Virtual whiteboard for Jan'22 math+econ+code

Alfred Galichon

January 13, 2022

1 Day 1

1.1 1a. The diet problem

Consider N_{ij}=amount of nutrient i that is brought by one dollar's worth of food j.

Look for q_j = \$ invested in food j
such that we minimize total cost
$\min \sum_j q_j$

subject to the constraint that the minimal intake in each food i is met
$\sum_j N_{ij} q_j \geq d_i$ for each nutrient i
where d_i is the minimum quantity of nutrient i.

To summarize, we need to solve
$\min_{q \geq 0} \sum_j q_j$
s.t. $\sum_j N_{ij} q_j \geq d_i$

In matrix form this is
$\min_{q \geq 0} c^\top q$
s.t. $Nq \geq d$
where $c_j = 1$.

1.1.1 Duality worked out by hand

$\min_{q \geq 0} \sum_j c_j q_j$
s.t. $\sum_j N_{ij} q_j \geq d_i$
transform into an unconstrained minimization problem
$\min_{q \geq 0} \sum_j c_j q_j + \sum_i F\left(d_i - \sum_j N_{ij} q_j\right)$
where $F(z) = 0$ if $z \leq 0$, and $F(z) = +\infty$ if $z > 0$.
How to represent F? How about
$F(z) = \max_{\pi \geq 0} \{\pi z\}$.

$\min_{q \geq 0} \sum_j c_j q_j + \sum_i \max_{\pi_i \geq 0} \pi_i \left(d_i - \sum_j N_{ij} q_j\right)$

$\min_{q\geq 0} \sum_j c_j q_j + \max_{\pi_i \geq 0} \sum_i \pi_i \left(d_i - \sum_j N_{ij} q_j \right)$

$\min_{q_j\geq 0} \max_{\pi_i \geq 0} \sum_j c_j q_j + \sum_i \pi_i \left(d_i - \sum_j N_{ij} q_j \right)$
Assuming $\min\max = \max\min$, we can reformulate this as
$\max_{\pi_i \geq 0} \min_{q_j\geq 0} \sum_j c_j q_j + \sum_i \pi_i \left(d_i - \sum_j N_{ij} q_j \right)$
$\max_{\pi_i \geq 0} \min_{q_j\geq 0} \sum_j c_j q_j + \sum_i \pi_i d_i - \sum_{i,j} \pi_i N_{ij} q_j$
going through the same steps in the opposite direction:
$\max_{\pi_i \geq 0} \sum_i \pi_i d_i + \min_{q_j\geq 0} \sum_j c_j q_j - \sum_{i,j} \pi_i N_{ij} q_j$
$\max_{\pi_i \geq 0} \sum_i \pi_i d_i + \min_{q_j\geq 0} \sum_j q_j \left(c_j - \sum_i \pi_i N_{ij} \right)$
$\max_{\pi_i \geq 0} \sum_i \pi_i d_i + \sum_j \min_{q_j\geq 0} q_j \left(c_j - \sum_i \pi_i N_{ij} \right)$
But $\left(\min_{q_j\geq 0} q_j w_j \right) = -\infty$ if $w_j < 0$ and $= 0$ if $w_j \geq 0$

Therefore
$\max_{\pi_i \geq 0} \sum_i \pi_i d_i$
$s.t.\ c_j \geq \sum_i \pi_i N_{ij}$
that is
$\max_{\pi \geq 0} \pi^\top d$
$s.t.\ N^\top \pi \leq c$

Theorem. Consider the "primal problem"
$\min_{q\geq 0} \sum_j c_j q_j$
s.t. $\sum_j N_{ij} q_j \geq d_i$ $[\pi_i \geq 0]$
and the "dual problem"
$\max_{\pi_i \geq 0} \sum_i \pi_i d_i$
$s.t.\ \sum_i \pi_i N_{ij} \leq c_j$ $[q_j \geq 0]$

Then if either of them is feasible [i.e. there is a variable that meets the constraints] then

(1) the value of the primal problem is equal to the value of the dual problem

(2) complementary slackness:

Assume q is a solution to the primal problem and π is a solution to the dual problem

$\pi_i > 0$ implies $d_i = \sum_j N_{ij} q_j$
$q_j > 0$ implies $c_j = \sum_i \pi_i N_{ij}$

Theorem. If q is feasible for the primal and π is feasible for the dual and if complementary slackness holds, then q is optimal for the primal and π is optimal for the dual.

1.2 1b. The optimal assigment problem

Joint surplus matrix

$\Phi_{xy} = x^\top A y = \sum_{k,l} A_{kl} x_k y_l$

Assume n_x men of type x and m_y women of type y.

Becker-Shapley-Shubik's model of matching.
Assume if man x matches with woman y, then:
x gets surplus α_{xy}
y gets surplus γ_{xy}

Assume utility is transferable, ie if w_{xy} is the transfer from the woman to the man (either positive or negative),
x gets surplus $\alpha_{xy} + w_{xy}$
y gets surplus $\gamma_{xy} - w_{xy}$
This is the Transferable Utility assumption.

w_{xy} is determined at equilibrium.
Regardless of what w_{xy} is, the joint surplus
$\Phi_{xy} = (\alpha_{xy} + w_{xy}) + \left(\gamma_{xy} - w_{xy}\right) = \alpha_{xy} + \gamma_{xy}$ is the same.

Roadmap:
1. Optimality
2. Equilibrium

1.2.1 Optimality

A matching is a $\mu_{xy} \geq 0$ which is the number of men of type x matched with women of type y.
Constraint on μ_{xy}:
$\sum_y \mu_{xy} = n_x$
$\sum_x \mu_{xy} = m_y$

Optimal matching consists of
$\max_{\mu \geq 0} \sum_{xy} \mu_{xy} \Phi_{xy}$
s.t.
$\sum_y \mu_{xy} = n_x \ \ [u_x]$
$\sum_x \mu_{xy} = m_y \ \ [v_y]$

Duality:
$\max_{\mu \geq 0} \sum_{xy} \mu_{xy} \Phi_{xy} + \sum_x \min_{u_x} u_x \left(n_x - \sum_y \mu_{xy}\right) + \sum_y \min_{v_y} v_y \left(m_y - \sum_x \mu_{xy}\right)$
$\min_{u_x, v_y} \sum_x n_x u_x + \sum_y m_y v_y + \max_{\mu \geq 0} \sum_{xy} \mu_{xy} \left(\Phi_{xy} - u_x - v_y\right)$
that is, the dual problem is
$\min_{u_x, v_y} \sum_x n_x u_x + \sum_y m_y v_y$
s.t. $u_x + v_y \geq \Phi_{xy} \ \ \left[\mu_{xy} \geq 0\right]$

By complementary slackness, we have $\mu_{xy} > 0 \implies u_x + v_y = \Phi_{xy}$.

1.2.2 Interpretation as a stable outcome.

We now consider the decentralized version of the problem.

