Relaxing Conditional Independence

Robert A. Miller

Tilburg University

November 2023

Job Matching and Occupational Choice (Miller, 1984)

Individual payoffs and choices

- Relaxing the conditional independence assumption is tantamount to reopening the multiple integration challenge.
- Consider what happens when unobserved beliefs evolve endogenously.
- The payoff from working job $m \in \{1, 2, \ldots\}$ at time $t \in \{0, 1, \ldots\}$ is:

$$x_{mt} \equiv \psi_t + \xi_m + \sigma_m \epsilon_{mt} \tag{1}$$

  where:
  - ψ_t is a lifecycle trend shaping term that plays no role in the analysis;
  - ξ_m is a job match parameter drawn from $N\left(\gamma_m, \delta_m^2\right)$;
  - ε_{mt} is an idiosyncratic *iid* disturbance drawn from $N(0, 1)$
- Every period t the individual chooses a job $m \in \{1, 2, \ldots\}$ where:
  - $d_t = (d_{1t}, d_{2t}, \ldots)$ denotes her choice, $d_{mt} \in \{0, 1\}$ and $\sum_{m=1}^{\infty} d_{mt} = 1$.
  - her realized lifetime utility is $\sum_{t=0}^{\infty} \sum_{m=1}^{\infty} \beta^t d_{mt} x_{mt}$

- At $t = 0$ the individual sees $\left(\gamma_m, \delta_m^2\right)$ for each m.
- After making her choice, she also sees ψ_t, and $d_{mt} x_{mt}$ for all m.
- Her posterior beliefs for job m at time $t \in \{0, 1, \ldots\}$ are $N\left(\gamma_{mt}, \delta_{mt}^2\right)$ where:

$$\begin{aligned}\gamma_{mt,+1} &= \frac{\delta_m^{-2}\gamma_m + \sigma_m^{-2}\sum_{s=0}^{t}\left(x_{ms} - \psi_s\right) d_{ms}}{\delta_m^{-2} + \sigma_m^{-2}\sum_{s=0}^{t} d_{ms}} \\ &= \gamma_{mt} + \left(x_{mt} - \psi_t\right) / \left(\sigma_m^2 \delta_{mt}^{-2} + 1\right) d_{mt}\end{aligned}$$

and

$$\delta_{m,t+1}^{-2} = \delta_m^{-2} + \sigma_m^{-2}\sum_{s=0}^{t} d_{ms} = \delta_{mt}^{-2} + \sigma_m^{-2} \tag{2}$$

- She maximizes the sum of expected payoffs, sequentially choosing d_t given her beliefs $N\left(\gamma_{mt}, \delta_{mt}^2\right)$ for each $m \in M$.

Corollary (from Theorem 2 in Gittens and Jones,1974)

At each $t \in \{1, 2, \ldots\}$ it is optimal to select the $m \in M$ maximizing:

$$DAI_m(\gamma_{mt}, \delta_{mt}) \equiv \sup_{\tau \geq t} \left\{ \frac{E\left[\sum_{r=t}^{\tau} \beta^{r-t}(x_{mr} - \psi_r) \,|\gamma_{mt}, \delta_{mt}\right]}{E\left[\sum_{r=t}^{\tau} \beta^{r-t} \,|\gamma_{mt}, \delta_{mt}\right]} \right\}$$

- If τ is fixed and there is perfect foresight, the fundamental ratio is:
  - the discounted sum of benefits $\sum_{r=t}^{\tau} \beta^{r-t}(x_{mr} - \psi_r)$
  - divided by the discounted sum of time $\sum_{r=t}^{\tau} \beta^{r-t}$.
- For example if project A yields 5 and takes 2 periods to complete, and B yields 3 but only takes 1 period, do A first if and only if:

$$\begin{aligned} & 5 + 3\beta^2 > 3 + 5\beta \\ \iff\ & 5(1-\beta) > 3(1-\beta)(1+\beta) \\ \iff\ & DAI_A \equiv 5 \big/ (1+\beta) > 3 \equiv DAI_B \end{aligned}$$

Optimization

Bayesian learning with a normal distribution

- Define $D(\sigma)$ is the (standard) DAI for a (hypothetical) job whose
  - fixed match parameter ξ is drawn from $N(0,1)$
  - whose random component in the payoff is $\sigma\varepsilon_t$.

