

1 Day 1

Linear programming duality

The problem is

$$\begin{aligned} V_P = \max_{x \geq 0} \quad & x^\top c \\ \text{s.t.} \quad & Ax \leq d \end{aligned}$$

We would like to write this as

$$\max_{x \geq 0} x^\top c + F(d - Ax)$$

where $F(u) = 0$ if $u \geq 0$, $F(u) = -\infty$ otherwise. The simplest choice is

$$F(u) = \min_{y \geq 0} \{y^\top u\} = \min_{y_j \geq 0} \left\{ \sum_j y_j u_j \right\}.$$

Thus rewrite the problem as

$$\begin{aligned} V_P &= \max_{x \geq 0} x^\top c + \min_{y \geq 0} \{y^\top (d - Ax)\} \\ &= \max_{x \geq 0} \min_{y \geq 0} \{x^\top c + y^\top d - y^\top Ax\} \end{aligned}$$

By the minimax theorem, if there are feasible solutions, then

$$\begin{aligned} V_P &= \min_{y \geq 0} \max_{x \geq 0} \{x^\top c + y^\top d - y^\top Ax\} \\ &= \min_{y \geq 0} y^\top d + \max_{x \geq 0} \{x^\top c - y^\top Ax\} \\ &= \min_{y \geq 0} y^\top d + \max_{x \geq 0} \{x^\top c - x^\top A^\top y\} \\ &= \min_{y \geq 0} y^\top d + \max_{x \geq 0} \{x^\top (c - A^\top y)\} \end{aligned}$$

Now, we remark that

$$\begin{aligned} \max_{x \geq 0} \{x^\top (c - A^\top y)\} &= +\infty \text{ if } c_i > (A^\top y)_i \text{ for some } i \\ &= 0 \text{ if } c \leq A^\top y \end{aligned}$$

we have derived the dual problem

$$\begin{aligned} V_P = V_D := \min_{y \geq 0} \quad & y^\top d \\ \text{s.t. } & c \leq A^\top y \end{aligned}$$

Further, if x^* an optimal solution to primal and y^* an optimal solution to the dual problem, we have

$$(x^*)^\top (c - A^\top y^*) = 0$$

but $c \leq A^\top y^*$ and $x^* \geq 0$ therefore we have for each i

$$x_i^* (c_i - (A^\top y^*)_i) = 0$$

therefore

$$x_i^* > 0 \implies (A^\top y^*)_i = c_i$$

and similarly

$$y_j^* > 0 \implies (Ax^*)_j = d_j$$

This is **complementary slackness**.

1.1 Gradient matrix

∇_{ax} $a \in A, x \in Z$

$$(\nabla f)_{xy} = \sum_z \nabla_{(xy)z} f_z = f_y - f_x$$

Mass balance equation. For every z ,

$$\begin{aligned} q_z &= (\text{total mass arriving to } z \text{ from other nodes}) \\ &\quad - (\text{total mass departing from } z \text{ to other nodes}) \\ &= \sum_x \mu_{xz} - \sum_y \mu_{zy} \end{aligned}$$

This can be expressed in a matrix way! Indeed,

$$q_z = \sum_{a \in A} \nabla_{az} \mu_a$$

hence mass balance rewrites in a matrix notation as

$$q = \nabla^\top \mu.$$

By the way, if q satisfies mass balance, then

$$\begin{aligned} q_z &= \sum_x \mu_{xz} - \sum_y \mu_{zy} \\ \sum_{z \in Z} q_z &= \sum_{z \in Z} \sum_{x \in Z: xz \in A} \mu_{xz} - \sum_{z \in Z} \sum_{y: zy \in A} \mu_{zy} \\ &= \sum_{a \in A} \mu_a - \sum_{a \in A} \mu_a \\ &= 0 \end{aligned}$$

1.2 An equilibrium problem

Introduce p_z =price of the commodity at z .

Consider a trader operating on the arc xy .

trader's profit = $p_y - p_x - c_{xy}$

Assume that there is free entry of traders on any arc. Absence of rent implies that for any arc $xy \in A$ we have

$$p_y - p_x \leq c_{xy}$$

which can be written

$$\nabla p \leq c$$

No arbitrage condition.

Now assume at equilibrium, a quantity $\mu_{xy} > 0$ is shipped from x to y . This implies that the traders at arc xy break even, hence

$$p_y - p_x - c_{xy} = 0$$

To summarize, an equilibrium on the network is given by $(p_z)_{z \in Z}$ and $(\mu_{xy})_{xy \in A}$ such that:

- (i) balance of mass holds: $\nabla^\top \mu = q$
- (ii) no arbitrage holds: $\nabla p \leq c$.
- (iii) $\mu_{xy} > 0 \implies (\nabla p)_{xy} = p_y - p_x = c_{xy}$.

1.3 An optimal shippment problem

Consider the Soviet problem. They decide on μ_{xy} subject to mass balance $\nabla^\top \mu = q$.

