DATA AND COMPUTATION 2 STUDY GUIDE

AMY RHEE

1. Logistic Regression

Motivation

Suppose we have a consumer facing some options $j \in \mathcal{J}_0 = \mathcal{J} \cup \{0\}$, set of options including the default option of choosing nothing, $\{0\}$.

Let U be the vector of systematic utilities for our consumer, where each entry corresponds to each option j, and let ε be the vector of utility shocks for our consumer per j.

The preferred option is the one that attains the maximum in

$$\max_{j \in \mathcal{J}} \{U_j + \varepsilon_j, \varepsilon_0\}$$

Choice Probabilities

Let s_j be the probability of choosing option j where

$$s_j = \sigma_j(U) = \mathbb{P}[U_j + \varepsilon_j \geq U_z + \varepsilon_z \text{ for all } z \in \mathcal{J}_0]$$

Note that $s \equiv \sigma(U)$ is a vector of choice probabilities such that its components are positive[1] and sum to one, i.e. $\sum_{j \in \mathcal{J}_0} s_j = 1$.

Some properties of choice probabilities:

(1) $\sigma_j(U)$ is increasing in U_j

(2) $\sigma_j(U)$ is weakly decreasing for $U_{j'}$ for $j' \neq j$

(3) For a constant c, we can replace one component U with U_j and still have $\sigma(U + c) = \sigma(U) \implies$ We can normalize the utility of one of the alternatives j, e.g. the default option, and thus have $U_0 = 0$.

The last property's consequence means that we can only consider σ as a mapping from $\mathbb{R}^{|\mathcal{J}|}$ to $\{s_j\}_{j \in \mathcal{J}}$ and recover the default choice probability by $s_0 = 1 - \sum_{j \in \mathcal{J}} s_j$.

Furthermore, define the expected indirect utility of consumers

$$\begin{aligned} G(U) &= \mathbb{E}[\max_{j \in \mathcal{J}} \{U_j + \varepsilon_j, \varepsilon_0\}] \\ \implies \frac{\partial G}{\partial U_j} &= \mathbb{P}[U_j + \varepsilon_j \geq U_z + \varepsilon_z \text{ for all } z \in \mathcal{J}_0] = s_j \\ \implies \sigma(U) &= \nabla G(U) \end{aligned}$$

Gumbel Distribution

The CDF is

$$F(z) = \exp(-\exp(-x + \gamma)), \text{ where } \gamma = 0.5772... \text{ (Euler's constant)}$$

Properties of Gumbel:

(1) The mean of the Gumbel distribution is 0 (since we've shifted by Euler's constant)

(2) $\varepsilon_1, ..., \varepsilon_n \sim_i id$ Gumbel $\implies \max\{u_i + \varepsilon_i\}$ has the same distribution as $\log(\sum_{i=1}^n \exp u_i) + \varepsilon$, where $\varepsilon \sim$ Gumbel

(3) Standard Gumbel: $F(\exp(-\exp(-x))$ with mean γ

Date: March 27, 2023.

[1] Probability of being indifferent between two options is zero, since the distribution of ε is continuous, so we have a strict inequality

Social Welfare Function & Choice Probabilities

Since utility shocks ε have a Gumbel distribution, we can rewrite the expected indirect utility of consumers (McFadden's surplus function) as:

$$G(U) = \log\left(1 + \sum_{j\in\mathcal{J}} \exp(U_j)\right)$$

As a result, the choice probability of alternative j is proportional to the exponential of the systematic utility associated with U, i.e.

$$\sigma_j(U) = \frac{\exp U_j}{1 + \sum_{j'\in\mathcal{J}} \exp(U_{j'})}$$

Suppose the random utility shock is scaled by a factor T. Then, the choice probability takes the form of the soft-max operator, i.e.

$$\sigma_j(U) = \frac{\exp U_j/T}{1 + \sum_{j'\in\mathcal{J}} \exp(U_{j'}/T)}$$

And as $T \to 0,$ we have $\sigma_j(U) \to \max_{j\in\mathcal{J}}\{U_j, 0\}$.