Corollary (Proposition 4 of Miller, 1984)

In this model:

$$DAI_m(\gamma_{mt}, \delta_{mt}) = \gamma_{mt} + \delta_{mt} D\left[\left(\frac{\sigma_m}{\delta_m}\right)^2 + \sum_{s=0}^{t-1} d_{ms}\right]$$

- We can prove $D(\cdot)$ is a decreasing function, implying that $DAI_m(\gamma_{mt}, \delta_{mt}) \uparrow$ as:
  - γ_{mt} and δ_{mt} and $\beta \uparrow$
  - σ_m and $\sum_{s=0}^{t-1} d_{ms} \downarrow$.

Probability Distribution of Spell Lengths

Hazard rate for spell length

- Assuming $(\gamma_m, \delta_m, \sigma_m) = (\gamma, \delta, \sigma)$ for all m, a world in which all differences between jobs are match specific, it suffices to only keep track of the current job match. (Why?)
- Define h_t as the discrete hazard at t periods as the probability a spell ends after t periods conditional on surviving that long. Then:

$$\begin{aligned}
h_t &\equiv \Pr\left\{\gamma_t + \delta_t D\left[\left(\frac{\sigma}{\delta}\right)^2 + t, \beta\right] \leq \gamma + \delta D\left[\left(\frac{\sigma}{\delta}\right)^2, \beta\right]\right\} \\
&= \Pr\left\{\frac{\gamma_t - \gamma}{\sigma} \leq \frac{\delta}{\sigma} D\left[\left(\frac{\sigma}{\delta}\right)^2, \beta\right] - \frac{\delta_t}{\sigma} D\left[\left(\frac{\sigma}{\delta}\right)^2 + t, \beta\right]\right\} \\
&= \Pr\left\{\rho_t \leq \alpha^{-1/2} D(\alpha, \beta) - (\alpha + t)^{-1/2} D(\alpha + t, \beta)\right\}
\end{aligned}$$

where $\rho_t \equiv (\gamma_t - \gamma)/\sigma$ and $\alpha \equiv (\sigma/\delta)^2$ which implies:

$$\frac{\delta_t}{\sigma} = \frac{\left[\delta^{-2} + t\sigma^{-2}\right]^{-1/2}}{\sigma} = \left[\left(\frac{\delta}{\sigma}\right)^{-2} + t\right]^{-1/2} = (\alpha + t)^{-1/2}$$

Probability Distribution of Spell Lengths

Relating the hazard rate to the distribution of normalized match qualities

- Define the probability distribution of transformed means of spells surviving at least t periods as:

$$\Psi_t(\rho) \equiv \Pr\{\rho_t \leq \rho\} = \Pr\left\{\sigma^{-1}(\gamma_t - \gamma) \leq \rho\right\} = \Pr\{\gamma_t \leq \gamma + \rho\sigma\}$$

- To help fix ideas note that $\Psi_0(\rho) = 0$ for all $\rho < 0$ and $\Psi_0(0) = 1$.
- From the definition of h_t and $\Psi_t(\rho)$:

$$\begin{aligned} h_t &= \Pr\left\{\rho_t \leq \alpha^{-1/2} D(\alpha, \beta) - (\alpha + t)^{-1/2} D(\alpha + t, \beta)\right\} \\ &= \Psi_t\left[\alpha^{-1/2} D(\alpha, \beta) - (\alpha + t)^{-1/2} D(\alpha + t, \beta)\right] \end{aligned}$$

- To derive the discrete hazard, we recursively compute $\Psi_t(\rho)$.

