit.

- The random coefficient logit model (Berry, Levinsohn and Pakes, 1995) may be viewed as an interpolant between the random characteristics model and the logit model. In this case,

$$\varepsilon = (1-\lambda) Z\epsilon + \lambda\eta$$

where $\epsilon \sim \mathbf{P}_\epsilon$, η is an EV1 distribution independent from the previous term, and λ is a interpolation parameter ($\lambda = 1$ is the logit model, and $\lambda = 0$ is the pure characteristics model).

- In this case, one may compute the Emax operator as

$$\begin{aligned} G(U) &= \mathbb{E}\left[\max_{y\in\mathcal{Y}_0}\left\{U_y + (1-\lambda)(Z\epsilon)_y + \lambda\eta_y\right\}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{Y}_0}\left\{U_y + (1-\lambda)(Z\epsilon)_y + \lambda\eta_y\right\}|\epsilon\right]\right] \\ &= \mathbb{E}\left[\lambda\log\sum_{y\in\mathcal{Y}_0}\exp\left(\frac{U_y + (1-\lambda)(Z\epsilon)_y}{\lambda}\right)\right] \end{aligned}$$

- Recall

$$G\left(U\right)=\mathbb{E}\left[\lambda\log\sum_{y\in\mathcal{Y}_0}\exp\left(\frac{U_y+\left(1-\lambda\right)\left(Z\epsilon\right)_y}{\lambda}\right)\right].$$

- The demand map in the random coefficients logit model is obtained by derivation of the expression of the Emax, i.e.

$$\sigma_y\left(U\right)=\mathbb{E}\left[\frac{\exp\left(\frac{U_y+(1-\lambda)(Z\epsilon)_y}{\lambda}\right)}{\sum_{y'\in\mathcal{Y}_0}\exp\left(\frac{U_{y'}+(1-\lambda)(Z\epsilon)_{y'}}{\lambda}\right)}\right].$$

- Cf. Bonnet, G. and Shum (2017). Let $u_i = T \log \sum_z \exp\left(\left(U_z + \varepsilon_z^i \right) / T \right)$. One has

$$\begin{cases} s_y = \sum_{i=1}^N \frac{1}{N} \exp\left(\left(U_y - u_i + \varepsilon_y^i \right) / T \right) \\ \frac{1}{N} = \sum_y \frac{1}{N} \exp\left(\left(U_y - u_i + \varepsilon_y^i \right) / T \right) \end{cases}.$$

- As a result, (u_i, U_y) are the solution of the regularized OT problem

$$\min_{u,U} \sum_{i=1}^N \frac{1}{N} u_i - \sum s_y U_y + \sum_{i,y} \frac{1}{N} \exp\left(\left(U_y - u_i + \varepsilon_y^i \right) / T \right).$$

- Consider the IPFP algorithm for solving the latter problem:

$$\begin{cases} \exp\left(u_i^{k+1}/T\right) = \sum_z \exp\left(\left(U_z^k + \varepsilon_z^i\right)/T\right) \\ \exp U_y^{k+1}/T = \frac{Ns_y}{\sum_{i=1}^N \exp\left(\left(-u_i^{k+1}+\varepsilon_y^i\right)/T\right)} \end{cases}$$

- This rewrites as

$$\exp U_y^{k+1}/T = \frac{Ns_y}{\sum_{i=1}^N \frac{\exp\left(\varepsilon_y^i/T\right)}{\sum_z \exp\left(\left(U_z^k+\varepsilon_z^i\right)/T\right)}}, \text{ i.e.}$$

$$U_y^{k+1} = T\log s_y - T\log \sum_{i=1}^N \frac{1}{N} \frac{\exp\left(\varepsilon_y^i/T\right)}{\sum_z \exp\left(\left(U_z^k + \varepsilon_z^i\right)/T\right)}$$

which is the contraction mapping algorithm of Berry, Levinsohn and Pakes (1995, appendix 1).

Section 4

The inversion theorem

THEOREM (G AND SALANIÉ)

Consider a solution $\left(u\left(\varepsilon\right), v_y\right)$ to the dual Monge-Kantorovich problem with cost $\Phi\left(\varepsilon, y\right) = \varepsilon_y$, that is:

$$\min_{u,v} \int u\left(\varepsilon\right) d\mathbf{P}\left(\varepsilon\right) + \sum_{y\in\mathcal{Y}_0} v_y s_y \tag{7}$$
$$s.t.\ u\left(\varepsilon\right) + v_y \geq \Phi\left(\varepsilon, y\right)$$

Then:
(i) $U = \sigma^{-1}\left(s\right)$ is given by $U_y = v_0 - v_y$.
(ii) The value of Problem (7) is $-G^*\left(s\right)$.

PROOF.

