

Maximum Simulated Likelihood

Francesco Ruggieri

The University of Chicago

April 16, 2020

What is Maximum (Non-Simulated) Likelihood?

- $(X_1 \sim P_1, \dots, X_N \sim P_N)$ is a collection of random vectors
 - Not necessarily independent and not necessarily identically distributed
- Each P_i depends on some common parameter vector $\theta \in \Theta$
- For convenience, assume that $(X_1 \sim P_1, \dots, X_N \sim P_N)$ are independent
- The **likelihood function** of θ is the joint distribution of (X_1, \dots, X_N) under θ and evaluated at \mathbf{x} :

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^N P_i(x_i; \theta)$$

- The **log-likelihood function** of θ is

$$LL(\theta|\mathbf{x}) = \log f(\mathbf{x}|\theta) = \sum_{i=1}^N \log P_i(x_i; \theta)$$

What is Maximum (Non-Simulated) Likelihood?

- Then a **maximum likelihood estimator** of θ is

$$\hat{\theta}_{ML} \in \arg \max_{\tilde{\theta}} L(\tilde{\theta} | \mathbf{x}) \iff \hat{\theta}_{ML} \in \arg \max_{\tilde{\theta}} LL(\tilde{\theta} | \mathbf{x})$$

- What if there is **no closed form** for $\{P_i(\cdot; \theta)\}_{i=1}^N$?
- Those masses/densities could be approximated with simulation...

What is Maximum Simulated Likelihood?

- The **simulated likelihood function** of θ is a simulated approximation to the joint distribution of (X_1, \dots, X_N) under θ and evaluated at \mathbf{x} :

$$SL(\theta|\mathbf{x}) = \check{f}(\mathbf{x}|\theta) = \prod_{i=1}^N \check{P}_i(x_i; \theta)$$

where \check{P}_i is a simulated approximation to P_i

- The **simulated log-likelihood function** of θ is

$$SLL(\theta|\mathbf{x}) = \log \check{f}(\mathbf{x}|\theta) = \sum_{i=1}^N \log \check{P}_i(x_i; \theta)$$

- Then a **maximum simulated likelihood estimator** of θ is

$$\hat{\theta}_{MSL} \in \arg \max_{\tilde{\theta}} SL(\tilde{\theta}|\mathbf{x}) \iff \hat{\theta}_{MSL} \in \arg \max_{\tilde{\theta}} SLL(\tilde{\theta}|\mathbf{x})$$

A Simple Example: Mixed Logit

- Consider the indirect utility function with **random coefficients**:

$$V_{ij} = X'_{ij}\beta_i + \varepsilon_{ij} \quad \text{with} \quad \beta_i \sim f(\beta; \theta)$$

- $\varepsilon_{ij} \stackrel{\text{iid}}{\sim}$ T1EV, so individual choice probabilities (conditional on β_i) are

$$P(X_{ij})|\beta_i = \frac{e^{X'_{ij}\beta_i}}{\sum_{k=0}^J e^{X'_{ik}\beta_i}}$$

- β_i is random, so let us integrate it out:

$$P(X_{ij}; \theta) = \int \frac{e^{X'_{ij}\beta_i}}{\sum_{k=0}^J e^{X'_{ik}\beta_i}} f(\beta_i; \theta) d\beta_i$$

A Simple Example: Mixed Logit

- Oh no! That integral may **not** have a **closed form**!
- But wait, if we assumed $f(\beta; \theta)$, we could **simulate** it...
 - ① Draw β^r from $f(\beta; \theta)$
 - ② Compute $P(X_{ij}) | \beta^r$
 - ③ Repeat (1) and (2) R times, with $R \gg 0$
 - ④ Compute the simple average of $\{P(X_{ij}) | \beta^r\}_{r=1}^R$
- A **simulated approximation** to the individual choice probability is

$$\check{P}(X_{ij}; \theta) = \frac{1}{R} \sum_{r=1}^R \frac{e^{X'_{ij}\beta^r}}{\sum_{k=0}^J e^{X'_{ik}\beta^r}}$$

A Simple Example: Mixed Logit

- Almost done...
- Now we can compute the **simulated log likelihood**:

$$SLL(\theta|\mathbf{x}) = \sum_{i=1}^N \sum_{j=0}^J D_{ij} \log \check{P}(X_{ij}; \theta)$$

where $D_{ij} = 1$ if individual i chose good j , and 0 otherwise

- A **maximum simulated likelihood estimator** of θ is

$$\hat{\theta}_{MSL} \in \arg \max_{\tilde{\theta}} SLL(\tilde{\theta}|\mathbf{x})$$