Probability Distribution of Spell Lengths

Inequalities relating to normalized match qualities after one period

- By definition every match survives at least one period, and hence:

$$\Psi_1(\rho) = \Pr\{\gamma_1 \leq \gamma + \rho\sigma\}$$

- From the Bayesian updating rule for γ_t:

$$\begin{aligned}
\gamma_1 &\leq \gamma + \rho\sigma \\
&\Leftrightarrow \frac{\delta^{-2}\gamma + \sigma^{-2}(x_1 - \psi_1)}{\delta^{-2} + \sigma^{-2}} \leq \gamma + \rho\sigma \\
&\Leftrightarrow \delta^{-2}\gamma + \sigma^{-2}(\xi + \sigma\epsilon) \leq (\gamma + \rho\sigma)\left(\delta^{-2} + \sigma^{-2}\right) \\
&\Leftrightarrow \alpha\gamma + \xi + \sigma\epsilon \leq (\gamma + \rho\sigma)(\alpha + 1) \\
&\Leftrightarrow (\xi - \gamma) + \sigma\epsilon \leq \sigma(\alpha + 1)\rho \\
&\Leftrightarrow \delta^{-1}(\xi - \gamma) + \alpha^{1/2}\epsilon \leq \alpha^{1/2}(\alpha + 1)\rho
\end{aligned}$$

Probability Distribution of Spell Lengths

Computing the distribution of normalized match qualities after one period

- Since every match survives at least one period, we can calculate $\Psi_1(\rho)$ for all matches:

$$\Psi_1(\rho) \equiv \Pr\{\gamma_1 \leq \gamma + \rho\sigma\} \equiv \Pr\{\rho_1 \leq \rho\}$$

- Appealing to the inequalities from the previous slide:

$$\begin{aligned}
\Psi_1(\rho) &= \Pr\{\gamma_1 \leq \gamma + \rho\sigma\} \\
&= \Pr\left\{\delta^{-1}(\xi - \gamma) + \alpha^{1/2}\epsilon \leq \alpha^{1/2}(\alpha+1)\rho\right\} \\
&= \Pr\left\{\epsilon' + \alpha^{1/2}\epsilon \leq \alpha^{1/2}(\alpha+1)\rho\right\} \\
&= \Pr\left\{(\alpha+1)^{1/2}\epsilon'' \leq \alpha^{1/2}(\alpha+1)\rho\right\} \\
&= \Phi\left[\alpha^{1/2}(\alpha+1)^{1/2}\rho\right]
\end{aligned}$$

where ϵ, ϵ' and ϵ'' are independent standard normal random variables.

Probability Distribution of Spell Lengths

Solving for the one period hazard rate and the probability distribution of survivors

- The spell ends if:

$$\rho_1 < \alpha^{-1/2} D(\alpha, \beta) - (\alpha+1)^{-1/2} D(\alpha+1, \beta) \equiv \rho_1^*$$

- Therefore the proportion of spells ending after one period is:

$$\begin{aligned} h_1 &= \Psi_1 \left[\alpha^{-1/2} D(\alpha, \beta) - (\alpha+1)^{-1/2} D(\alpha+1, \beta) \right] \\ &= \Phi \left\{ \begin{array}{l} \left[\alpha^{1/2} (\alpha+1)^{1/2} \right] \\ \times \left[\alpha^{-1/2} D(\alpha, \beta) - (\alpha+1)^{-1/2} D(\alpha+1, \beta) \right] \end{array} \right\} \\ &> 1/2 \ \ (\text{because } D(\cdot) \text{ is decreasing in } \alpha) \end{aligned}$$

- Intuitively, half the time a (negative) signal reduces the posterior mean below its prior, making a new job with same initial characteristics as the old job look more attractive because:
  - the new job has greater information value.
  - it also has higher mean.