Logistic Regression Suppose individual i's utility to decision j is pinned down by

$$U_{ij} = \sum_k \phi_{ij}^k \lambda_k + \varepsilon_{ij}$$

where $\varepsilon_{ij} \sim_{iid}$ Gumbel. Then, the conditional probability that individual i's optimal choice is j is

$$\pi_{ij} = \frac{\exp(\sum_k \phi_{ij}^k \lambda_k + \varepsilon_{ij})}{\sum_{j'} \exp\left(\sum_k \phi_{ij}^k \lambda_k + \varepsilon_{ij}\right)}$$

And therefore, the (population) conditional log-likelihood associated with j is

$$\begin{aligned} l_{ij}(\lambda) &= \log \pi_{ij} \\ &= \sum_k \phi_{ij}^k \lambda_k - \log \sum_{j'} \exp\left(\sum_k \phi_{ij}^k \lambda_k\right) \end{aligned}$$

Note that this is simply logistic regression!

Logistic Regression as MLE

Denote $J(i)$ as the actual choice of i and $\hat{\pi}_{ij} = \mathbf{1}\{j = J(i)\}$. In words, $\hat{\pi}_{ij}$ takes value 1 if individual i chose option j (so when $j = J(i)$), and 0 otherwise. Note that $\sum_j \hat{\pi}_{ij} = 1 \ \forall i$, and $\sum_i \hat{\pi}_{ij}$ is the number of individuals who chose option j.

The sample log-likelihood is then

$$\begin{aligned} l(\lambda) &= \sum_i l_{i,J(i)}(\lambda) \\ &= \hat{\pi}^T \Phi\lambda - \sum_i \log \sum_{j'} \exp\left((\Phi\lambda)_{ij'}\right) \\ \implies \frac{\partial l}{\partial \lambda} &= (\hat{\pi} - \pi^{(\lambda)})^T \Phi = 0, \text{ where } \pi_{ij}^{\lambda} = \frac{\exp(\Phi\lambda)_{ij}}{\sum_{j'} \exp(\Phi\lambda)_{ij}} \end{aligned}$$

We can rewrite the FOC as equating predicted moments (LHS) and observed moments (RHS), i.e.

$$\sum_{ij} \pi_{ij}^{\lambda} \Phi_{ij}^k = \sum_i \hat{\pi}_{ij} \Phi_{ij}^k$$

Logistic regression can be cast as the same MLE problem

$$\max_{\lambda} \{\hat{\pi}^T \Phi\lambda - \sum_i \log \sum_{j'} \exp(\Phi\lambda)_{ij'}\}$$

Thus, the FOC is also $\sum_{ij} \pi_{ij}^{\lambda} \Phi_{ij}^k = \sum_i \hat{\pi}_{ij} \Phi_{ij}^k$.

2. GLM

Motivation

GLM encompasses making predictions on the conditional mean of a random variable y, i.e. $\mathbb{E}[y|x]$, given an explanatory random vector x.

$$g(\mathbb{E}[y|x]) = x^T\beta$$
$$\implies \mathbb{E}[y|x] = g^{-1}(x^T\beta)$$

We denote $g : \mathbb{R} \to \mathbb{R}$ as the link function, and it is always increasing and continuous (i.e. invertible).

We often also specify

$$Var(y|x) = \mathbb{V}(g^{-1}(x^T\beta))$$

We want to estimate β, so

$$\mathbb{E}[y|x] = g^{-1}(x^T\beta)$$
$$\mathbb{E}[xy|x] = xg^{-1}(x^T\beta)$$
$$\mathbb{E}[xy] = \mathbb{E}[xg^{-1}(x^T\beta)]$$

Define auxilary function

$$F(z) = \int_0^z g^{-1}(t)dt \implies F'(z) = g^{-1}(z)$$

Then,

$$\mathbb{E}[xy] = \mathbb{E}[xF'(x^T\beta)]$$
$$\implies \frac{\partial}{\partial\beta_k}\mathbb{E}[xy] = \mathbb{E}[\frac{\partial}{\partial\beta_k}F(x^T\beta)], \text{ since } \mathbb{E}[xF'(x^T\beta)] = \nabla_\beta\mathbb{E}[F(x^T\beta)]$$
$$\implies \frac{\partial}{\partial\beta_k}\mathbb{E}[xy] = \mathbb{E}[g^{-1}(x^T\beta)x_k]$$

We can interpret the last line as a FOC, meaning that we're looking for $\beta \in \mathbb{R}^k$ such that

$$\nabla_\beta\{\mathbb{E}[x^T\beta y] - \mathbb{E}[F(x^T\beta)]\} = 0$$

We are maximizing a concave function, since $F = g^{-1}$ is increasing in z, meaning that F is convex and thus $-\mathbb{E}[F(x^T\beta)]$ is concave.