Definition: (μ, u, v) is a stable outcome if

(1) μ is a matching:

$\sum_y \mu_{xy} = n_x$ $[u_x]$

$\sum_x \mu_{xy} = m_y$ $[v_y]$

(2) Stability: we have for all x and y that

$u_x + v_y \geq \Phi_{xy}$

[otherwise we would have $u_x + v_y < \Phi_{xy}$ and xy would be a blocking pair, ie there is a way for x to have more than u_x and y to have more than v_y if x and y match together]

(3) Feasibility: if $\mu_{xy} > 0$, then $u_x + v_y = \Phi_{xy}$.

Note that (1) means that μ is feasible for the primal problem

(2) means that (u, v) is feasible for the dual problem

(3) means complementary slackness.

Hence μ is a solution to the primal problem

$\max_{\mu \geq 0} \sum_{xy} \mu_{xy} \Phi_{xy}$

s.t.

$\sum_y \mu_{xy} = n_x$ $[u_x]$

$\sum_x \mu_{xy} = m_y$ $[v_y]$

and (u, v) is a solution to the dual problem

$\min_{u_x, v_y} \sum_x n_x u_x + \sum_y m_y v_y$

s.t. $u_x + v_y \geq \Phi_{xy}$ $\left[\mu_{xy} \geq 0\right]$

Recovering the transfers. We had that if w_{xy} is the transfer from y to x, then

x gets surplus $\alpha_{xy} + w_{xy}$

y gets surplus $\gamma_{xy} - w_{xy}$

Hence the payoff of x at equilibrium is

$u_x = \max_y \{\alpha_{xy} + w_{xy}\}$

and

$v_y = \max_x \left\{\gamma_{xy} - w_{xy}\right\}$

Hence

$u_x \geq \alpha_{xy} + w_{xy}$ for all x and y

$v_y \geq \gamma_{xy} - w_{xy}$ for all x and y

This yields

$\gamma_{xy} - v_y \leq w_{xy} \leq u_x - \alpha_{xy}$

We have $u_x - \alpha_{xy} \geq \gamma_{xy} - v_y$ – indeed $u_x + v_y \geq \alpha_{xy} + \gamma_{xy}$

When $\mu_{xy} > 0$, we have $w_{xy} = \gamma_{xy} - v_y = u_x - \alpha_{xy}$

When partners can remain unmatched.

Assume that individuals get utility zero if they remain unmatched.

Then a feasible matching imposes

$\sum_y \mu_{xy} \leq n_x$
$\sum_x \mu_{xy} \leq m_y$

An optimal matching solves
$\max_{\mu \geq 0} \sum_{xy} \mu_{xy} \Phi_{xy}$
s.t.
$\sum_y \mu_{xy} \leq n_x \;\; [u_x \geq 0]$
$\sum_x \mu_{xy} \leq m_y \;\; [v_y \geq 0]$

Duality:
$\max_{\mu \geq 0} \sum_{xy} \mu_{xy} \Phi_{xy} + \sum_x \min_{u_x \geq 0} u_x \left(n_x - \sum_y \mu_{xy} \right) + \sum_y \min_{v_y \geq 0} v_y \left(m_y - \sum_x \mu_{xy} \right)$
$\min_{u_x \geq, v_y \geq} \sum_x n_x u_x + \sum_y m_y v_y + \max_{\mu \geq 0} \sum_{xy} \mu_{xy} \left(\Phi_{xy} - u_x - v_y \right)$
that is, the dual problem is
$\min_{u_x \geq, v_y \geq 0} \sum_x n_x u_x + \sum_y m_y v_y$
s.t. $u_x + v_y \geq \Phi_{xy} \;\; \left[\mu_{xy} \geq 0 \right]$

By complementary slackness, we have $\mu_{xy} > 0 \implies u_x + v_y = \Phi_{xy}$.
The dual is
$\min_{u_x \geq 0, v_y \geq 0} \sum_x n_x u_x + \sum_y m_y v_y$
s.t. $u_x + v_y \geq \Phi_{xy} \;\; \left[\mu_{xy} \geq 0 \right]$

Alternatively, the primal can be expresses as
$\max_{\mu \geq 0} \sum_{xy} \mu_{xy} \Phi_{xy}$
s.t.
$\sum_y \mu_{xy} + \mu_{x0} = n_x \;\; [u_x \geq 0]$
$\sum_x \mu_{xy} + \mu_{0y} = m_y \;\; [v_y \geq 0]$
where μ_{x0} and μ_{0y} act as slackness variable.

Stability interpretation: a stable outcome (μ, u, v) in the problem with singles is such that

(1) μ is a feasible partial matching:
$\sum_y \mu_{xy} + \mu_{x0} = n_x$
$\sum_x \mu_{xy} + \mu_{0y} = m_y$
(2) Stability holds
$u_x + v_y \geq \Phi_{xy}$
$u_x \geq 0, v_y \geq 0$
(3) Complementary slackness
$\mu_{xy} > 0 \implies u_x + v_y = \Phi_{xy}$
$\mu_{x0} > 0$ i.e $\left(\sum_y \mu_{xy} < n_x \right) \implies u_x = 0$
$\mu_{0y} > 0$ i.e. $\left(\sum_x \mu_{xy} < m_y \right) \implies v_y = 0$

1.2.3 Indivisibilities (finite population)

When $n_x = 1$ for each x and $m_y = 1$ for each y, we should in principle impose an integrality constraint, that is

$$\mu_{xy} \in \{0, 1\}.$$

Then the problem is no longer a linear programming problem, but an integer programming problem.

However, in the bipartite case, one can abstract away from the integrality constraint.

1.2.4 Computation

Consider the problem

$\max_{\mu \geq 0} \sum_{xy} \mu_{xy} \Phi_{xy}$
s.t.
$\sum_y \mu_{xy} \leq n_x \ \ [u_x \geq 0]$
$\sum_x \mu_{xy} \leq m_y \ \ [v_y \geq 0]$
this is of the form
$\max_{\mu \geq 0} \mu^\top \Phi$
$M\mu \leq \binom{n}{m}$

How to convert a matrix into a vector? Take a matrix M, we call $vec\left(M\right)$ its vectorized version,

This can be done by:

* stack the columns: Matlab, Julia, R, Fortran column-major ordering, or Fortran ordering.

* stack the rows: done by NumPy by default, as well as C as well as some other languages: row-major ordering, or C ordering – primary convetion in this course.