Probability Distribution of Spell Lengths

Recursively computing the distribution of normalized match qualities

- Continuing in this line of reasoning:

$$\Psi_2(\rho) = \frac{\int_{-\infty}^{\infty} \Phi \left[\begin{array}{c} \alpha^{1/2} (\alpha+1)^{1/2} \times \\ \left(\rho - \epsilon \left[(\alpha+1)(\alpha+2) \right]^{-1/2} \right) \end{array} \right] d\Phi(\epsilon) - h_1}{1 - h_1}$$

and more generally (from page 1112 of Miller, 1984):

$$\Psi_{t+1}(\rho) \equiv \frac{\int_{-\infty}^{\infty} \Psi_t \left(\rho - \epsilon \left[(\alpha+t)(\alpha+t+1) \right]^{-1/2} \right) d\Phi(\epsilon) - h_t}{1 - h_t}$$

- Suppose the sample comprises a cross section of spells $n \in \{1, \ldots, N\}$, some of which are completed after τ_n periods, and some of which are incomplete lasting at least τ_n periods. Let:

$$\rho(n) \equiv \begin{cases} \tau_n \text{ if spell is complete} \\ \{\tau_n, \tau_{n+1}, \ldots\} \text{ if spell is incomplete} \end{cases}$$

- Let $p_\tau(\alpha_n, \beta_n)$ denote the unconditional probability of individual n with discount factor β_n working τ periods in a new job with information factor α_n before switching to another new job in the same occupation:

$$p_\tau(\alpha_n, \beta_n) \equiv h_\tau(\alpha_n, \beta_n) \prod_{s=1}^{\tau-1} [1 - h_s(\alpha_n, \beta_n)]$$

- Then the joint probability of spell duration times observed in the sample is:

$$\prod_{n=1}^{N} \sum_{\tau \in \rho(n)} p_\tau(\alpha_n, \beta_n)$$

Maximum Likelihood Estimation

The likelihood function and structural estimates

- Suppose the information and discount factors depend on X_n, some individual socio-economic factors;

$$\begin{aligned} \alpha_n &\equiv AX_n \\ \beta_n &\equiv BX_n \end{aligned}$$

  where A and B are the structural parameters to be estimated. Then the likelihood is:

$$L_N(A, B) \equiv \prod_{n=1}^{N} \sum_{\tau \in \rho(n)} p_\tau(AX_n, BX_n)$$

- Briefly, the structural estimates show that:
  1. individuals care about the future and the information value from job experimentation;
  2. the occupational dummy variables are significant, suggesting that the choice of different occupations is systematic;
  3. educational groups have different beliefs and learning rates.

- The integration in the job matching example is:
  1. quite cumbersome
  2. suggestive of how quickly integration becomes unmanageable if jobs differ *exante* as well as *expost*.
- CCP estimators can be exploited to ameliorate this problem.
- Recalling Mr. Zurcher's problem:
  - Replace the existing engine ($d_{1t} = 1$), or keep it for at least one more period ($d_{2t} = 1$).
  - Bus mileage x_t follows the update rule $x_{t+1} = d_{1t} + d_{2t} \left(x_t + 1 \right)$.
  - Transitory iid choice-specific shocks, ϵ_{jt}, are T1EV.
  - Zurcher sequentially maximizes expected discounted sum of payoffs:

$$E \left\{ \sum_{t=1}^{\infty} \beta^{t-1} \left[d_{2t} (\theta_1 x_t + \theta_2 s + \epsilon_{2t}) + d_{1t} \epsilon_{1t} \right] \right\}$$

- Now suppose s, the bus make, is unobserved.