In order to minimize costs, they try to achieve the minimum of $\sum_{xy \in A} \mu_{xy} c_{xy}$. Thus, they do:

$$\begin{aligned} \min_{\mu \geq 0} \quad & \mu^\top c \\ \text{s.t.} \quad & \nabla^\top \mu = q \quad [p] \end{aligned}$$

this is a linear programming problem (primal). Let's compute its dual.

$$\begin{aligned} \max_p \quad & p^\top q \\ \text{s.t.} \quad & \nabla p \leq c \quad [\mu \geq 0] \end{aligned}$$

Theorem: if μ and p are respectively solutions to the primal and the dual problems, then they also solve the equilibrium problem above. Indeed, $\nabla^\top \mu = q$ and $\nabla p \leq c$ are immediately satisfied, and by complementary slackness, $\mu_{xy} > 0$ implies

$$p_y - p_x = c_{xy}.$$

2 Day 2

Solve the central planner's problem

$$\begin{aligned} \max_{\pi_{xy} \geq 0} \quad & \sum_{xy} \pi_{xy} \Phi_{xy} \\ \text{s.t.} \quad & \sum_y \pi_{xy} = p_x \quad [u_x] \\ & \sum_x \pi_{xy} = q_y \quad [v_y] \end{aligned}$$

by the maxmin formulation

$$\begin{aligned} & \max_{\pi_{xy} \geq 0} \min_{u_x, v_y} \sum_{xy} \pi_{xy} \Phi_{xy} + \sum_x p_x u_x + \sum_y q_y v_y - \sum_{xy} \pi_{xy} (u_x + v_y) \\ = & \min_{u_x, v_y} \sum_x p_x u_x + \sum_y q_y v_y + \max_{\pi_{xy} \geq 0} \sum_{xy} \pi_{xy} (\Phi_{xy} - u_x - v_y) \end{aligned}$$

this is

$$\begin{aligned} \min_{u_x, v_y} \quad & \sum_x p_x u_x + \sum_y q_y v_y \\ \text{s.t.} \quad & u_x + v_y \geq \Phi_{xy} \quad [\pi_{xy} \geq 0] \end{aligned}$$

Remark: by complementary slackness, $\pi_{xy} > 0 \implies u_x + v_y = \Phi_{xy}$.

Remark 2: if (u, v) then $(u + c, v - c)$ is also a solution for any constant c .

Take a solution (u, v) of the dual problem. Then for any feasible solution we have

$$v_y \geq \max_x \{\Phi_{xy} - u_x\}$$

but for any optimal solution (u, v) we have

$$v_y = \max_x \{\Phi_{xy} - u_x\}.$$

But the same logic implies

$$u_x = \max_y \{\Phi_{xy} - v_y\}.$$

2.1 Interpretation 1: stable matching

(π_{xy}, u_x, v_y) is a stable matching if:

- $\pi \geq 0$ and $\sum_y \pi_{xy} = p_x$ and $\sum_x \pi_{xy} = q_y$
- $\forall x, y, u_x + v_y \geq \Phi_{xy}$
- If $u_x + v_y > \Phi_{xy} \implies \pi_{xy} = 0$; or in other words $\pi_{xy} > 0 \implies u_x + v_y = \Phi_{xy}$

If $u_x + v_y < \Phi_{xy}$ then xy would be a blocking pair, this should be ruled out

2.2 Interpretation 2: Wage equilibrium

Assume x is a worker and y is a firm, and interpret u_x as the wage of worker x , and v_y as the profit of firm y .

Then firm's problem is

$$\begin{aligned} v_y &= \max_x \{\Phi_{xy} - u_x\} \\ \pi_{xy} &> 0 \implies v_y = \Phi_{xy} - u_x \\ \pi &\geq 0, \sum_y \pi_{xy} = p_x, \sum_x \pi_{xy} = q_y \end{aligned}$$

Rewrite this as

$$\begin{aligned} v_y &\geq \Phi_{xy} - u_x \forall x, y \\ \pi_{xy} &> 0 \implies v_y = \Phi_{xy} - u_x \\ \pi &\geq 0, \sum_y \pi_{xy} = p_x, \sum_x \pi_{xy} = q_y \end{aligned}$$

2.3 Case with unassigned agents

Assume that people don't have to match, and if they remain unmatched they get utility 0

π_{x0} = mass of men of type x remaining unassigned

π_{0y} = mass of women of type y remaining unassigned

$$\begin{aligned} \sum_y \pi_{xy} + \pi_{x0} &= p_x \\ \sum_x \pi_{xy} + \pi_{0y} &= q_y \end{aligned}$$

Optimal assignment problem

$$\begin{aligned} &\max_{\pi \geq 0} \sum_{xy} \pi_{xy} \Phi_{xy} \\ s.t. \quad &\sum_y \pi_{xy} + \pi_{x0} = p_x \\ &\sum_x \pi_{xy} + \pi_{0y} = q_y \end{aligned}$$

rewrite this as

$$\begin{aligned} &\max_{\pi \geq 0} \sum_{xy} \pi_{xy} \Phi_{xy} \\ s.t. \quad &\sum_y \pi_{xy} \leq p_x \quad [u_x \geq 0] \\ &\sum_x \pi_{xy} \leq q_y \quad [v_y \geq 0] \end{aligned}$$

whose dual is

$$\begin{aligned} \min_{u_x \geq 0, v_y \geq 0} \quad & \sum_x p_x u_x + \sum_y q_y v_y \\ \text{s.t.} \quad & u_x + v_y \geq \Phi_{xy} \quad [\pi_{xy} \geq 0] \end{aligned}$$

3 Day 3

Compute primal problem

$$\begin{aligned} \max_{\pi \geq 0} \quad & \left\{ \sum_{ij} \pi_{ij} \Phi_{ij} - \sigma \sum_{ij} \pi_{ij} \log \pi_{ij} \right\} \\ \text{s.t.} \quad & \sum_j \pi_{ij} = p_i \\ & \sum_i \pi_{ij} = q_j \end{aligned}$$