$\sigma^{-1}(s) = \arg\max_{U:U_0=0} \left\{ \sum_{y\in\mathcal{Y}} s_y U_y - G(U) \right\}$, thus, letting $v = -U$, v is the solution ot

$$\min_{v:v_0=0} \left\{ \sum_{y\in\mathcal{Y}_0} s_y v_y + G(-v) \right\}$$

which is exactly problem (7). □

- It follows from the inversion theorem that the problem of demand inversion in the pure characteristics model is a semi-discrete transport problem, a point made in Bonnet, G and Shum (2017).
- Indeed, the correspondence is:
    - an alternative y is a fountain
    - the characteristics of an alternative is a fountain location
    - the systematic utility associated with alternative y is minus the price of fountain y
    - the market share of altenative y coindides with the capacity of fountain y
    - the random vector ϵ is the location of an inhabitant

- Cf. Chiong, G, and Shum (2016). Cf. Bonnet, G., O'Hara and Shum (2017). Simulated models are identified by

$$\max \sum_{i,y} \pi_{iy} \varepsilon_{iy}$$

$$s.t. \left\{ \begin{array}{l} \sum_{i=1}^{N} \pi_{iy} = s_y \\ \sum_y \pi_{iy} = \frac{1}{N} \end{array} \right. .$$

- As a result, (u_i, U_y) are the solution of the dual problem

$$\min_{u,U} \sum_{i=1}^{N} \frac{1}{N} u_i - \sum s_y U_y$$

$$s.t.\ u_i - U_y \geq \varepsilon_{iy}$$

$$U_0 = 0$$

- We shall code the AR simulator for the probit model and then invert it using the inversion theorem.
- Take a vector of systematic utilities:
```
U_y = c(1.6, 3.2, 1.1,0)
```
- Simulate the market shares using the AR simulator:
```
epsilon_iy = matrix(rnorm(nbDraws*nbY),ncol=nbY) %*%
SqrtCovar
u_iy = t(t(epsilon_iy)+U_y)
ui = apply(X = u_iy, MARGIN = 1, FUN = max)
s_y = apply(X = u_iy - ui, MARGIN = 2,FUN = function(v)
(length(which(v==0)))) / nbDraws
```

- To invert the market share, simply run the optimal assignment problem:

```
A1 =
kronecker(matrix(1,1,nbY),sparseMatrix(1:nbDraws,1:nbDraws))
A2 =
kronecker(sparseMatrix(1:nbY,1:nbY),matrix(1,1,nbDraws))
A = rbind2(A1,A2)
result = gurobi (
list(A=A,obj=c(epsilon_iy),modelsense="max",
rhs=c(rep(1/nbDraws,nbDraws),s_y) ,sense="="),
params=list(OutputFlag=0) )
Uhat_y = - result$pi[(1+nbDraws):(nbY+nbDraws)] +
result$pi[(nbY+nbDraws)]
```

- Cf. Bonnet, G., O'Hara and Shum (2017). Let $u_i = T \log \sum_z \exp\left((U_z + \varepsilon_z^i) / T \right)$. One has

$$\begin{cases} s_y = \sum_{i=1}^{N} \frac{1}{N} \exp\left((U_y - u_i + \varepsilon_y^i) / T \right) \\ \frac{1}{N} = \sum_y \frac{1}{N} \exp\left((U_y - u_i + \varepsilon_y^i) / T \right) \end{cases}.$$

- As a result, (u_i, U_y) are the solution of the regularized OT problem

$$\min_{u,U} \sum_{i=1}^{N} \frac{1}{N} u_i - \sum s_y U_y + \sum_{i,y} \frac{1}{N} \exp\left((U_y - u_i + \varepsilon_y^i) / T \right).$$

- Consider the IPFP algorithm for solving the latter problem:

$$\begin{cases} \exp\left(u_i^{k+1}/T\right) = \sum_z \exp\left(\left(U_z^k + \varepsilon_z^i\right)/T\right) \\ \exp U_y^{k+1}/T = \frac{Ns_y}{\sum_{i=1}^N \exp\left(\left(-u_i^{k+1} + \varepsilon_y^i\right)/T\right)} \end{cases}$$

- This rewrites as

$$\exp U_y^{k+1}/T = \frac{Ns_y}{\sum_{i=1}^N \frac{\exp\left(\varepsilon_y^i/T\right)}{\sum_z \exp\left(\left(U_z^k + \varepsilon_z^i\right)/T\right)}}, \text{ i.e.}$$

$$U_y^{k+1} = T \log s_y - T \log \sum_{i=1}^N \frac{1}{N} \frac{\exp\left(\varepsilon_y^i/T\right)}{\sum_z \exp\left(\left(U_z^k + \varepsilon_z^i\right)/T\right)}$$

which is exactly the contraction mapping algorithm of Berry, Levinsohn and Pakes (1995, appendix 1).