Motivating Example

ML Estimation when CCP's are known (infeasible)

- To show how the EM algorithm helps, consider the infeasible case where $s \in \{1, \ldots, S\}$ is unobserved but $p(x, s)$ is known.
- Let π_s denote population probability of being in unobserved state s.
- Supposing β is known the ML estimator for this "easier" problem is:

$$\{\hat{\theta}, \hat{\pi}\} = \arg\max_{\theta,\pi} \sum_{n=1}^{N} \ln \left[\sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} l(d_{nt}|x_{nt}, s, p, \theta) \right]$$

where $p \equiv p(x, s)$ is a string of probabilities assigned/estimated for each (x, s) and $l(d_{nt}|x_{nt}, s_n, p, \theta)$ is derived from our representation of the conditional valuation functions and takes the form:

$$\frac{d_{1nt} + d_{2nt} \exp(\theta_1 x_{nt} + \theta_2 s + \beta \ln [p(0, s)] - \beta \ln [p(x_{nt} + 1, s)]}{1 + \exp(\theta_1 x_{nt} + \theta_2 s + \beta \ln [p(0, s)] - \beta \ln [p(x_{nt} + 1, s)])}$$

- Maximizing over the sum of a log of summed products is computationally burdensome.

Motivating Example

Why EM is attractive (when CCP's are known)

- The EM algorithm is a computationally attractive alternative to directly maximizing the likelihood.
- Denote by $d_n \equiv (d_{n1}, \ldots, d_{nT})$ and $x_n \equiv (x_{n1}, \ldots, x_{nT})$ the full sequence of choices and mileages observed in the data for bus n.
- At the m^{th} iteration:

$$\begin{aligned} q_{ns}^{(m+1)} &= \Pr\left\{ s \,\middle|\, d_n, x_{n,} \theta^{(m)}, \pi_s^{(m)}, p \right\} \\ &= \frac{\pi_s^{(m)} \prod_{t=1}^{T} l(d_{nt} | x_{nt}, s, p, \theta^{(m)})}{\sum_{s'=1}^{S} \pi_{s'}^{(m)} \prod_{t=1}^{T} l(d_{nt} | x_{nt}, s', p, \theta^{(m)})} \end{aligned}$$

$$\pi_s^{(m+1)} = N^{-1} \sum_{n=1}^{N} q_{ns}^{(m+1)}$$

$$\theta^{(m+1)} = \arg\max_{\theta} \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{t=1}^{T} q_{ns}^{(m+1)} \ln[l(d_{nt} | x_{nt}, s, p, \theta)]$$

Motivating Example

Steps in our algorithm when s is unobserved and CCP's are unknown

Our algorithm begins by setting initial values for $\theta^{(1)}$, $\pi^{(1)}$, and $p^{(1)}(\cdot)$:

Step 1 Compute $q_{ns}^{(m+1)}$ as:

$$q_{ns}^{(m+1)} = \frac{\pi_s^{(m)} \prod_{t=1}^{T} l\left[d_{nt} | x_{nt}, s, p^{(m)}, \theta^{(m)}\right]}{\sum_{s'=1}^{S} \pi_s^{(m)} \prod_{t=1}^{T} l\left(d_{nt} | x_{nt}, s', p^{(m)}, \theta^{(m)}\right)}$$

Step 2 Compute $\pi_s^{(m+1)}$ according to:

$$\pi_s^{(m+1)} = \frac{\sum_{n=1}^{N} q_{ns}^{(m+1)}}{N}$$

Step 3 Update $p^{(m+1)}(x, s)$ using one of two rules below.

Step 4 Obtain $\theta^{(m+1)}$ from:

$$\theta^{(m+1)} = \arg\max_{\theta} \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{t=1}^{T} q_{ns}^{(m+1)} \ln\left[l\left(d_{nt} | x_{nt}, s_n, p^{(m+1)}, \theta\right)\right.$$

Motivating Example

Updating the CCP's

- Take a weighted average of decisions to replace engine, conditional on x, where weights are the conditional probabilities of being in unobserved state s.