Write Lagrangian

$$\begin{aligned} \max_{\pi \geq 0} \left\{ \sum_{ij} \pi_{ij} \Phi_{ij} - \sigma \sum_{ij} \pi_{ij} \log \pi_{ij} \right\} + \min_{u_i, v_j} \left\{ \sum_i u_i \left(p_i - \sum_j \pi_{ij} \right) + \sum_j v_j \left(q_j - \sum_i \pi_{ij} \right) \right\} \\ \min_{u_i, v_j} \sum_i p_i u_i + \sum_j q_j v_j + \max_{\pi \geq 0} \left\{ \sum_{ij} \pi_{ij} (\Phi_{ij} - u_i - v_j) - \sigma \sum_{ij} \pi_{ij} \log \pi_{ij} \right\} \end{aligned}$$

FOC in the inner problem. We have

$$\Phi_{ij} - u_i - v_j = \sigma (1 + \log \pi_{ij})$$

that is

$$\pi_{ij} = \Pi_{ij}(u_i, v_j) := \exp \left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma} \right)$$

and we have

$$\pi_{ij} (\Phi_{ij} - u_i - v_j) = \sigma \pi_{ij} + \sigma \pi_{ij} \log \pi_{ij}$$

therefore

$$\pi_{ij} (\Phi_{ij} - u_i - v_j) - \sigma \pi_{ij} \log \pi_{ij} = \sigma \pi_{ij}$$

hence the previous problem becomes

$$\min_{u_i, v_j} \sum_i p_i u_i + \sum_j q_j v_j + \sum_{ij} \sigma \Pi_{ij}(u_i, v_j)$$

where $\Pi_{ij}(u_i, v_j) := \exp\left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma}\right)$, that is

$$\min_{u_i, v_j} F(u, v) := \sum_i p_i u_i + \sum_j q_j v_j + \sum_{ij} \sigma \exp\left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma}\right)$$

F is smooth and convex but ** not ** strictly convex because

$$F(u + c, v - c) = F(u, v)$$

Consider the problem of

$$\min_{(\theta_i)_{1 \leq i \leq n}} F(\theta)$$

1. Gradient descent / tatonnement

$$\theta^{t+1} = \theta^t - \epsilon \nabla F(\theta^t)$$

2. Newton descent

$$\theta^{t+1} = \theta^t - \epsilon (D^2 F(\theta^t))^{-1} \nabla F(\theta^t)$$

3. Coordinate descent.

For each t

For each i

Fix θ_j^t for $j \neq i$ and consider the problem

$$\min_{\theta_i \in \mathbb{R}} F(\theta_i, \theta_{-i}^t)$$

and call it θ_i^t

4. [later on] proximal gradient descent – will talk about it later

3.1 Gradient descent for regularized OT

$$F(u, v) = \sum_i p_i u_i + \sum_j q_j v_j + \sigma \sum_{ij} \exp\left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma}\right)$$

$$\frac{\partial F}{\partial u_i}(u, v) = p_i - \sum_j \underbrace{\exp\left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma}\right)}_{\pi_{ij}}$$

$$\frac{\partial F}{\partial v_j}(u, v) = q_j - \sum_i \underbrace{\exp\left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma}\right)}_{\pi_{ij}}$$

therefore (u, v) is optimal iff

$$\begin{aligned}\sum_j \Pi_{ij}(u_i, v_j) &= p_i \\ \sum_i \Pi_{ij}(u_i, v_j) &= q_j\end{aligned}$$

Gradient descent:

$$\begin{aligned}u_i^{t+1} &= u_i^t + \epsilon \left(\sum_j \exp \left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma} \right) - p_i \right) \\ v_j^{t+1} &= v_j^t + \epsilon \left(\sum_i \exp \left(\frac{\Phi_{ij} - u_i - v_j - \sigma}{\sigma} \right) - q_j \right)\end{aligned}$$

Coordinate descent.

We have u_i^{t+1} such that

$$\frac{\partial F((u_i^{t+1}; u_{-i}^t), v^t)}{\partial u_i^{t+1}} = 0$$

that is

$$p_i = \sum_j \exp \left(\frac{\Phi_{ij} - u_i^{t+1} - v_j^t - \sigma}{\sigma} \right)$$

Introduce $A_i = \exp(-u_i/\sigma)$ and $B_j = \exp(-v_j/\sigma)$, and

$$K_{ij} = \exp \left(\frac{\Phi_{ij} - \sigma}{\sigma} \right)$$

we can rewrite the algorithm as

$$p_i = \sum_j K_{ij} A_i^{t+1} B_j^t,$$

thus

$$A_i^{t+1} = \frac{1}{\sum_j K_{ij} B_j^t}$$

Similarly, optimality wrt v_j^{t+1} / B_j^{t+1} yields

$$B_j^{t+1} = \frac{1}{\sum_i K_{ij} A_i^{t+1}}.$$

This is the IPFP algorithm / Sinkhorn's algorithm.

Solution when $\Phi = 0$:

$$\pi_{ij} = p_i q_j$$

3.2 The log-sum-exp trick

We have that

$$\sigma \log \left(\exp \left(\frac{a}{\sigma} \right) + \exp \left(\frac{b}{\sigma} \right) \right) \rightarrow_{\sigma \rightarrow 0^+} \max \{a, b\}$$

because of this $\sigma \log \left(\exp \left(\frac{a}{\sigma} \right) + \exp \left(\frac{b}{\sigma} \right) \right)$ is called smooth-max of a and b .