Therefore,

$$\beta^* = \arg\max_{\beta\in\mathbb{R}^k}\{\mathbb{E}[x^T\beta y] - \mathbb{E}[F(x^T\beta)]\}$$

Our objective function is a generalized linear model.

Example 1: OLS

In least squares, our model is $y = x^T\beta + \varepsilon$, where $\mathbb{E}[\varepsilon|x] = 0$. Then, our link function is $g(z) = z$. In addition, assuming $\mathbb{E}[\varepsilon^2|x] = \sigma^2$, we have that $Var(y|x) = \sigma^2$.

3. Poisson Regression

Let $z \sim Pois(\theta)$, where $\theta \in (0, \infty)$, and $\mathbb{E}[z] = \mathbb{V}(z) = \theta$ for $z = 1, 2, 3, ...$

The Poisson PMF is

$$f(z) = \pi_{z|\theta} = \frac{\exp(-\theta)\theta^z}{z!}$$

Furthermore, suppose that conditional on x, r.v. y has a Poisson distribution with parameter $\theta = \exp(x^T\beta)$. Then,

$$\mathbb{E}[y|x] = \exp(x^T\beta) \implies g(z) = \ln(z) \text{ and } F(z) = \int_{-\infty}^z g^{-1}(t)dt = \exp(z)$$

And we get $Var(y|x) = \exp(x^T\beta)$.

A GLM with the log-link function is

$$\max_{\beta\in\mathbb{R}^k} \mathbb{E}[xy]^T\beta - \mathbb{E}[\exp(x^T\beta)]$$

Maximizing the sample likelihood with respect to β thus yields Poisson regression form

$$\sum_i -\exp(x_i^T\beta) - x_i^T\beta y_i - \ln(y_i!)$$
$$\frac{\partial}{\partial\beta}\sum_i -\exp(x_i^T\beta) - x_i^T\beta y_i - \ln(y_i!) = 0$$
$$\implies \sum_i (y_i - \exp(x_i^T\beta))x_i = 0$$

Define X as the $N \times J$ matrix of stacked rows x_i^T on top of each other. Then, our maximization problem is

$$\max_\beta y^T X\beta - 1^T \exp(X\beta)$$

And furthermore, if we define $\bar{y} = \exp(X\beta)$ as the predictor of y, then we have

$$\sum_i y_i X_{ik} = \sum_i \bar{y}_i X_{ik} \text{ for all } k$$
$$\mathbb{E}[y_i x_i] = \mathbb{E}[\bar{y}_i x_i]$$

4. GLM with Fixed Effects

Logistic regression = Poisson regression (GLM) + i-level fixed effects

Recall that the parameter vector λ from a logistic model solves

$$\max_\lambda \hat{\pi}^T\Phi\lambda - \sum_i \log\sum_{j'}\exp(\Phi\lambda)_{ij'}$$

This is almost a GLM, but the problem is the $log(\cdot)$. To make the connection between Logistic regression and GLM, we need to add individual i-fixed effects.

Recall that Poisson regression's problem is

$$\text{Population value: } \max_{\beta\in\mathbb{R}^k}\mathbb{E}[x^T\beta y] - \mathbb{E}[\exp(x^T\beta)]$$
$$\text{Sample analogue: } \max_\beta \frac{1}{n}\sum_{i=1}^n x_i^T\beta y_i - \frac{1}{n}\sum_{i=1}^n \exp(x_i^T\beta)$$
$$\implies \max_\lambda \sum_{i=1}^n\sum_{j\in\mathcal{J}}\sum_{k=1}^K \hat{\pi}_{ij}\phi_{ij}^k\lambda_k - \sum_{i=1}^n\sum_{j\in\mathcal{J}}\exp\phi_{ij}^{(\lambda)}$$

To simplify notation, note that we've defined: $\phi_{ij}^{(\lambda)} = \sum_k \phi_{ij}^k\lambda_k$.