Step 3A Update CCP's with:

$$p^{(m+1)}(x,s) = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} d_{1nt} q_{ns}^{(m+1)} I(x_{nt} = x)}{\sum_{n=1}^{N}\sum_{t=1}^{T} q_{ns}^{(m+1)} I(x_{nt} = x)}$$

- Or in a stationary infinite horizon model use identity from model that likelihood returns CCP of replacing the engine:

Step 3B Update CCP's with:

$$p^{(m+1)}(x_{nt}, s_n) = l(d_{nt1} = 1 | x_{nt}, s_n, p^{(m)}, \theta^{(m)})$$

Monte Carlo

Finite horizon renewal problem

- Suppose $s \in \{0, 1\}$ equally weighted.
- There are two observed state variables
  1. total accumulated mileage:

$$x_{1t+1} = \begin{cases} \Delta_t \text{ if } d_{1t} = 1 \\ x_{1t} + \Delta_t \text{ if } d_{2t} = 1 \end{cases}$$

  2. permanent route characteristic for the bus, x_2, that systematically affects miles added each period.
- We assume $\Delta_t \in \{0, 0.125, \ldots, 24.875, 25\}$ is drawn from:

$$f(\Delta_t | x_2) = \exp\left[-x_2(\Delta_t - 25)\right] - \exp\left[-x_2(\Delta_t - 24.875)\right]$$

and x_2 is a multiple 0.01 drawn from a discrete equi-probability distribution between 0.25 and 1.25.

- Let θ_{0t} be an aggregate shock (denoting cost fluctuations say).
- The difference in current payoff from retaining versus replacing the engine is:

$$u_{2t}(x_{1t}, s) - u_{1t}(x_{1t}, s) \equiv \theta_{0t} + \theta_1 \min\{x_{1t}, 25\} + \theta_2 s$$

- Denoting the observed state variables by $x_t \equiv (x_{1t}, x_2)$, this translates to:

$$\begin{aligned} v_{2t}(x_t, s) - v_{1t}(x_t, s) &= \theta_{0t} + \theta_1 \min\{x_{1t}, 25\} + \theta_2 s \\ &\quad + \beta \sum_{\Delta_t \in \Lambda} \left\{ \ln \left[\frac{p_{1t}(0, s)}{p_{1t}(x_{1t} + \Delta_t, s)} \right] \right\} f(\Delta_t | x_2) \end{aligned}$$

Monte Carlo

Table 1 of Arcidiacono and Miller (2011, page 1854)

MONTE CARLO FOR THE OPTIMAL STOPPING PROBLEM[a]

		s Observed			*s* Unobserved		Time Effects	
	DGP	FIML	CCP	Ignoring *s* CCP	FIML	CCP	*s* Observed CCP	*s* Unobserved CCP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
θ_0 (intercept)	2	2.0100	1.9911	2.4330	2.0186	2.0280		
		(0.0405)	(0.0399)	(0.0363)	(0.1185)	(0.1374)		
θ_1 (mileage)	−0.15	−0.1488	−0.1441	−0.1339	−0.1504	−0.1484	−0.1440	−0.1514
		(0.0074)	(0.0098)	(0.0102)	(0.0091)	(0.0111)	(0.0121)	(0.0136)
θ_2 (unobs. state)	1	0.9945	0.9726		1.0073	0.9953	0.9683	1.0067
		(0.0611)	(0.0668)		(0.0919)	(0.0985)	(0.0636)	(0.1417)
β (discount factor)	0.9	0.9102	0.9099	0.9115	0.9004	0.8979	0.9172	0.8870
		(0.0411)	(0.0554)	(0.0591)	(0.0473)	(0.0585)	(0.0639)	(0.0752)
Time (minutes)		130.29	0.078	0.033	275.01	6.59	0.079	11.31
		(19.73)	(0.0041)	(0.0020)	(15.23)	(2.52)	(0.0047)	(5.71)

[a]Mean and standard deviations for 50 simulations. For columns 1–6, the observed data consist of 1000 buses for 20 periods. For columns 7 and 8, the intercept (θ_0) is allowed to vary over time and the data consist of 2000 buses for 10 periods. See the text and the Supplemental Material for additional details.