Take the first set of constraints $\sum_y \mu_{xy} \leq n_x$. If μ is understood as a matrix, this is

$$\mu 1_Y \leq n$$

by pre-multuplying by the identity this yields

$$I_X \mu 1_Y \leq n$$

And we need to look at $vec\left(I_X \mu 1_Y\right) = matrix.vec\left(\mu\right)$

Fundamental identity is (assuming row-major ordering)

$$vec\left(AXB^\top\right) = \left(A \otimes B\right) vec\left(X\right).$$

Here, our constraints become

$$vec\left(I_X \mu 1_Y\right) = \left(I_X \otimes 1_Y^\top\right) vec\left(\mu\right) \leq n$$

Similarly, we can vectorize the other constraints $\sum_x \mu_{xy} \leq m_y$ into $1_X^\top \mu I_Y \leq m$, that is

$$vec\left(1_X^\top \mu I_Y\right) = \left(1_X^\top \otimes I_Y\right) vec\left(\mu\right) \leq m$$

Hence our optimal assignment problem becomes in a vectorized fashion

$$\begin{aligned}
\max_{vec(\mu)\geq 0} \quad & vec\left(\mu\right)^{\top} vec\left(\Phi\right) \\
s.t. \quad & \left(I_X \otimes 1_Y^{\top}\right) vec\left(\mu\right) \leq n \\
& \left(1_X^{\top} \otimes I_Y\right) vec\left(\mu\right) \leq m
\end{aligned}$$

that is

$$\begin{aligned}
\max_{v\geq 0} \quad & v^{\top} vec\left(\Phi\right) \\
s.t. \quad & Mv \leq \begin{pmatrix} n \\ m \end{pmatrix}
\end{aligned}$$

where

$$M = \begin{pmatrix} I_X \otimes 1_Y^{\top} \\ 1_X^{\top} \otimes I_Y \end{pmatrix}$$

is called the margining matrix.

2 Day 2

About yesterday's exercises

$$\begin{pmatrix} 1/4 & 1/4-1/8 & 1/4+1/8 \\ & 2/8+1/8 & 1/8-1/8 \\ & & \end{pmatrix}$$

networkx
algorithms to detect loops

Birkhoff-von Neumann

$\min \sum_{xy} \mu_{xy} c_{xy}$
$M\mu = \binom{n}{m}$
If $c_{xy} \in \{0,1\}$, then $\Gamma = \{xy : c_{xy} = 0\}$

$0 = \min \sum_{xy} \mu_{xy} c_{xy}$
$M\mu = \binom{n}{m}$
iff for every xy such that $\mu_{xy} > 0$, then $xy \in \Gamma$
this means that there is a matching between n and m "compatible" with Γ.

The dual to
$\min \sum_{xy} \mu_{xy} c_{xy}$
$M\mu = \binom{n}{m}$
is
$\max_{u,v} \sum_x u_x - \sum_y v_y$
s.t. $u_x - v_y \leq c_{xy}$

$$\Gamma = \begin{pmatrix} 1 & 1 & 1 & 1 \\ & & & 1 \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

2.1 2a. One-dimensional matching

We now assume that types of workers x and firms y belong in $\mathbb{R}$.

Assume that there is the same total mass of workers and firms.

The total mass of workers and firms is normalized to one.

$n(x)$ is the density of probability associated with the distribution of workers

$m(y)$ is the density of probability associated with the distribution of firms

A matching is a distribution of probability $\mu(x,y)$ over pairs x, y. It should satisfy

$\int \mu(x,y)\, dy = n(x)$
$\int \mu(x,y)\, dx = m(y)$

Assume that the economic value created by a CEO of type x with a firm of type y is

$\Phi(x, y)$.

The optimal assignment problem is

$\max_\mu \int \Phi(x, y) \mu(x, y) \, dx dy$

s.t.

$\int \mu(x, y) \, dy = n(x), x \in R$

$\int \mu(x, y) \, dx = m(y), y \in R$

CEO application (Gabaix and Landier, Tervio):

x is the CEO's talent = extra % return on asset the CEO generates

y is the firm's size (market cap)

In that case, we get that

$\Phi(x, y) = xy$

Agenda:

Goal = predict CEO compensation

Solve for the primal problem

From the solution to the primal problem, we will deduce the solution to the dual problem

The dual problem will give us predictions for CEO compensation

Intuition: If $T(x)$ is the firm that CEO x is matched with, it makes sense to assume that

$T(x)$ is increasing.

If that is the case, what is $T(x)$?

We know that if $X \sim n$ is a random variable distributed according to n, then we $T(X) \sim m$.

We can introduce the cumulative distribution functions (CDF) $F_n(x) = \int_{-\infty}^{x} n(z) \, dz$ and $F_m(y) = \int_{-\infty}^{y} m(z) \, dz$ associated with n and m respectively, and we have

$$\Pr(T(X) \leq y) = F_m(y)$$

take $y = T(x)$ for a fixed value of x, we get

$$\Pr(T(X) \leq T(x)) = F_m(T(x))$$

thus

$$\begin{aligned} \Pr(X \leq x) &= F_m(T(x)) \\ F_n(x) &= F_m(T(x)) \end{aligned}$$

therefore

$$T(x) = F_m^{-1}(F_n(x))$$

In particular, this means that the median CEO – ie the CEO x such $F_n(x) = 1/2$ is matched with the median firm indeed $F_n(x) = 1/2 = F_m(T(x))$ – thus $T(x)$ is the median firm.

Let's see how we can solve for the dual problem. The dual problem is
$\min_{u(x),v(y)} \int u(x) n(x) dx + \int v(y) m(y) dy$
s.t. $u(x) + v(y) \geq \Phi(x,y) = xy$

Assume (u,v) is a ** feasible solution ** to the dual problem.
Then for every x and every y,
$u(x) + v(y) \geq xy$
thus
$v(y) \geq \max_x \{xy - u(x)\}$ holds for for every y.
Claim: for any ** optimal solution ** to the dual problem, we have
$v(y) = \max_x \{xy - u(x)\}$ – interpreted as firm y's problem
Obviously, this is symmetric, so we have also
$u(x) = \max_y \{xy - v(y)\}$ – interpreted as worker x's problem.

Let's write down optimality conditions in the firm y's problem
$y = u'(x)$.

But we are in the case where $T(x)$ is known and $T(x) = F_m^{-1}(F_n(x))$.
We have therefore
$u'(x) = F_m^{-1}(F_n(x))$
as a result
$u(x) = \int^x F_m^{-1}(F_n(z)) dz + cte.$
Deduce v (either by $v(y) = \max_x \{xy - u(x)\}$, or $v(y) = \int^y F_n^{-1}(F_m(z)) dz + cte$)

Theorem. When Φ is supermodular, i.e. $\partial^2\Phi(x,y)/\partial x\partial y \geq 0$, then the positive assortative matching solution $T(x) = F_m^{-1}(F_n(x))$ is optimal.
Until now we have assumed $\Phi(x,y) = xy$, so $\partial^2\Phi(x,y)/\partial x\partial y = 1 \geq 0$.

We have that if u and v are optimal dual solutions, then
$v(y) = \max_x \{\Phi(x,y) - u(x)\}$ – interpreted as firm y's problem
Obviously, this is symmetric, so we have also
$u(x) = \max_y \{\Phi(x,y) - v(y)\}$ – interpreted as worker x's problem.