Suppose we add $i-$dummy variables to regressors ϕ_{ij}. Then, our problem is

$$\max_{\lambda,(u_i)}\sum_{i=1}^n\sum_{j\in\mathcal{J}}\hat{\pi}_{ij}(\phi_{ij}^{(\lambda)} - u_i) - \sum_{i=1}^n\sum_{j\in\mathcal{J}}\exp(\phi_{ij}^{(\lambda)} - u_i)$$

And corresponding optimality condition for u_i is

$$-\sum_j \hat{\pi}_{ij} + \sum_j \exp(\phi_{ij}^{(\lambda)} - u_i) = 0$$
$$-1 + \sum_j \exp(\phi_{ij}^{(\lambda)} - u_i) = 0$$
$$\sum_j \exp(\phi_{ij}^{(\lambda)} - u_i) = 1$$
$$\sum_j \exp(\phi_{ij}^{(\lambda)} = \exp(u_i)$$
$$\implies u_i^* = \log\sum_j \exp\phi_{ij}^{(\lambda)}$$

We interpret the last line as the expected utility of the consumer.

Once we substitute the optimal $u_i^* = \log \sum_j \exp \phi_{ij}^{(\lambda)}$ into our original problem, the problem becomes MLE for the multinomial logit model:

$$
\begin{aligned}
&\max_{\lambda,u} \sum_{ij} \hat{\pi}_{ij}(\Phi\lambda)_{ij} - \sum_i u_i \\
\iff &\max_{\lambda} \sum_{ij} \hat{\pi}_{ij}(\Phi\lambda)_{ij} - \sum_i \log \sum_j \exp(\Phi\lambda)_{ij} \\
\iff &\max_{\lambda} \sum_{ij} \hat{\pi}_{ij}\phi_{ij}^{(\lambda)} - \sum_i \log \sum_j \exp(\Phi\lambda)_{ij} \\
&\implies -\sum_j \hat{\pi}_{ij} + \sum_j \exp(\phi_{ij}^{(\lambda)}) = 0 \\
&\implies 1 = \sum_j \exp(\phi_{ij}^{(\lambda)}) \equiv \sum_j \exp(\pi_{ij}^{(\lambda)})
\end{aligned}
$$

Let's rewrite the Poisson regression with different notation. Let $l \in \{1, ..., N\}$ denote our sample. We define the following new notation: $l \iff (i,j)$; $n \iff n \times J$; $y \iff \hat{\pi}_{ij}$; $\beta \iff (\lambda_k, u_i)$; $X\beta \iff X_1\lambda + X_2 u$. Note that X_1 is an $nJ \times K$ matrix, where $(X_1)_{ij,k} = \phi_{ij}^k$ regressor, and X_2 is an $nJ \times n$ matrix, where $(X_2 u)_{ij} = -u_i$.

$$
\frac{1}{N}\sum_l x_l^T \beta y_l - \frac{1}{N}\sum_l \exp(x_l^T\beta)
$$

How do we find what matrix X_2 is?

$$
\begin{aligned}
vec_R(u\mathbf{1}_J^T) &= vec_R\left(\begin{bmatrix} u_1 & \cdots & u_1 \\ \vdots & \ddots & \vdots \\ u_n & \cdots & u_n \end{bmatrix}\right) \\
vec_R(I_n u\mathbf{1}_J^T) &= vec_R\left(\begin{bmatrix} u_1 & \cdots & u_1 \\ \vdots & \ddots & \vdots \\ u_n & \cdots & u_n \end{bmatrix}\right) \\
vec_R(I_n u\mathbf{1}_J^T) &= (I_n \otimes \mathbf{1}_J) vec_R(u) \\
\implies X_2 &= -I_n \otimes \mathbf{1}_J
\end{aligned}
$$

Therefore, we define the big matrix X blockwise:

$$
\begin{aligned}
X &= [\Phi \quad (-I_n \otimes \mathbf{1}_J)] \\
\implies X\beta &= vec\left([(\Phi\lambda)_{ij} - u_i]_{ij}\right)
\end{aligned}
$$