The idea is that for any $c \in \mathbb{R}$

$$\sigma \log \left(\exp \left(\frac{a+c}{\sigma} \right) + \exp \left(\frac{b+c}{\sigma} \right) \right) = c + \sigma \log \left(\exp \left(\frac{a}{\sigma} \right) + \exp \left(\frac{b}{\sigma} \right) \right)$$

Take $c = -\max \{a, b\}$ will get

$$\begin{aligned} & \sigma \log \left(\exp \left(\frac{a}{\sigma} \right) + \exp \left(\frac{b}{\sigma} \right) \right) - \max \{a, b\} \\ = & \sigma \log \left(\exp \left(\frac{\min \{0, a-b\}}{\sigma} \right) + \exp \left(\frac{\min \{0, b-a\}}{\sigma} \right) \right) \end{aligned}$$

But we have

$$0 \leq \sigma \log \left(\exp \left(\frac{\min \{0, a-b\}}{\sigma} \right) + \exp \left(\frac{\min \{0, b-a\}}{\sigma} \right) \right) \leq \sigma \log 2$$

For practical purposes, we will use

$$\sigma \log \left(\exp \left(\frac{a}{\sigma} \right) + \exp \left(\frac{b}{\sigma} \right) \right) = \sigma \log \left(\exp \left(\frac{a+c}{\sigma} \right) + \exp \left(\frac{b+c}{\sigma} \right) \right) - c$$

with $c = -\max \{a, b\}$, thus

$$\begin{aligned} & \sigma \log \left(\exp \left(\frac{a}{\sigma} \right) + \exp \left(\frac{b}{\sigma} \right) \right) \\ = & \max \{a, b\} + \sigma \log \left(\exp \left(\frac{\min \{0, a-b\}}{\sigma} \right) + \exp \left(\frac{\min \{0, b-a\}}{\sigma} \right) \right) \end{aligned}$$

Back to the IPFP algorithm. We had

$$\begin{aligned} u_i^{t+1} &= -\sigma \log \left(\frac{1}{p_i} \sum_j \exp \left(\frac{\Phi_{ij} - v_j^t - \sigma}{\sigma} \right) \right) \\ u_i^{t+1} &= \sigma \log p_i - \sigma \log \left(\sum_j \exp \left(\frac{\Phi_{ij} - v_j^t - \sigma}{\sigma} \right) \right) \end{aligned}$$

3.3 Discrete choice

Consider

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

If the distribution of (ε_y) has a density, then

$$\sum_{y \in \mathcal{Y}_0} \sigma_y(U) = 1$$

When the distribution of (ε_y) has a density, then $\Pr(U_y + \varepsilon_y = U_z + \varepsilon_z) = 0$ for $y \neq z$, therefore

$$\sigma_y(U) = \Pr(U_y + \varepsilon_y > U_z + \varepsilon_z \forall z \in \mathcal{Y}_0 \setminus \{y\})$$

and we have

$$\sum_{y \in \mathcal{Y}_0} \Pr(U_y + \varepsilon_y > U_z + \varepsilon_z \forall z \in \mathcal{Y}_0 \setminus \{y\}) \leq 1 \leq \sum_{y \in \mathcal{Y}_0} \sigma_y(U)$$

For instance when $\mathcal{Y}_0 = \{0, 1\}$

$$\Pr(U_1 + \varepsilon_1 > U_0 + \varepsilon_0) + \Pr(U_0 + \varepsilon_0 \geq U_1 + \varepsilon_1) = 1$$

but then

$$\Pr(U_1 + \varepsilon_1 \geq U_0 + \varepsilon_0) + \Pr(U_0 + \varepsilon_0 \geq U_1 + \varepsilon_1) = 1$$

3.4 Daly-Zachary-Williams

Compute the overall indirect utility of the consumers (social welfare). It is

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

This is called the Emax operator. It is a convex function. Let's see how a change in U_y affects the social welfare. This is given by

$$\begin{aligned} \frac{\partial G}{\partial U_y}(U) &= E \left[\frac{\partial}{\partial U_y} \max_{z \in \mathcal{Y}} \{U_z + \varepsilon_z, \varepsilon_0\} \right] \\ &= E \left[1 \left\{ y \in \arg \max_{z \in \mathcal{Y}} \{U_z + \varepsilon_z, \varepsilon_0\} \right\} \right] \\ &= \sigma_y(U) \end{aligned}$$

3.5 The Logit model

If (ε_y) are iid Gumbel ie if their joint cdf

$$\begin{aligned} F_\varepsilon(a) &= \Pr(\varepsilon_y \leq a_y \forall y) = \prod_{y \in \mathcal{Y}} \exp(-\exp(-a_y + \gamma)) \\ &= \exp\left(-e^\gamma \sum_{y \in \mathcal{Y}} e^{-a_y}\right) \end{aligned}$$

then

Proposition: One has

$$Z = \max_y \{U_y + \varepsilon_y\} =_D \log \sum_y \exp(U_y) + \varepsilon$$

therefore

$$\begin{aligned} \max_y \{U_y + \sigma \varepsilon_y\} &= \sigma \max_y \left\{ \frac{U_y}{\sigma} + \varepsilon_y \right\} \\ &= {}_D \sigma \log \sum_y \exp\left(\frac{U_y}{\sigma}\right) + \sigma \varepsilon \end{aligned}$$