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

- Consider a dynamic program with **finite horizon**, $t \in \{0, \dots, T\}$
- In each period, agents choose **one of K possible alternatives**
- Alternatives are mutually exclusive, so **decisions** are

$$d_k(t) = \begin{cases} 0 & \text{if } k \text{ is not chosen at time } t \\ 1 & \text{if } k \text{ is chosen at time } t \end{cases}$$

- A **current period reward** is associated with choice k at time t , $R_k(t)$
- The **state space** at time t is $S(t)$

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

The **value function** depends on both state space and time horizon:

$$V(S(t), t) = \max \{V_1(S(t), t), \dots, V_K(S(t), t)\}$$

where each **choice value function** obeys the Bellman equation

$$V_k(S(t), t) = \begin{cases} R_k(S(t), t) + \delta \mathbb{E}[V(S(t+1), t+1) | S(t), d_k(t) = 1] & \text{if } t < T \\ R_k(S(t), t) & \text{if } t = T \end{cases}$$

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

- Labor supply is discrete: **zero (0), part-time (1), full-time (2)**
- Per-period reward functions are given by

$$R_0(S(t), t) = \gamma + \varepsilon_{0t}$$

$$R_1(S(t), t) = w_1(x_t) + \varepsilon_{1t}$$

$$R_2(S(t), t) = w_2(x_t) + \varepsilon_{2t}$$

where γ is a constant and w are wage offers

- The **state space** is

$$S(t) = \{x_t, \varepsilon_{0t}, \varepsilon_{1t}, \varepsilon_{2t}\}$$

where x_t denotes experience and ε_t are shocks to the value of non-market time or labor productivity shocks

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

- The **state space evolves** according to

$$x_{t+1} = x_t + 0.5 \times d_1(t) + 1 \times d_2(t)$$
$$f(\varepsilon_{t+1} | S(t), d_k(t)) = f(\varepsilon_{t+1} | x_t, d_k(t))$$

- Working:
 - part time increases the experience stock by 0.5
 - full time increases the experience stock by 1
- We also assume shocks are **serially independent**
 - Issues become even more clear if you relax this assumption

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

- Recall that the value function is

$$V(S(t), t) = \max \{V_0(S(t), t), V_1(S(t), t), V_2(S(t), t)\}$$

- Then, at time $t - 1$ and **for all choices k** , agents must compute the **expected maximum** of the choice-specific value functions

$$\mathbb{E}[\max \{V_0(S(t), t), V_1(S(t), t), V_2(S(t), t)\} | S(t-1), d_k(t-1)]$$

where $\mathbb{E}[\cdot]$ is a **three-variate multiple integral** with respect to the joint distribution of ε

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

- Solve the model by **backward recursion**
- At time T there is no future, so choice is based on per-period reward
- At time $T - 1$, the expected maximum must be computed for all k possible choices and all possible realizations of x_{T-1}
- When you reach time 0, you need to have computed **all possible** choice-specific value functions (*paths* of possible state realizations)!
- This becomes even worse if the ε shocks are not serially independent
- How to **avoid computing** all of these three-variate multiple integrals with no closed form?

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

- Before finding out, one step back...
- **Why are we doing this?**
- Our goal is NOT to solve the model per se, but solve the model in order to **estimate structural parameters**, for instance δ and γ
- In practice, we have **a panel** with **observed labor supply choices**
- If someone with 9 years of experience worked part time in $t = 2007$

$$\mathbb{P}(d_1(t) = 1, w_{1t} | x_t = 9) = \mathbb{P}(w_{1t}, V_1(S(t), t) \geq V_0(S(t), t), V_1(S(t), t) \geq V_2(S(t), t))$$

- This is **one term** in the likelihood function!

MSL and Dynamic Discrete Choice: Keane & Wolpin (1994)

- If we could compute integrals, we would multiply those $\mathbb{P}(\cdot)$ across i and over t , and obtain the maximum likelihood estimator
- But since we cannot, Keane and Wolpin (1994) proposes a method based on **simulation**
 - 1 Take a draw r from the joint distribution of $\varepsilon \equiv (\varepsilon_0, \varepsilon_1, \varepsilon_2)$
 - 2 Calculate $V_0^r(S(t), t), V_1^r(S(t), t), V_2^r(S(t), t)$
 - 3 Pick the maximum among $V_0^r(S(t), t), V_1^r(S(t), t), V_2^r(S(t), t)$
 - 4 Repeat (1)–(3) R times, with $R \gg 0$
 - 5 Compute the simple average of R maximum choice-specific values
 - 6 Perform (1)–(5) in each t and for every possible $S(t)$

Thank you!