By first order condition in the firm's $T(x)$ problem, we have
$u'(x) = \partial_x\Phi(x,T(x))$
Now let's verify that if Φ is supermodular, then T is increasing. Deriving the above wrt x, we have
$u''(x) = \partial^2_{xx}\Phi(x,T(x)) + \partial^2_{xy}\Phi(x,T(x)) T'(x)$
therefore

$$T'(x) = \frac{u''(x) - \partial^2_{xx}\Phi(x,T(x))}{\partial^2_{xy}\Phi(x,T(x))}$$

The denominator is positive $\partial^2_{xy}\Phi(x, T(x)) > 0$ by supermodularity

We have $u''(x) - \partial^2_{xx}\Phi(x, T(x)) \geq 0$ by second order conditions. Indeed, $\partial^2_{xx}\Phi(x, y) - u''(x) \leq 0$.

Hence

$$T'(x) \geq 0.$$

Exercise. Model of marriage with taxes.

If single individual get gross income w, then gets net amount $N(w)$ where N is increasing and concave.

Assume x and y are the gross incomes of two marital partners. Then their combined gross income is $x + y$, and we create two fictious personnas which make $\frac{x+y}{2}$ each. Thus their combined net income is $2N\left(\frac{x+y}{2}\right)$.

Thus we can consider a matching market where the matching surplus is

$\Phi(x, y) = 2N\left(\frac{x+y}{2}\right)$

1) Is there surplus to matching? i.e. do we have $2N\left(\frac{x+y}{2}\right) \geq N(x) + N(y)$?

2) Derive the matching equilibrium on this market.

Solution.

1) $N\left(\frac{x+y}{2}\right) \geq \frac{N(x)+N(y)}{2}$ because N is concave.

2) We have that the matching surplus $\Phi(x, y) = 2N\left(\frac{x+y}{2}\right) - N(x) - N(y)$. The cross-derivative is

$$\frac{\partial^2\Phi(x, y)}{\partial x \partial y} = \frac{1}{2}N''\left(\frac{x+y}{2}\right) \leq 0$$

by concavity.

How can we use what we saw before to tackle this? Define $\tilde{x} = x$ and $\tilde{y} = -y$ and the surplus becomes

$$\tilde{\Phi}(\tilde{x}, \tilde{y}) = \Phi(\tilde{x}, -\tilde{y})$$

We have x and y are matched if and only if

$$F_X(x) = 1 - F_Y(y).$$

2.2 2b. Semi-discrete optimal transport

Inhabitant's problem

$$u(x) = \max_j \left\{x^\top y_j - v_j\right\}$$

I would like to get a formula for the market share of fountain j.

Let us compute the aggregate indirect welfare of the consumer. This is

$$\int_{\mathcal{X}} u(x)\, n(x)\, dx$$

where $n(x)$ is the density of inhabitants at x. Expressed as a function of the prices, this is

$$F(v) = \int_{\mathcal{X}} \max_j \left\{ x^\top y_j - v_j \right\} n(x)\, dx$$

Claim:

$$\frac{\partial F}{\partial v_j} = -D_j(v)$$

Thus the demand for fountain j is given

$$D_j(v) = -\frac{\partial F(v)}{\partial v_j}.$$

Recall that fountain j has fixed capacity q_j. Therefore the market-clearing prices v_j of the fountains are given by

$$D_j(v) = q_j,$$

that is

$$\frac{\partial F(v)}{\partial v_j} + q_j = 0$$

which we can rewrite as

$$\frac{\partial}{\partial v_j} \left\{ F(v) + \sum_k q_k v_k \right\} = 0$$

that is v is obtained by minimizing $S(v) := F(v) + \sum_k q_k v_k$ over v. Thus, v is a solution to

$$\min_v \left\{ F(v) + \sum_k q_k v_k \right\}$$

that is

$$\min_v \left\{ \int_{\mathcal{X}} \max_j \left\{ x^\top y_j - v_j \right\} n(x)\, dx + \sum_k q_k v_k \right\}$$

but we can view this as

$$\begin{aligned} \min_v \quad & \left\{ \int_{\mathcal{X}} u(x)\, n(x)\, dx + \sum_k q_k v_k \right\} \\ s.t. \quad & u(x) \geq \max_j \left\{ x^\top y_j - v_j \right\} \end{aligned}$$

which reformates as

$$\begin{aligned} \min_v \quad & \left\{ \int_{\mathcal{X}} u(x)\, n(x)\, dx + \sum_k q_k v_k \right\} \\ s.t. \quad & u(x) + v_j \geq x^\top y_j \;\; \forall x, \forall j \end{aligned}$$

Tommaso' suggestion.

$$v_j^{t+1} = v_j^t + \varepsilon \left(D_j \left(v \right) - q_j \right).$$

We have $S\left(v\right) = F\left(v\right) + \sum_k q_k v_k$, and therefore

$$\frac{\partial S\left(v\right)}{\partial v_j} = -D_j\left(v\right) + q_j$$

thus this algorithm amounts to

$$v_j^{t+1} = v_j^t - \varepsilon \frac{\partial S\left(v\right)}{\partial v_j}.$$

This is gradient descent! ie

$$v^{t+1} = v^t - \varepsilon \nabla S\left(v\right).$$

Another possibility would be **coordinate descent**.
$q_j = D_j\left(v_j^{t+1}, v_{-j}^t\right)$
Parallel version: Jacobi
Sequential version: Gauss-Seidel

Note that
$q_j = D_j\left(v_j^{t+1}, v_{-j}^t\right)$ is equivalent to
$v_j^{t+1} = \arg\min_{v_j} S\left(v_j, v_{-j}^t\right)$

Final remark. Back to the central planner problem.

The assinment we found consists in mapping inhabitant x with fountain y_j such that $j \in \arg\max_j \left\{x^\top y_j - v_j\right\}$.

Noting that $u\left(x\right) = \max_j \left\{x^\top y_j - v_j\right\}$, we have that $\nabla u\left(x\right) = y_j$. Thus $T\left(x\right) = \nabla u\left(x\right)$ is the optimal assignment from inhabitants to fountains, in the sense that it solves the primal problem

$$\begin{aligned}
\max_{\mu(x,y_j)} \quad & -\int \left|x - y_j\right|^2 \mu\left(x, y_j\right) dx \\
s.t. \quad & \int \mu\left(x, y_j\right) dx = q_j \\
& \sum_j \mu\left(x, y_j\right) = n\left(x\right).
\end{aligned}$$

Exercise. Implement a coordinate descent (Gauss-Seidel) version of the algorithm.