Proof of the proposition. Let's compute the c.d.f. of Z . We have

$$\begin{aligned} F_Z(z) &= \Pr(Z \leq z) = \Pr\left(\max_y \{U_y + \varepsilon_y\} \leq z\right) = \Pr(U_y + \varepsilon_y \leq z, \forall y) \\ &= \Pr(\varepsilon_y \leq z - U_y : \forall y) = \exp\left(-e^\gamma \sum_{y \in \mathcal{Y}} e^{U_y - z}\right) \end{aligned}$$

Now let's compute the cdf of $Z' = \log \sum_y \exp(U_y) + \varepsilon$, where ε is a Gumbel. We have

$$\begin{aligned} F_{Z'}(z) &= \Pr\left(\log \sum_y \exp(U_y) + \varepsilon \leq z\right) = \Pr\left(\varepsilon \leq z - \log \sum_y \exp(U_y)\right) \\ &= \exp\left(-\exp\left(\log \sum_y \exp(U_y) - z + \gamma\right)\right) = \exp\left(-e^{-\gamma} \sum_{y \in \mathcal{Y}} e^{U_y - z}\right) \end{aligned}$$

Thus these two cdfs are the same!

As a result, we have that in the logit model,

$$G(U) = \log \sum_{y \in \mathcal{Y}_0} \exp(U_y)$$

and if we assume $U_0 = 0$, we will get

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

Let's deduce the market shares (choice probabilities) in the logit model. We have

$$\begin{aligned} \sigma_y(U) &= \frac{\partial}{\partial U_y} \log \left(\sum_{y \in \mathcal{Y}_0} \exp(U_y) \right) \\ &= \frac{\exp(U_y)}{\sum_{y \in \mathcal{Y}_0} \exp(U_y)} \end{aligned}$$

and if $U_0 = 0$,

$$\sigma_y(U) = \frac{\exp(U_y)}{1 + \sum_{y \in \mathcal{Y}} \exp(U_y)} \text{ and } \sigma_0(U) = \frac{1}{1 + \sum_{y \in \mathcal{Y}} \exp(U_y)}$$

which is Gibbs distribution.

3.6 Extending the logit model

Let's assume that η_i for $i = 1, \dots, n$ are i.i.d. Gumbel random variables. We would like to combine the η_i in order to create a model for some ε_y . We will take

$$\varepsilon_y = \max_i \{a_{iy} + \eta_i\} - \log \sum_i \exp(a_{iy}) \quad (1)$$

We have that the each of the ε_y is a Gumbel random variable, BUT they are not independent. Let's compute their c.d.f.

We have

$$\begin{aligned} F_\varepsilon(c) &= \Pr(\varepsilon_y \leq c_y \forall y) \\ &= \Pr \left(\max_i \{a_{iy} + \eta_i\} - \log \sum_i \exp(a_{iy}) \leq c_y, \forall y \right) \\ &= \Pr \left(a_{iy} + \eta_i \leq c_y + \log \sum_i \exp(a_{iy}), \forall y, \forall i \right) \\ &= \Pr \left(\eta_i \leq c_y - a_{iy} + \log \sum_i \exp(a_{iy}), \forall y, \forall i \right) \\ &= \Pr \left(\eta_i \leq c_y + \min_y \left\{ -a_{iy} + \log \sum_i \exp(a_{iy}) \right\}, \forall i \right) \\ &= \exp \left(-e^\gamma \sum_i e^{-c_y} e^{-\min_y \{ -a_{iy} + \log \sum_i \exp(a_{iy}) \}} \right) \end{aligned}$$

that is

$$\begin{aligned}
F_\varepsilon(c) &= \exp\left(-e^\gamma \sum_i e^{-c y} e^{-\min_y \{-a_{iy} + \log \sum_i \exp(a_{iy})\}}\right) \\
&= \exp\left(-e^\gamma \sum_i e^{-c y} e^{\max_y \{a_{iy} - \log \sum_i \exp(a_{iy})\}}\right) \\
&= \exp\left(-\sum_i e^\gamma \max_y \left\{e^{-c y} \frac{e^{a_{iy}}}{\sum_i \exp(a_{iy})}\right\}\right)
\end{aligned}$$

therefore, we get that

$$F_\varepsilon(c) = \exp(-g(e^{-a}))$$

where

$$g(b) = \sum_i e^\gamma \max_y \left\{b_y \frac{e^{a_{iy}}}{\sum_i \exp(a_{iy})}\right\}$$

We have that $g(b)$ is positive homogenous of degree one and is such that $\exp(-g(e^{-a}))$ is a c.d.f.

This is what is called the Generalized Extreme Value model of McFadden, also called the Multivariate Extreme Value model. The representation (1) is called Pickand's representation.

Definition. The distribution of ε belongs to the GEV distribution if there is a function $g(b)$ such that $g(b)$ is positive homogenous of degree one and is such that $\exp(-g(e^{-a}))$ is a c.d.f.