Poisson regression of $\hat{\pi}_{ij}$ on X yields

$$
\begin{aligned}
&\max_{\lambda,u}\{-\sum_{ij}\exp\left((\Phi\lambda)_{ij} - u_i\right) + \sum_{ij}\hat{\pi}_{ij}\left((\Phi\lambda)_{ij} - u_i\right)\} \\
\implies &\max_{\lambda,u}\{-\sum_{ij}\exp\left((\Phi\lambda)_{ij} - u_i\right) + \sum_{ij}\hat{\pi}_{ij}(\Phi\lambda)_{ij} - \sum_i u_i\}
\end{aligned}
$$

5. Vectorization and Kronecker Products

Vectorization: How we represent an $n \times n$ matrix $(\Phi\lambda)_{ij}$ as a vector of dimension $\mathbb{R}^{n^2}$

Row-Major Order: $vec_R\left(\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}\right) = \begin{bmatrix} M_{11} \\ M_{12} \\ M_{21} \\ M_{22} \end{bmatrix}$

– **e.g.** $vec_R\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

Column-Major Order: $vec_C\left(\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}\right) = \begin{bmatrix} M_{11} \\ M_{21} \\ M_{12} \\ M_{22} \end{bmatrix}$

Kronecker Product: Let A be an $n \times m$ matrix, B be a $p \times q$ matrix. Then, the Kronecker product $A \otimes B$ is an $np \times mq$ matrix defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & \ddots & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & \cdots & a_{mn}B \end{bmatrix}$$

Fundamental Vectorization Identity: $vec_R(AXB) = (A \otimes B^T)vec_R(X)$

Transpose with Kronecker Products: $A^T \otimes B^T = (A \otimes B)^T$

6. Linear Optimization Basics

Linear Programming Problem

Let $c \in \mathbb{R}^n, d \in \mathbb{R}^m,$ and $A \in \mathbb{R}^{m \times n}$. We define the following as the **primal** linear programming problem

$$V_p = \max_{x \in \mathbb{R}^n_+ c^T x} \text{ s.t. } Ax = d$$

The set of x's that satisfy the constraint are called feasible solutions, i.e. $\{x \in \mathbb{R}^n_+ : Ax = d\}$. The set of solutions to the problem are called optimal solutions.

Primal $\iff$ Dual

We can recast the primal problem as

$$V_p = \max_{x \in \mathbb{R}^n_+} c^T x + L_p(d - Ax)$$

where $L_p(\cdot)$ is a penalty function s.t.

$$\begin{aligned} L_p(z) &= 0 \text{ if } z = 0 \\ &= -\infty \text{ else} \end{aligned}$$

The simplest choice of such penalty function is given by $L_p(z) = \min_{y \in \mathbb{R}^m}\{z^T y\}$. Adopting this penalty function, we then have

$$V_p = \max_{x \in \mathbb{R}^n, x \geq 0} \min_{y \in \mathbb{R}^m} \{c^T x + (d - Ax)^T y\}$$

By the **minimax inequality** $\max_x \min_y \leq \min_y \max_x$

$$\begin{aligned} \implies V_p &\leq \min_{y \in \mathbb{R}^m} \max_{x \in \mathbb{R}^n, x \geq 0} \{c^T x + (d - Ax)^T y\} = \min_{y \in \mathbb{R}^m} \max_{x \in \mathbb{R}^n, x \geq 0} \{x^T(c - A^T y) + d^T y\} \\ &\leq \min_{y \in \mathbb{R}^m} \{d^T y + L_d(c - A^T y)\} \equiv V_d \end{aligned}$$

where $L_d(\cdot)$ is a penalty function s.t.

$$\begin{aligned} L_d(z) &= 0 \text{ if z} = 0 \\ &= +\infty \text{ else} \end{aligned}$$

Therefore, we define the following as the **dual program**

$$V_d = \min_{y \in \mathbb{R}^m} d^T y \text{ s.t. } A^T y \geq c$$

Duality Theorem

The value of the primal $V_p = \max_{x \in \mathbb{R}^n_+} c^T x + L_p(d - Ax)$ and the value of the dual $V_d = \min_{y \in \mathbb{R}^m}\{d^T y + L_d(c - A^T y)\}$ have the following properties:

(1) V_p and V_d satisfy **weak duality**: $V_p \leq V_d$

(2) If the primal or the dual program have an optimal solution, then both programs have an optimal solution and **strong duality** holds: $V_p = V_d$

(3) If <u>both</u> primal and dual are feasible, then $V_p = V_d < \infty$ (finite solutions)

(4) If $x^* \in \mathbb{R}^n_+$ is an optimal primal solution, and $y^* \in \mathbb{R}^m$ is an optimal dual solution, then **complementary slackness** holds: $x_i^* > 0 \implies (A^T y^*)_i = c_i$

Corollary: If q is a feasible primal solution and y is a feasible dual solution, then q is an optimal primal solution and y is an optimal dual solution $\iff$ complementary slackness holds, i.e. $q_j > 0$ and $y_i > 0 \implies$ constraints bind.

Examples

(1) Consider the problem

$$\min_{q\geq 0} \sum_k q_j \text{ s.t. } Nq \geq r$$

Assume that $n = 1, J = 1, N = -1,$ and $r = 1$. Then, our problem is not feasible. Now, consider the dual

$$\max_{y\geq 0} y \text{ s.t. } Ny \geq 1$$

The dual is feasible, meaning that $V_p = V_d$, even though the primal is not feasible.

$$\implies V_p = V_d = +\infty$$

In general, minimizing or maximizing over the empty set yields (by limit argument)

$$\min_{\emptyset} = +\infty \text{ and } \max_{\emptyset} = -\infty$$

(2)

7. Optimal Assignment and Becker Model

Motivation

Consider the problem of assigning workers to firms. Each worker should work for one firm, and each firm should hire one worker.

Let n_x denote the number of workers of type $x \in X$, and let m_y denote the number of firms (jobs) of type $y \in Y$, where both X and Y are finite (discrete) sets. Assume that $\sum_x n_x = \sum_y m_y$, i.e. that the total number of workers = total number of jobs. Furthermore, assume that the total number of workers and jobs is normalized to one such that $\sum_x n_x = \sum_y m_y = 1$. Let ϕ_{xy} be the total output generated if a worker x matches with a firm y.

We want to know:

(1) Optimality: What is the optimal assignment that maximizes total output generated?

(2) Eq'm: In equilibrium, how many workers of each type are assigned to firms of each type? What are the equilibrium wages?

Define μ_{xy} as the mass of workers of type x assigned to type y firms. Assume that all jobs are filled and all workers are assigned, i.e. $\sum_x \mu_{xy} = m_y$ and $\sum_y \mu_{xy} = n_x$. Then, total output is $\sum_{xy} \mu_{xy}\phi_{xy}$.

Primal $\iff$ Dual

The social planner wants to maximize total output

$$\begin{aligned} &\max_{\mu\geq 0} \sum_{xy} \mu_{xy}\phi_{xy} \\ &\text{s.t. } \sum_x \mu_{xy} = m_y \quad [u_x] \\ &\sum_y \mu_{xy} = n_x \quad [v_y] \end{aligned}$$

The quantities in $[\cdot]$ denote lagrange multipliers corresponding to each constraint.

We can rewrite the primal problem in $\max_\mu \min_{u,v}$ form:

$$\max_{\mu\geq 0}\{\sum_{xy} \mu_{xy}\phi_{xy} + \min_{u,v}\{\sum_x u_x(n_x - \sum_y \mu_{xy}) + \sum_y v_y(m_y - \sum_x \mu_{xy})\}\}$$

The corresponding dual problem is

$$\begin{aligned} &\min_{u,v} \sum_x n_x u_x + \sum_y m_y v_y \\ &\text{s.t. } u_x + v_y \geq \phi_{xy} \quad [\mu_{xy} \geq 0] \end{aligned}$$

Computing Optimal Assignment Problem

Primal:

$$\max_{\mu\geq 0} \Phi^T \mu \text{ s.t. } M\mu = \begin{bmatrix} n \\ m \end{bmatrix}$$

Dual:

$$\min_{u,v} \sum_x u_x n_x + \sum_y v_y m_y \text{ s.t. } M^T \begin{bmatrix} u \\ v \end{bmatrix} \geq \Phi$$