3 Day 3

3.1 3a. Optimal transport with entropic regularization

Consider the problem

$$\begin{aligned}\max_{\mu\geq 0} \quad & \sum_{xy}\mu_{xy}\Phi_{xy}-\sigma\sum_{xy}\mu_{xy}\ln\mu_{xy}\\ s.t. \quad & \sum_{y}\mu_{xy}=n_x\ \ [u_x]\\ & \sum_{x}\mu_{xy}=m_y\ \ [v_y]\end{aligned}$$

Reformulate into

$$\max_{\mu\geq 0}\sum_{xy}\mu_{xy}\Phi_{xy}-\sigma\sum_{xy}\mu_{xy}\ln\mu_{xy}+\min_{(u_x)}\sum_{x}u_x\left(n_x-\sum_{y}\mu_{xy}\right)+\min_{(v_y)}\sum_{y}v_y\left(m_y-\sum_{x}\mu_{xy}\right)$$

that is

$$\min_{u_x,v_y}\max_{\mu\geq 0}\sum_{xy}\mu_{xy}\Phi_{xy}-\sigma\sum_{xy}\mu_{xy}\ln\mu_{xy}+\sum_{x}u_x\left(n_x-\sum_{y}\mu_{xy}\right)+\sum_{y}v_y\left(m_y-\sum_{x}\mu_{xy}\right)$$

hence

$$\min_{u_x,v_y}\max_{\mu\geq 0}\sum_{xy}\mu_{xy}\Phi_{xy}-\sigma\sum_{xy}\mu_{xy}\ln\mu_{xy}+\sum_{x}u_xn_x-\sum_{xy}\mu_{xy}u_x+\sum_{y}v_ym_y-\sum_{xy}\mu_{xy}v_y$$

hence

$$\min_{u_x,v_y}\sum_{x}u_xn_x+\sum_{y}v_ym_y+\max_{\mu\geq 0}\sum_{xy}\mu_{xy}\left(\Phi_{xy}-u_x-v_y\right)-\sigma\sum_{xy}\mu_{xy}\ln\mu_{xy}$$

let's compute the soft penalization $\max_{\mu\geq 0}\sum_{xy}\mu_{xy}\left(\Phi_{xy}-u_x-v_y\right)-\sigma\sum_{xy}\mu_{xy}\ln\mu_{xy}$. This is

$$\sum_{xy}\max_{\mu_{xy}\geq 0}\left\{\mu_{xy}\left(\Phi_{xy}-u_x-v_y\right)-\sigma\mu_{xy}\ln\mu_{xy}\right\}$$

Set $a_{xy}=\Phi_{xy}-u_x-v_y$ and compute

$$\max_{\mu\geq 0}\left\{\mu a-\sigma\mu\ln\mu\right\}$$

We have by first order conditions that

$$a=\sigma\left(1+\ln\mu\right)$$

hence

$$\mu a=\sigma\mu+\sigma\mu\ln\mu$$

thus

$$\mu a - \sigma\mu \ln \mu = \sigma\mu$$

but μ can be obtained by $a = \sigma\left(1 + \ln \mu\right)$

$$\mu = \exp\left(\frac{a}{\sigma} - 1\right)$$

Therefore we have

$$\max_{\mu \geq 0} \left\{\mu a - \sigma\mu \ln \mu\right\} = \sigma \exp\left(\frac{a}{\sigma} - 1\right)$$

Hence the value of the problem is

$$\min_{u_x, v_y} \sum_x u_x n_x + \sum_y v_y m_y + \sum_{xy} \sigma \exp\left(\frac{\Phi_{xy} - u_x - v_y - \sigma}{\sigma}\right)$$

Let's replace v_y by $v_y + \sigma$, and this becomes

$$\min_{u_x, v_y} F\left(u, v\right) := \sum_x u_x n_x + \sum_y v_y m_y + \sum_{xy} \sigma \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma}\right).$$

Let's look at the first order conditions
$n_x - \sum_y \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma}\right) = 0$
$m_y - \sum_x \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma}\right) = 0$

How to recover the primal solution μ_{xy} from the dual solution (u_x, v_y)? Well, we have

$$\mu_{xy} = \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma}\right).$$

How to solve for this problem?
* gradient descent:
$u_x^{t+1} = u_x^t + \epsilon \frac{\partial F}{\partial u_x} = u_x^t + \epsilon \left(n_x - \sum_y \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma}\right)\right)$
$v_y^{t+1} = v_y^t + \epsilon \frac{\partial F}{\partial v_y} = v_y^t + \epsilon \left(m_y - \sum_x \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma}\right)\right)$
* maximizing with respect to u_x only means setting the corresponding foc to zero, ie

$$n_x = \sum_y \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma}\right)$$

that is

$$n_x = \exp\left(-\frac{u_x}{\sigma}\right) \sum_y \exp\left(\frac{\Phi_{xy} - v_y}{\sigma}\right)$$

that is

$$\exp\left(\frac{-u_x}{\sigma}\right) = \frac{n_x}{\sum_y \exp\left(\frac{\Phi_{xy} - v_y}{\sigma}\right)}$$

ie

$$u_x = \sigma \ln \left(\frac{1}{n_x} \sum_y \exp \left(\frac{\Phi_{xy} - v_y}{\sigma} \right) \right)$$

This gives us the following coordinate algorithm

$$\left\{ \begin{array}{c} u_x^{t+1} = \sigma \ln \left(\frac{1}{n_x} \sum_y \exp \left(\frac{\Phi_{xy} - v_y^t}{\sigma} \right) \right) \\ v_y^{t+1} = \sigma \ln \left(\frac{1}{m_y} \sum_x \exp \left(\frac{\Phi_{xy} - u_x^{t+1}}{\sigma} \right) \right) \end{array} \right.$$

or, setting $A_x = \exp(-u_x/\sigma)$ and $B_y = \exp(-v_y/\sigma)$

$$\left\{ \begin{array}{c} A_x^{t+1} = \frac{n_x}{\sum_y \exp\left(\frac{\Phi_{xy}}{\sigma}\right) B_y^t} \\ B_y^{t+1} = \frac{m_y}{\sum_x \exp\left(\frac{\Phi_{xy}}{\sigma}\right) A_x^{t+1}} \end{array} \right.$$

This is called the Iterated Proportional Fitting Procedure (IPFP), or Sinkhorn's algorithm.

The log-sum-exp trick. We want to compute $\log\left(e^a + e^b\right)$ and potentially a or b – or both – are very large; hence the exponentials will blow up.

Idea= for any constant c, $\log\left(e^a + e^b\right) = c + \log\left(e^{a-c} + e^{b-c}\right)$. Indeed,

$$c + \log\left(e^{a-c} + e^{b-c}\right) = c + \log\left(e^{-c}\left(e^a + e^b\right)\right) = c - c + \log\left(e^a + e^b\right).$$

Take $c = \max(a, b)$, then $a - c \leq 0$ and $b - c \leq 0$ and at least one of these terms is zero. Thus

$$\log\left(e^a + e^b\right) = \max\{a, b\} + \log\left(e^{\min\{0, a-b\}} + e^{\min\{0, b-a\}}\right)$$

3.1.1 Generalized linear models

Classical linear regression

$Y = X\beta + \varepsilon$

expresses as

$E[Y|X = x] = x^\top \beta$

We generalize this into

$E[Y|X = x] = g^{-1}(X\beta)$

In particular $g = \ln$, yields

$E[Y|X = x] = \exp(X\beta)$

How to estimate β then? Based on the observation of (x_i, y_i).