Theorem (McFadden 1978). If ε has a GEV distribution associated with homogeneous function g , one has

$$Z = \max_y \{U_y + \varepsilon_y\} =_D \log g(e^U) + \varepsilon$$

Proof of the theorem. Let's compute the c.d.f. of Z . We have

$$\begin{aligned}
F_Z(z) &= \Pr(Z \leq z) = \Pr\left(\max_y \{U_y + \varepsilon_y\} \leq z\right) = \Pr(U_y + \varepsilon_y \leq z, \forall y) \\
&= \Pr(\varepsilon_y \leq z - U_y : \forall y) = \exp(-g(e^{U_y - z})) \\
&= \exp(-g(e^{-z} e^{U_y})) \\
&= \exp(-e^{-z} g(e^U))
\end{aligned}$$

Now let's compute the cdf of $Z' = \log g(e^U) + \varepsilon$, where ε is a Gumbel. We have

$$\begin{aligned}
F_{Z'}(z) &= \Pr(\log g(e^U) + \varepsilon \leq z) = \Pr(\varepsilon \leq z - \log g(e^U)) \\
&= \exp(-e^{-z} g(e^U))
\end{aligned}$$

Thus these two cdfs are the same!

Consequence: We have a closed-form expression for G and σ_y which is

$$\begin{aligned} G(U) &= \log g(e^U) \\ \sigma_y(U) &= \frac{\partial_y g(e^U)}{g(e^U)} e^{U_y} \end{aligned}$$

3.7 Tomorrow

Random coefficient logit model of Berry Levinsohn Pakes.

$$\begin{aligned} U_y + \varepsilon_y \\ \varepsilon_y = \xi_y^\top \eta + \epsilon_y \end{aligned}$$

The inversion of this model is the problem of going from (s_y) to U_y

is an entropic regularized OT problem. More specifically,

$$\begin{aligned} \max_{\pi \geq 0} \quad & \sum_{iy} \pi_{iy} \underbrace{(\xi_y^\top \eta_i)}_{\Phi_{iy}} - \sum_{iy} \pi_{iy} \log \pi_{iy} \\ \text{s.t.} \quad & \sum_i \pi_{iy} = s_y \quad [U_y] \\ & \sum_y \pi_{iy} = \frac{1}{n} \end{aligned}$$

4 Day 4

An announcement:

Looking for RAing opportunity at the frontier of economics, ML and computation?

Equiprice is hiring!

$$\begin{aligned} \Phi(x, y) &= x^\top Ay \\ A &= \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix} \\ y' &= Ay \end{aligned}$$

4.1 Inverting demand maps

Recall from B06 that (DZW theorem)

$$\sigma(U) = \nabla G(U)$$

where $G(U) = E[\max\{U_y + \varepsilon_y\}]$.

Now we are looking for U such that

$$\sigma(U) = s,$$

that is

$$\nabla G(U) = s$$

This can be solved by a convex optimization problem.

If $s = 0$ this is $\nabla G(U) = 0$ hence that is the foc associated with $\min_U G(U)$.

In general, this is

$$G^*(s) = \max_U \{s^\top U - G(U)\}$$

hence

$$\sigma^{-1}(s) = \arg \max_U \{s^\top U - G(U)\}$$

and we have that

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

Example: logit model (with normalization $U_0 = 0$). $G(U) = \log \left(1 + \sum_{y \in \mathcal{Y}} e^{U_y} \right)$.

A straightfoward calculation shows that

$$G^*(s) = \sum_{y \in \mathcal{Y}} s_y \log s_y + s_0 \ln s_0$$

where $s_0 = 1 - \sum_{y \in \mathcal{Y}} s_y$. Cupid's paper Galichon-Salanie.

In this case,

$$\begin{aligned} U_y &= \sigma_y^{-1}(s) = \frac{\partial G^*}{\partial s_y}(s) = (1 + \log s_y) - (1 + \log s_0) \\ &= \log \frac{s_y}{s_0} \end{aligned}$$

4.2 More on G^*

In the general case. We have

$$\begin{aligned} G^*(s) &= \max_U \{s^\top U - G(U)\} \\ &= \max_{U: U_0=0} \left\{ \sum_{y \in \mathcal{Y}} s_y U_y - E \left[\max_{y \in \mathcal{Y}_0} \{U_y + \varepsilon_y\} \right] \right\} \\ -G^*(s) &= \min_{U: U_0=0} \left\{ - \sum_{y \in \mathcal{Y}} s_y U_y + E \left[\max_{y \in \mathcal{Y}_0} \{U_y + \varepsilon_y\} \right] \right\} \end{aligned}$$

Define $v_y = -U_y$, we have

$$-G^*(s) = \min_{(v_y): v_0=0} \left\{ \sum_{y \in \mathcal{Y}} s_y v_y + E \left[\max_{y \in \mathcal{Y}_0} \{-v_y + \varepsilon_y\} \right] \right\}$$

This can be recast as $-G^*(s) =$

$$\min_{\substack{(v_y): v_0=0 \\ u(\varepsilon)}} \left\{ \sum_{y \in \mathcal{Y}} s_y v_y + E[u(\varepsilon)] \right\}$$

$$u(\varepsilon) \geq -v_y + \varepsilon_y$$

equivalently

$$\min_{\substack{(v_y): v_0=0 \\ u(\varepsilon)}} \left\{ \sum_{y \in \mathcal{Y}} s_y v_y + E_P[u(\varepsilon)] \right\}$$

$$u(\varepsilon) + v_y \geq \varepsilon_y$$

In the discrete case, if we sample P , into $\varepsilon^1, \dots, \varepsilon^N$, we have, setting $u(\varepsilon^i) = u_i$

$$\min_{\substack{(v_y): v_0=0 \\ u_i}} \left\{ \frac{1}{N} \sum_{i=1}^N u_i + \sum_{y \in \mathcal{Y}} s_y v_y \right\}$$

$$u_i + v_y \geq \varepsilon_y^i$$

Theorem (Galichon-Salanie 2011). The $U_y = -v_y$ are identified by an optimal transport problem of matching

$\varepsilon \sim P$ (“distribution of firms”) and $Y \sim s$ (“distribution of workers”) in order to optimize total surplus $\Phi_{\varepsilon y} = \varepsilon_y$.