How to find M? First, let's vectorize our constraints for the primal problem. If we expand out the constraints from matrix form, we get

$$\sum_x \mu_{xy} = m_y \implies \mathbf{1}_X^T \cdot \mu = m^T \implies (\mathbf{1}_X^T \otimes I_Y) vec(\mu) = vec(m)$$
$$\sum_y \mu_{xy} = n_x \implies \mu \cdot \mathbf{1}_Y = n \implies (I_X \otimes \mathbf{1}_Y^T) vec(\mu) = vec(n)$$

Therefore,

$$M = \begin{bmatrix} (I_X \otimes \mathbf{1}_Y^T) \\ (\mathbf{1}_X^T \otimes I_Y) \end{bmatrix}$$

Note that since we have an equality constraint for the primal problem, there is no sign restriction for the Lagrange multiplier $p = [u \;\; v]^T$, and thus there should be no sign constraint on the decision variables (u, v) for the dual problem.

Feasibility

$$\mu_{xy} = \frac{n_x m_y}{\sum n_x}$$
$$\sum_y \mu_{xy} = n_x \frac{\sum_y m_y}{\sum_x n_x} = n_x$$
$$\sum_x \mu_{xy} = m + y \frac{\sum_x n_x}{\sum_x n_x} = m_y$$

Duality Theorem: Discrete Case Consider feasible primal solution $\mu_{xy} > 0$ such that $\sum_y \mu_{xy} = n_x$ and $\sum_x \mu_{xy} = m_y$. Consider feasible dual solution (u, v) such that $u_x + v_y \geq \phi_{xy}$ for all x, y.

(1) The value of the primal (social planner's) problem μ coincides with the value of the dual problem (u, v)

(2) μ is a primal solution and (u, v) is a dual solution $\iff$ (by complementary slackness) $\mu_{xy} > 0 \implies u_x + v_y = \phi_{xy}$

Interpreting Equilibrium

Suppose that if x and y match, then worker gets utility α_{xy} and firm gets utility γ_{xy}. Then, we define total surplus as $\phi_{xy} = \alpha_{xy} + \gamma_{xy}$. We define the worker's problem and firm's problem separately as follows

$$\text{Worker: } \max_y \alpha_{xy} \implies x \text{ chooses } y \text{ s.t. } y \in \arg\max_y \alpha_{xy}$$
$$\text{Firm: } \max_x \gamma_{xy} \implies y \text{ chooses } x \text{ s.t. } x \in \arg\max_x \gamma_{xy}$$

In the decentralized model with two agents, we need a separate market clearing condition for equilibrium. In order to facilitate market clearing, we need prices, namely wages w_{xy}.

Thus, we recast the problems by adding wage variable to allow for transferable utility. Note that total surplus remains the same because wages are paid out by the firms and collected by the workers.

$$\text{Worker: } \max_y \alpha_{xy} + w_{xy} \equiv u_x$$
$$\text{Firm: } \max_x \gamma_{xy} - w_{xy} \equiv v_y$$

Recall that u_x, v_y are indirect utilities, i.e. utilities associated with optimal income. Note that if x, y are matched, then $u_x = \alpha_{xy} + w_{xy}$ and $v_y = \gamma_{xy} - w_{xy}$.

Equilibrium Conditions

Let μ_{xy} be the mass of xy pairs formed in equilibrium.

(1) Market clears: $\sum_y \mu_{xy} = n_x$ and $\sum_x \mu_{xy} = m_y$

(2) Everyone optimizes: If x and y are matched, i.e. $\mu_{xy} > 0$, then $y \in \arg\max_y \alpha_{xy} + w_{xy}$ and $x \in \arg\max_x \gamma_{xy} - w_{xy}$. We can rewrite this condition as

$$u_x \geq \alpha_{xy} + w_{xy}, \text{ w/ equality if } \mu_{xy} > 0$$
$$v_y \geq \gamma_{xy} - w_{xy}, \text{ w/ equality if } \mu_{xy} > 0$$

Eq'm Conditions $\implies$ Complementary Slackness

The second equilibrium condition implies

$$\begin{aligned} & u_x \geq \alpha_{xy} + w_{xy}, \text{ w/ equality if } \mu_{xy} > 0 \\ & v_y \geq \gamma_{xy} - w_{xy}, \text{ w/ equality if } \mu_{xy} > 0 \\ \implies & u_x + v_y \geq \alpha_{xy} + \gamma_{xy} = \phi_{xy} \\ \implies & u_x + v_y = \phi_{xy} \text{ if } \mu_{xy} > 0 \end{aligned}$$

Therefore, by the complementary slackness theorem (duality theorem: discrete case), we know that the optimal value of the primal problem is equivalent to the optimal value of the dual problem.