We are going to build a method that will match the moments $E[Xy]$

Idea is to compute

$$\min_\beta \left\{ \frac{1}{n} \sum_i e^{X_i \beta} - \frac{1}{n} \sum_i y_i X_i \beta \right\}$$

Population version is

$$\min_{\beta}\left\{E\left[e^{X\beta}\right]-\hat{E}\left[YX\beta\right]\right\}$$

foc yield

$$E\left[Xe^{X\beta}\right]=\hat{E}\left[YX\right]$$

but by assumption, $E\left[e^{X\beta}\right]=E\left[Y|X\right]$, thus for the optimal value of β

$$E\left[XY\right]=\hat{E}\left[YX\right]$$

3.2 3b. Random utility models

Decision maker i

Decision maker needs to choose among alternatives $y\in\mathcal{Y}$.

Utility of decision maker i taking decision y is

$$U_y+\varepsilon_{iy}$$

where ε_{iy} is an individual-specific term which is distributed according to distribution P, and U is a vector called the vector of systematic utilities.

Decision maker i's problem is

$$\max_{y\in\mathcal{Y}}\left\{U_y+\varepsilon_{iy}\right\}$$

Market share of y

$$Q_y\left(U\right)=P\left(U_y+\varepsilon_{iy}\geq U_z+\varepsilon_{iz}\forall z\in\mathcal{Y}\right)$$

$$\begin{aligned}\frac{\partial Q_y}{\partial U_y} &\geq 0\\ \frac{\partial Q_y}{\partial U_z} &\leq 0, z\neq y\end{aligned}$$

Compute the mean indirect utilities of all the decision makers. That is

$$\frac{1}{N}\sum_i\max_{y\in\mathcal{Y}}\left\{U_y+\varepsilon_{iy}\right\}$$

In a large population, by the law of large numbers, this becomes

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

We have

$$\frac{\partial G\left(U\right)}{\partial U_y}=P\left(U_y+\varepsilon_{iy}\geq U_z+\varepsilon_{iz}\forall z\in\mathcal{Y}\right)$$

indeed,

$$\begin{aligned}\frac{\partial G(U)}{\partial U_y} &= E_P\left[\frac{\partial}{\partial U_y}\max_{y\in\mathcal{Y}}\{U_y+\varepsilon_y\}\right] \\ &= E_P\left[1\left\{y\in\arg\max_{y\in\mathcal{Y}}\{U_y+\varepsilon_y\}\right\}\right]\end{aligned}$$

This means that

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

Assume that we observe (q_y) the market share of each y and we are after (U_y). How do we solve the demand inversion problem, ie how do we recover (U_y) such that

$$Q_y(U)=q_y, \forall y\in\mathcal{Y}$$

We replace Q by its values as the gradient of G, and we have

$$\frac{\partial G(U)}{\partial U_y}=q_y, \forall y\in\mathcal{Y}$$

therefore the U's that we are looking for need to solve

$$\max_{(U_y)}\left\{\sum_{y\in\mathcal{Y}}q_yU_y-G(U)\right\}.$$

This expression plays an important role, and we define

$$G^*(q)=\max_{(U_y)}\left\{\sum_{y\in\mathcal{Y}}q_yU_y-G(U)\right\}$$

and we call G^* the **generalized entropy of choice**.

Example. The logit model.

Assume (ε_y) are iid random variables, and that the c.d.f of ε_y is $\exp(-\exp(-(z+\gamma)))$ where γ =Euler's constant.

Then we have

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

(remember $G(U)=E_P[\max_y\{U_y+\varepsilon_y\}]$

More generally, we have

$$E_P\left[\max_y\{U_y+\sigma\varepsilon_y\}\right]=\sigma\log\sum_{y\in\mathcal{Y}}\exp\left(\frac{U_y}{\sigma}\right).$$

Exercise for tomorrow. Show that in the case of the logit model,

$$\begin{aligned}G^*(q) &= \sum_{y\in\mathcal{Y}}q_y\ln q_y \text{ if } \sum_{y\in\mathcal{Y}}q_y=1 \text{ and } q_y>0 \text{ for all } y \\ &= +\infty\end{aligned}$$

4 Day 4

4.1 4a. Random utility models (ctd)

4.2 Nonparametric inversion

Compute

$$G^*(q) = \max_{U_y}\left\{\sum_y q_y U_y - \log \sum \exp U_y\right\}$$

Assume an interior solution. By FOC

$$q_y = \frac{e^{U_y}}{\sum_y \exp U_y}$$

thus $\sum q_y = 1$ and $q_y > 0$ necessarily. Taking logs,

$$\ln q_y = U_y - \log \sum_y \exp U_y$$

and multiplying by q_y and summing over y, we get

$$\begin{aligned}
\sum_y q_y \ln q_y &= \sum_y q_y U_y - \left(\sum_y q_y\right) \log \sum_y \exp U_y \\
&= \sum_y q_y U_y - \log \sum \exp U_y = G^*(q).
\end{aligned}$$

Now, let's interpret $G^*(q)$. When $q_y = \frac{e^{U_y}}{\sum_y \exp U_y}$, we have that

$$G^*(q) = \sum_y q_y U_y - G(U)$$

thus

$$G(U) = \sum_y q_y U_y - G^*(q).$$

Recall

$$G(U) = E_P\left[\max_y \{U_y + \varepsilon_y\}\right]$$

Introduce $Y^\varepsilon = \arg\max_y \{U_y + \varepsilon_y\}$ – this is the alternative chosen by an agent with ε.

$$\begin{aligned}
G(U) &= E_P\left[\max_y \{U_y + \varepsilon_y\}\right] = E_P\left[U_{Y^\varepsilon} + \varepsilon_{Y^\varepsilon}\right] \\
&= E_P\left[U_{Y^\varepsilon}\right] + E_P\left[\varepsilon_{Y^\varepsilon}\right] \\
&= \sum_y U_y P\left(Y^\varepsilon = y\right) + E_P\left[\varepsilon_{Y^\varepsilon}\right] \\
&= \sum_y q_y U_y + E_P\left[\varepsilon_{Y^\varepsilon}\right]
\end{aligned}$$

compare this with $G(U) = \sum_y q_y U_y - G^*(q)$, we get that

$$-G^*(q) = E_P[\varepsilon_{Y^\varepsilon}].$$

How can we deduce U from G^*? Well recall that

$$G^*(q) = \max_{U_y} \left\{ \sum_y q_y U_y - \log \sum \exp U_y \right\}$$

Before we go on, we need to remark that a normalization is needed. Indeed, if U_y is a solution to

$$q_y = \frac{e^{U_y}}{\sum_y e^{U_y}}$$

then $(U_y + c)_y$ is also a solution for any constant c.