Primal version

$$\max_{(\varepsilon, Y)} E_\pi[\varepsilon_Y]$$

$$\sim \pi : \varepsilon \sim P, Y \sim s$$

and dual version

$$\min_{\substack{(v_y): v_0=0 \\ u_i}} \left\{ E_P[u(\varepsilon)] + \sum_{y \in \mathcal{Y}} s_y v_y \right\}$$

$$u(\varepsilon) + v_y \geq \varepsilon_y \quad \forall \varepsilon, y$$

Sampled version:

Primal version

$$\begin{aligned} \max_{\pi_{iy}} \quad & \sum_{iy} \pi_{iy} \varepsilon_y^i \\ \text{s.t.} \quad & \sum_{y \in \mathcal{Y}_0} \pi_{iy} = \frac{1}{N} \\ & \sum_i \pi_{iy} = s_y \end{aligned}$$

and the dual version

$$\begin{aligned} \min_{\substack{(v_y): v_0=0 \\ u_i}} \quad & \left\{ \frac{1}{N} \sum_{i=1}^N u_i + \sum_{y \in \mathcal{Y}} s_y v_y \right\} \\ & u_i + v_y \geq \varepsilon_y^i \end{aligned}$$

Probit example. $\varepsilon_{iy} = (Z\epsilon_i)_y$ where $\epsilon_i \sim N(0, I_K)$, Z is a $|\mathcal{Y}_0| \times K$ matrix $\text{cov}(\varepsilon_{iy}, \varepsilon_{iy'}) = (Z^\top Z)_{y,y'}$ in particular if we want to impose the covariance matrix Σ between alternatives, one should take $Z = \Sigma^{1/2}$.

Then we can identify the U_y by

$$\begin{aligned} \arg \min_{\substack{(U_y): v_0=0 \\ u_i}} \quad & \left\{ \frac{1}{N} \sum_{i=1}^N u_i - \sum_{y \in \mathcal{Y}} s_y U_y \right\} \\ & u_i - U_y \geq \sum_k Z_{yk} \epsilon_{ki} \end{aligned}$$

4.3 Random coefficient logit model

RCL model = pure characteristics + logit

in the sense that

Random coefficient logit model of Berry Levinsohn Pakes.

$$U_y + \varepsilon_y$$

$$\varepsilon_y = (Z\epsilon_i)_y + T\eta_y \text{ with } T > 0$$

where $\epsilon_i \sim \nu$ a fixed distribution over R^K

and Z is a $|\mathcal{Y}_0| \times K$ matrix.

and $(\eta_{iy})_y$ iid Gumbel

and ϵ_i and $(\eta_{iy})_y$ are independent.

We recall that

$$\sigma_y(U) = \frac{\partial G(U)}{\partial U_y}$$

where

$$G(U) = E \left[\max_{y \in \mathcal{Y}_0} \left\{ U_y + (Z\epsilon_i)_y + T\eta_y \right\} \right]$$

by the law of iterated expectations, we have

$$G(U) = E \left[E \left[\max_{y \in \mathcal{Y}_0} \{U_y + (Z\epsilon_i)_y + T\eta_y\} \mid \epsilon \right] \right]$$

by independence,

$$E \left[\max_{y \in \mathcal{Y}_0} \{U_y + (Z\epsilon_i)_y + T\eta_y\} \mid \epsilon \right] = T \log \left(1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{U_y + (Z\epsilon_i)_y}{T} \right) \right)$$

hnce

$$G(U) = E_\epsilon \left[T \log \left(1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{U_y + (Z\epsilon_i)_y}{T} \right) \right) \right]$$

hence

$$\sigma_y(U) = E_\epsilon \left[\frac{\exp \left(\frac{U_y + (Z\epsilon_i)_y}{T} \right)}{1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{U_y + (Z\epsilon_i)_y}{T} \right)} \right]$$

Now lets compute the inversion of the model. We have

$$\sigma^{-1}(s) = \arg \min_U \{G(U) - s^\top U\}$$

thus we are looking for U that solves

$$\min_U \left\{ E_\epsilon \left[T \log \left(1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{U_y + (Z\epsilon_i)_y}{T} \right) \right) \right] - s^\top U \right\}$$

that is $v_y = -U_y$ solves

$$\min_{v: v_0=0} \left\{ E_\epsilon \left[T \log \left(1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{(Z\epsilon_i)_y - v_y}{T} \right) \right) \right] + \sum_{y \in \mathcal{Y}_f} s_y v_y \right\} \quad (2)$$

this problem is equivalent to

$$\begin{aligned} \max_{(\epsilon, Y)} \quad & E_\pi [(Z\epsilon)_Y] - T E_\pi [\log \pi(\epsilon, Y)] \\ \sim \quad & \pi \in M(P_\epsilon, s) \end{aligned}$$