- We parameterize $u_{jt}(z_t)$ and $G(\epsilon_t)$ by θ, $f_{jt}(z_{t+1}|z_t)$ with α, and following our motivating example, we define two estimators.
- Given any CCP vector $\widehat{p}$, both solve a first order condition that maximizing the joint log likelihood also solves:

$$(\widehat{\theta}, \widehat{\pi}, \widehat{\alpha}) = \underset{(\theta,\pi,\alpha)}{\arg\max} \sum_{n=1}^{\mathcal{N}} \log \left[\sum_{s=1}^{\mathcal{S}} \pi_s L\left(d_n, x_n \left| x_{n1}, s ; \theta, \alpha, \widehat{p}\right.\right) \right]$$

where $L\left(d_n, x_n \left| x_{n1}, s ; \theta, \alpha, \widehat{p}\right.\right)$ is the likelihood of the (panel length) sequence (d_n, x_n):

$$\begin{aligned} & L\left(d_n, x_n \left| x_{n1}, s ; \theta, \alpha, \pi, p\right.\right) \\ = \quad & \prod_{t=1}^{\mathcal{T}} \sum_{j=1}^{J} d_{jnt} l_{jt}(x_{nt}, s, \theta, \pi, p) f_{jt}\left(x_{n,t+1} | x_{nt}, s, \alpha\right) \end{aligned}$$

The Estimators

Using the Likelihood to obtain the CCP's

- The difference between the estimators arises from how $\widehat{p}$ is defined.
- The first estimator is based on the fact that $l_{jt}(x_{nt}, s_n, \theta, \alpha, \pi, p)$ is the likelihood of observing individual n make choice j at time t given s_n.
- Accordingly define $\widehat{p}\,(x, s)$ to solve:

$$\widehat{p}_{jt}\,(x, s) = l_{jt}(x, s; \widehat{\theta}, \widehat{\alpha}, \widehat{\pi}, \widehat{p})$$

- The large sample properties are standard.

Appendix: The Estimators

An empirical approach to the CCP's

- Let $\widehat{L}_n(s_n = s)$ denote the joint likelihood of the data for n and being in unobserved state s evaluated at $\left(\widehat{\theta}, \widehat{\alpha}, \widehat{\pi}, \widehat{p}\right)$.

$$\widehat{L}_n(s_n = s) \equiv \widehat{\pi}_s L\left(d_n, x_n \left| x_{n1}, s ; \widehat{\theta}, \widehat{p}\right.\right)$$

- Also denote by $\widehat{L}_n$ the likelihood of observing (d_n, x_n) given parameter values $\left(\widehat{\pi}, \widehat{\theta}, \widehat{p}\right)$:

$$\begin{aligned} \widehat{L}_n &\equiv \sum_{s=1}^{\mathcal{S}} \widehat{\pi}_s L\left(d_n, x_n \left| x_{n1}, s ; \widehat{\theta}, \widehat{p}\right.\right) \\ &= \sum_{s=1}^{\mathcal{S}} \widehat{L}_n(s_n = s) \end{aligned}$$

- As an estimated sample approximation, $N^{-1} \sum_{n=1}^{N} \left[\widehat{L}_n(s_n = s) / \widehat{L}_n\right]$ is the fraction of the population in s.

- Similarly:
  1. $N^{-1}\sum_{n=1}^{N}\left[I(x_{nt}=x)\widehat{L}_n(s_n=s)/\widehat{L}_n\right]$ is the estimated fraction of the population in s with x at t.
  2. $N^{-1}\sum_{n=1}^{N}\left[d_{jnt}I(x_{nt}=x)\widehat{L}_n(s_n=s)/\widehat{L}_n\right]$ is the estimated fraction choosing j at t as well.
- We define:

$$\widehat{p}_{jt}(x,s)=$$

$$\left[\sum_{n=1}^{\mathcal{N}} d_{jnt}I(x_{nt}=x)\frac{\widehat{L}_n(s_n=s)}{\widehat{L}_n}\right]\Bigg/\left[\sum_{n=1}^{\mathcal{N}} I(x_{nt}=x)\frac{\widehat{L}_n(s_n=s)}{\widehat{L}_n}\right]$$

- Compared to the first one this estimator has similar properties but imposes less structure.