Assume there is a default option 0 and we call $Y_0 = Y \cup \{0\}$ the set of all alternatives including the default option, and normalize $U_0 = 0$. Then we are back to a unique solution. Consider incorporating the normalization to the problem. We have

$$G(U) = E\left[\max_{y \in Y} \{U_y + \varepsilon_y, 0 + \varepsilon_0\}\right]$$

In the logit model

$$G(U) = \log \sum_{y \in Y_0} \exp U_y = \log \left(1 + \sum_{y \in Y} \exp U_y\right)$$

now the market shares associated with the nondefault options $y \in Y$ are given by

$$\frac{\partial G(U)}{\partial U_y} = \frac{\exp U_y}{1 + \sum_{y \in Y} \exp U_y}$$

and now the generalized entropy of choice can be expressed by

$$\begin{aligned} G^*(q) &= \max_{U \in R^{Y_0}} \left\{ \sum_{y \in Y_0} q_y U_y - G(U) : U_0 = 0 \right\} \\ &= \max_{U \in R^Y} \left\{ \sum_{y \in Y} q_y U_y - G(U) \right\} \end{aligned}$$

so I can view G^* as a function of $(q_y)_{y \in Y}$ ie the market shares of nondefault alternatives.

In the logit model, we have

$$G^*\left((q)_{y\in Y}\right) = \max_{U\in R^Y}\left\{\sum_{y\in Y} q_y U_y - \log\left(1+\sum_y \exp U_y\right)\right\}$$

We are getting

$$G^*\left((q)_{y\in Y}\right) = \sum_{y\in Y} q_y \log q_y + q_0 \log q_0$$

where $q_0 = 1 - \sum_{y\in Y} q_y$, if $q_y > 0$ for all $y \in Y$ and $1 > \sum_{y\in Y} q_y$; $+\infty$ otherwise.

In other terms

$$G^*\left((q)_{y\in Y}\right) = \sum_{y\in Y} q_y \log q_y + \left(1-\sum_{y\in Y} q_y\right)\log\left(1-\sum_{y\in Y} q_y\right).$$

We have that

$$\frac{\partial G^*\left((q)_{y\in Y}\right)}{\partial q_y} = U_y$$

where U is such that $Q_y(U) = q_y$.

In the case of the entropy, we have

$$\begin{aligned}
U_y &= \frac{\partial G^*\left((q)_{y\in Y}\right)}{\partial q_y} \\
&= 1+\log q_y + \frac{\partial}{\partial q_y}\left\{q_0 \log q_0\right\} \\
&= 1+\log q_y + \frac{\partial}{\partial q_0}\left\{q_0 \log q_0\right\} \times \frac{\partial q_0}{\partial q_y} \\
&= 1+\log q_y + (1+\log q_0)\times -1 \\
&= \log q_y - \log q_0 \\
&= \log \frac{q_y}{q_0}
\end{aligned}$$

this is a well-known formula called the log-odds ratio formula.

A bit more on G^*. Assume now that P is no longer the iid Gumbel distribution – we don't have an explicit way to recover U from q. Instead we will resort to simulated methods.

We have seen that U is identified from (q_y) by

$$\min_U\left\{G(U) - \sum_y q_y U_y\right\} \tag{1}$$

What kind of problem is this? Recall the expression for $G(U)$:

$$G(U) = E_P \left[\max_y \{U_y + \varepsilon_y\} \right].$$

Take a sample of individuals $i \in \mathcal{I}$ where $|\mathcal{I}| = n$. Each i draws a vector $(\varepsilon_{iy})_{y \in Y}$ from the distribution P. The sample analog of G becomes

$$G_n(U) = \frac{1}{n} \sum_{i=1}^{n} \max_{y \in Y} \{U_y + \varepsilon_{iy}\}$$

Assume that we observe $(q_y)_y$. We would like to recover U_y. The sample analog of problem (1) is

$$\min_U \left\{ G_n(U) - \sum_y q_y U_y \right\}$$

that is

$$\min_U \left\{ \frac{1}{n} \sum_{i=1}^{n} \max_{y \in Y} \{U_y + \varepsilon_{iy}\} - \sum_y q_y U_y \right\}$$

Introduce $v_y = -U_y$, and we get

$$\min_{v_y} \left\{ \frac{1}{n} \sum_{i=1}^{n} \max_{y \in Y} \{-v_y + \varepsilon_{iy}\} + \sum_y q_y v_y \right\}$$

Claim: this rewrites as

$$\begin{array}{ll} \min_{u_i, v_y} & \left\{ \sum_{i=1}^{n} \frac{1}{n} u_i + \sum_y q_y v_y \right\} \\ s.t. & u_i + v_y \geq \varepsilon_{iy} \end{array}$$

Indeed, the constraint in the latter problem implies $u_i \geq \max_y \{-v_y + \varepsilon_{iy}\}$, but optimality implies that this holds as an equality.

Hence: inverting a discrete choice model is an optimal transport problem.

See G+Salanie – Cupids invisible hand.

4.3 Parametric estimation – characteristics approach / Logistic regression

For now, let's assume we observe no characteristics about the decision makers.

We do observe characteristics about alternatives.

$U_y = (\Phi\beta)_y$ where for each alternative y, Φ_{yk} is the k-th characteristics associated with y.

For example – y are car models, and Φ_{y1} is fuel efficiency, Φ_{y2} is number of seats, Φ_{y3} is a safety rating, We are assuming that $U_y = \sum_k \Phi_{yk}\beta_k$, where β_k are coefficients to be estimated.

Assume that the random utilities are iid Gumbel.

Then the overall indirect utility is

$$G(\Phi\beta) = \log \sum_y \exp(\Phi\beta)_y$$

Let's derive $G(\Phi\beta)$ with respect to β_k. We have

$$\begin{aligned}
\frac{\partial G(\Phi\beta)}{\partial \beta_k} &= \sum_y \frac{\partial G}{\partial U_y}(\Phi\beta)\frac{\partial U_y}{\partial \beta_k} \\
&= \sum_y Q_y(\Phi\beta)\Phi_{yk} \\
&= E_\beta[\Phi_{Yk}]
\end{aligned}$$

this is the moment of the k-th characteristics predicted by the model with parameter β.

I am observing the market shares $\hat{q}_y$ of the models actually purchased, so my estimation strategy will consist of matching those moments, i.e. looking for the value of the parameter β such that the predicted moments match with the observed moments. That is, look for β such that

$$\begin{aligned}
E_\beta[\Phi_{Yk}] &= \hat{E}[\Phi_{Yk}], \ i.e. \\
\sum_y Q_y(\Phi\beta)\Phi_{yk} &= \sum_y \hat{q}_y \Phi_{yk}
\end{aligned}$$

ie

$$\frac{\partial}{\partial \beta_k} G(\Phi\beta) = \sum_y \hat{q}_y \Phi_{yk}$$

which are moment matching conditions: predicted moments=observed moments.

We are going to view this as the first order conditions associated with a convex optimization problem and use that reformulation to estimate β.