Indeed the FOC in the (2) problem yield

$$s_y = \int f(\epsilon) \frac{\exp \left(\frac{-v_y + (Z\epsilon_i)_y}{T} \right)}{1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{-v_y + (Z\epsilon_i)_y}{T} \right)} d\epsilon$$

introduce

$$a(\epsilon) = -T \log \left(\frac{1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{-v_y + (Z\epsilon)_y}{T} \right)}{f(\epsilon)} \right)$$

so that the problem becomes

$$s_y = \int \exp \left(\frac{(Z\epsilon)_y - v_y - a(\epsilon)}{T} \right) d\epsilon$$

and we note that

$$\sum_y \exp \left(\frac{(Z\epsilon)_y - v_y - a(\epsilon)}{T} \right) = \sum_y \frac{f(\epsilon) \exp \left(\frac{(Z\epsilon)_y - v_y}{T} \right)}{1 + \sum_{y \in \mathcal{Y}} \exp \left(\frac{-v_y + (Z\epsilon)_y}{T} \right)} = f(\epsilon)$$

4.4 Parametric choice with no individual-specific observable heterogeneity

Assume

$$U_y = \sum_k \Phi_{yk} \beta_k$$

that is $U = \Phi\beta$.

Let us see how this works with maximum likelihood. The log-likelihood of the sample is

$$l(\beta) = N \sum_y \hat{s}_y \log \sigma_y(\Phi\beta)$$

and therefore the max-likelihood estimator is given by

$$\max_{\beta} l(\beta).$$

Let's see this in the logit model. In the logit model

$$\sigma_y(\Phi\beta) = \frac{\exp \left((\Phi\beta)_y \right)}{\sum_y \exp \left((\Phi\beta)_y \right)}$$

therefore we get

$$\max_{\beta} \sum_y \hat{s}_y (\Phi\beta)_y - \log \sum_y \exp \left((\Phi\beta)_y \right)$$

In the logit model, the max-likelihood problem is therefore a convex problem. In fact, this is

$$\max_{\beta} \left\{ \hat{s}^\top \Phi\beta - G(\Phi\beta) \right\}.$$

However, in the general case, there is no guarantee that

$$\frac{l(\beta)}{N} = \sum_y \hat{s}_y \log \sigma_y(\Phi\beta)$$

should be concave. In fact, we are no longer going to have

$$\sum_y \hat{s}_y \log \sigma_y(\Phi\beta) = \hat{s}^\top \Phi\beta - G(\Phi\beta)$$

Instead of doing max-likelihood in general, let's estimate β using the moment-matching estimator

$$\max_{\beta} \{ \hat{s}^\top \Phi\beta - G(\Phi\beta) \}$$

FOC of this problem wrt β_k :

$$(\hat{s}^\top \Phi)_k = (\sigma^\top(\Phi\beta) \Phi)_k$$

These are moment conditions; indeed

$$\sum_y \hat{s}_y \Phi_{yk} = \sum_y \sigma_y(\Phi\beta) \Phi_{yk}$$

Let's see how to compute this in practice in a probit model. In that case, recall that

$$\varepsilon_{iy} = (Z\epsilon_i)_y \text{ where } \epsilon_i \sim N(0, I_K), Z \text{ is a } |\mathcal{Y}_0| \times K \text{ matrix, } i = 1, \dots, N.$$

In that case

$$G_N(U) = \frac{1}{N} \sum_{i=1}^N \max_{y \in Y_0} \{ U_y + (Z\epsilon_i)_y \}$$

then the simulator for the moment-matching estimator

$$\max_{\beta \in R^K} \{ \hat{s}^\top \Phi\beta - G_N(\Phi\beta) \}$$

ie

$$\max_{\beta \in R^K} \left\{ \hat{s}^\top \Phi\beta - \frac{1}{N} \sum_{i=1}^N \max_{y \in Y_0} \{ (\Phi\beta)_y + (Z\epsilon_i)_y \} \right\}$$

which reformulates as

$$\begin{aligned} \max_{(u_i) \in R^N, \beta \in R^K} & \quad \left\{ \hat{s}^\top \Phi\beta - \frac{1}{N} \sum_{i=1}^N u_i \right\} \\ \text{s.t.} & \quad u_i - (\Phi\beta)_y \geq (Z\epsilon_i)_y \quad \forall i \in \{1, \dots, N\}, \forall y \in Y_0 \end{aligned}$$

4.5 Parametric choice with individual-specific observable heterogeneity

The analysis is left unchanged for the most part. In the logit case, the log-likelihood associated with observation i is

$$l_i(\beta) = \sum_{y \in \mathcal{Y}} \hat{\mu}_{iy} (\Phi\beta)_{iy} - \log \sum_{y \in \mathcal{Y}} \exp (\Phi\beta)_{iy}$$

and the max-likelihood rewrites as

$$\max_{\beta} \left\{ \sum_{i \in \mathcal{I}, y \in \mathcal{Y}} \hat{\mu}_{iy} (\Phi\beta)_{iy} - \sum_{i \in \mathcal{I}} \log \sum_{y \in \mathcal{Y}} \exp (\Phi\beta)_{iy} \right\}$$

With other random utility structures, this yields a moment matching procedure to estimate β , namely

$$\max_{\beta} \left\{ \hat{\mu}^\top \Phi\beta - \sum_{i \in \mathcal{I}} G((\Phi\beta)_i) \right\},$$

where G is the Emax operator associated with the distribution of the random utility.