Note that μ_{xy} is not unique. We could shift our multipliers $u_x \to u_x + c$ and $v_y \to v_y - c$ and we would get the same solution.

Complementary Slackness $\implies$ Eq'm (Market-Clearing Wage)

Assume that $\mu, (u, v)$ are optimal solutions to the primal and dual problems respectively. Then, by duality and complementary slackness, the following hold

(1) $\sum_y \mu_{xy} = n_x$ and $\sum_x \mu_{xy} = m_y$

(2) $u_x + v_y \geq \alpha_{xy} + \gamma_{xy}$

(3) $\mu_{xy} > 0 \implies u_x + v_y = \alpha_{xy} + \gamma_{xy}$

Since $\alpha_{xy} + w_{xy} \leq u_x$ and $\gamma_{xy} - w_{xy} \leq v_y$ by definition of indirect utilities, our requirements for the equilibrium wage are

(1) $w_{xy} \leq u_x - \alpha_{xy}$

(2) $w_{xy} \geq \gamma_{xy} - v_y$

This implies that we w_{xy} is s.t.

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

Note that the wage is not necessarily unique. Any wage that satisfies this inequality is an equilibrium wage. However, $\mu_{xy} > 0 \implies$ the interval is a singleton, i.e. existence of a unique equilibrium wage. Stated differently, given (u, v), the equilibrium wage w_{xy} is defined on the equilibrium path, i.e. when x, y are s.t. $\pi_{xy} > 0$. Note that all workers of the same type get the same indirect utility, but not necessarily the same wage, since workers of a given type x can match with different firms y.

To show that there exists an equilibrium wage w_{xy} that satisfies this constraint, we need to show that there exists an $x \in [\gamma_{xy} - v_y, u_x - \alpha_{xy}]$.

$$\begin{aligned} \text{In equilibrium, } & u_x + v_y \geq \phi_{xy} = \alpha_{xy} + \gamma_{xy} \\ & \gamma_{xy} - v_y \leq u_x - \alpha_{xy} \\ \mu_{xy} > 0 \implies & \gamma_{xy} - v_y = u_x - \alpha_{xy} \\ \implies & [\gamma_{xy} - v_y, u_x - \alpha_{xy}] = \{w^*_{xy}\} \text{ by definition of } u_x, v_y \end{aligned}$$

8. Optimization and Gradient Descent

$$\min_{a_x, b_y} \sum_{xy} \exp\left(\frac{\phi_{xy}}{2} - a_x - b_y\right) + \sum_x n_x a_x + \sum_y m_y b_y \equiv \min_{a_x, b_y} F(a, b)$$

Computation methods:

(1) Gradient Descent

$$a_x^{t+1} = a_x^t - \varepsilon \frac{\partial F}{\partial a_x}$$
$$b_y^{t+1} = b_y^t - \varepsilon \frac{\partial F}{\partial b_y}$$

(2) Coordinate Descent: F convex and smooth $\implies$ coordinate descent converges

$$b^{2t+1} \leftarrow \min_{b'} F(a^{2t}, b')$$
$$a^{2t+2} \leftarrow \min_{a'} F(a', b^{2t+1})$$

When $b = b^{2t+1}$, this means $\sum_x \exp(\phi_{xy}/2 - a_x - b_y) = m_y$, and when $a = a^{2t+2}$, this means $\sum_y \exp(\phi_{xy}/2 - a_x - b_y) = n_x$

$$\implies b_y = \log \frac{\sum_x \exp \phi_{xy}/2 - a_x - b_y}{m_y}$$
$$\implies a_x = \log \frac{\sum_y \exp \phi_{xy}/2 - a_x - b_y}{n_x}$$