This is the idea of generalized linear models (GLM). Let's see how it works. Introduce

$$\max_\beta \left\{ \sum_k \sum_y \beta_k \hat{q}_y \Phi_{yk} - G(\Phi\beta) \right\}$$

THe first order conditions for this problem are the moment-matching conditions above. The problem is a convex optimization problem, so gradient descent (or extensions) will lead to the estimator of β.

In the logit model, we have

$$\max_\beta \left\{ \sum_k \sum_y \beta_k \hat{q}_y \Phi_{yk} - \log \sum_y \exp\left((\Phi\beta)_y \right) \right\}$$

which is the familiar logistic regression.

Alternative approach via maximum likelihood. The probability that a decision maket picks up choice y is

$$l_y(\beta) = \frac{\partial G}{\partial U_y}(\Phi\beta)$$

Thus the likelihood of the sample is (up to rescaling)

$$\sum_y \hat{q}_y \ln l_y(\beta) = \sum_y \hat{q}_y \ln \frac{\partial G}{\partial U_y}(\Phi\beta)$$

and max-likelihood consists of

$$\max_\beta \sum_y \hat{q}_y \ln \frac{\partial G}{\partial U_y}(\Phi\beta).$$

In general (ie. for general structures of distribution of random utility), this is NOT a convex optimization problem.

HOWEVER, in the logit case, both approached coincide. Indeed, in that case,

$$l_y(\beta) = \frac{\partial G}{\partial U_y}(\Phi\beta) = \frac{e^{(\Phi\beta)_y}}{\sum_y e^{(\Phi\beta)_y}}$$

and thus

$$\sum_y \hat{q}_y \ln l_y(\beta) = \sum_y \hat{q}_y \left((\Phi\beta)_y - \log\left(\sum_y e^{(\Phi\beta)_y} \right) \right)$$

so in this case (but in this case only), this is

$$\sum_y \hat{q}_y \ln l_y(\beta) = \sum_y \hat{q}_y \left((\Phi\beta)_y - G(\Phi\beta) \right).$$

4.4 With characteristics about decision-makers observed.

Let us now assume we observe some characteristics such as income about decision maker i.

The likelihood that individual i will make decision y is now

$$\frac{\exp\left((\Phi\beta)_{iy}\right)}{\sum_y \exp\left((\Phi\beta)_{iy}\right)}$$

where $(\Phi\beta)_{iy} = \sum_k \Phi_{iy,k}\beta_k$. We now observe which decision maker makes which decision, that is we observe

$$\hat{\mu}_{iy} = 1\{i \text{ chooses } y\}.$$

Let us compute the log-likelihood of the sample. This is

$$\sum_{iy} \hat{\mu}_{iy} \log \frac{\exp\left((\Phi\beta)_{iy}\right)}{\sum_y \exp\left((\Phi\beta)_{iy}\right)}$$

which is

$$\sum_{iy} \hat{\mu}_{iy} \left((\Phi\beta)_{iy} - \log \sum_y \exp\left((\Phi\beta)_{iy}\right) \right)$$

first order condition is

$$\sum_{iy} \hat{\mu}_{iy} \Phi_{iy,k} = \sum_{iy} \frac{\exp\left((\Phi\beta)_{iy}\right)}{\sum_y \exp\left((\Phi\beta)_{iy}\right)} \Phi_{iy,k}$$

4.5 Reminders on Poisson regression

Assume that $\mu_{iy}|\Phi$ follows a Poisson distribution of parameter $\theta_{iy} = e^{(\Phi\beta)_{iy}}$.

$$\Pr\left(\mu_{iy}|\Phi_{iy}\right) = \frac{\theta_{iy}^{\mu_{iy}}}{\mu_{iy}!} e^{-\theta_{iy}}$$

Let's compute the log-likelihood of z. We have

$$l\left(\hat{\mu}_{iy}|\Phi_{iy}\right) = \hat{\mu}_{iy} \log \theta_{iy} - \theta_{iy} - \log\left(\hat{\mu}_{iy}!\right)$$

that is

$$l_{iy}\left(\hat{\mu}_{iy}|\Phi_{iy}\right) = \hat{\mu}_{iy} \left(\Phi\beta\right)_{iy} - e^{(\Phi\beta)_{iy}} - \log\left(\hat{\mu}_{iy}!\right)$$

Hence the conditional log-likelihood of the sample is

$$\sum_{iy} \hat{\mu}_{iy} \left(\Phi\beta\right)_{iy} - \sum_{iy} e^{(\Phi\beta)_{iy}} - \sum_{iy} \log\left(\hat{\mu}_{iy}!\right)$$

therefore the Poisson regression consists of the MLE for Poisson distribution that is

$$\max_{\beta} \left\{ \sum_{iy} \hat{\mu}_{iy} \left(\Phi\beta\right)_{iy} - \sum_{iy} e^{(\Phi\beta)_{iy}} \right\}$$

Question: How to reformulate a logistic regression as a Poisson regression?

Start from the Poisson regression and introduce an individual fixed effect. $\sum_k \Phi_{iyk}\beta_k$ is now replaced by $\sum_k \Phi_{iyk}\beta_k - u_i$, and the parameter β is replaced by (β, u).

The Poisson regression now becomes

$$\max_{\beta_k, u_i} \left\{ \sum_{iy} \hat{\mu}_{iy} \left(\sum_k \Phi_{yk}\beta_k - u_i \right) - \sum_{iy} e^{(\Phi\beta)_y - u_i} \right\}$$

Now we have to maximize over β_k the following expression

$$\max_{u_i} \left\{ \sum_{iy} \hat{\mu}_{iy} \left(\sum_k \Phi_{yk} \beta_k - u_i \right) - \sum_{iy} e^{(\Phi\beta)_y - u_i} \right\}$$

that is

$$\sum_{iyk} \hat{\mu}_{iy} \Phi_{yk} \beta_k + \max_{u_i} \left\{ \sum_{iy} \hat{\mu}_{iy} u_i - \sum_{iy} e^{(\Phi\beta)_{iy} + u_i} \right\}$$

by first order conditions, we have

$$\sum_y \hat{\mu}_{iy} = \sum_y e^{(\Phi\beta)_{iy} + u_i}$$

that is

$$1 = \sum_y e^{(\Phi\beta)_{iy} + u_i}$$

hence $u_i = -\log \left(\sum_y e^{(\Phi\beta)_{iy}} \right)$ and the expression becomes

$$\sum_{iyk} \hat{\mu}_{iy} \Phi_{yk} \beta_k + \max_{u_i} \left\{ \sum_{iy} \hat{\mu}_{iy} u_i - \sum_{iy} e^{(\Phi\beta)_y + u_i} \right\}$$

$$\max_{\beta} \left\{ \sum_k \sum_y \beta_k \hat{q}_y \Phi_{yk} - \log \sum_y \exp \left((\Phi\beta)_y \right) \right\}$$

Hence logistic regression = Poisson regression + consumer fixed effect.