Identifiability and Adaptive Control of Markov Chains

Ben Pence
Scott Moura

EECS 558 Project

December 12, 2009

Outline

- Stochastic system with unknown parameters
- Want to identify the parameters
- Adapt the control based on the parameter estimate

Identifiability
Estimation
Adaptive Control

Results With Identifiability
Mandl (1973)

Results W/O Identifiability
Borkar, Varaiya (1979,82)

Conclusions

Identifiability

Model: Controlled Markov Chain

$P_{ij}(u,\alpha)$ Transition Probability Matrix from state i to j; i,j∈S

$\alpha \in A$ α=Unknown Parameter, A=Parameter Space

$u \in U$ u=Control Action, U=Control Space

"Identifiability" condition:

If $\alpha \neq \beta$ in A, then

$$P_{ij}(u,\alpha) \neq P_{ij}(u,\beta) \qquad \text{For } \underline{\text{any}}\ u \in U$$

Identifiability

$$P_{ij} = \begin{bmatrix} 0.5-\alpha & 0.5+\alpha \\ 1 & 0 \end{bmatrix} \quad \begin{array}{c} u \in \{1,2\} \\ \alpha \in \{0.1, 0.2, 0.3\} \end{array}$$

☑ Identifiable
☐ Not Identifiable

$$P_{ij} = \begin{bmatrix} 0.5-2\alpha+\alpha u & 0.5+2\alpha-\alpha u \\ 1 & 0 \end{bmatrix} \quad \begin{array}{c} u \in \{1,2\} \\ \alpha \in \{0.1, 0.2, 0.3\} \end{array}$$

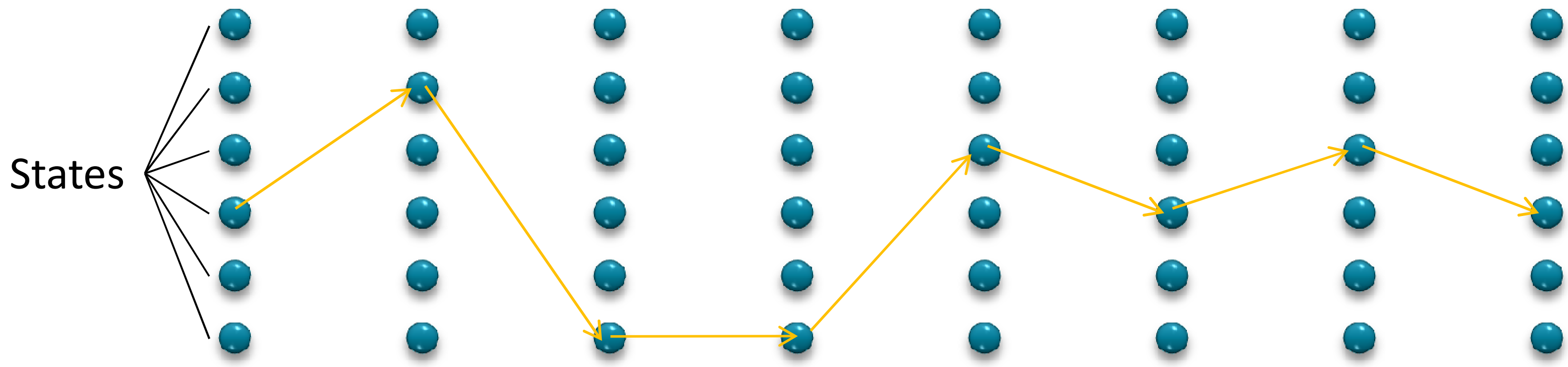
☐ Identifiable
☑ Not Identifiable

$$P_{ij} = \begin{bmatrix} 0.5-2\alpha+\alpha u & 0.5+2\alpha-\alpha u \\ 1 & 0 \end{bmatrix} \quad \begin{array}{c} u \in \{1,4\} \\ \alpha \in \{0.1, 0.2, 0.3\} \end{array}$$

☑ Identifiable
☐ Not Identifiable

Parameter Estimation

Maximum Likelihood Estimator



What value of α makes this trajectory most likely?

$$\hat{\alpha}_t = \arg\max_{\alpha} \prod_{s=0}^{t-1} P(x_s, x_{s+1}; u_s, \alpha)$$

Probability of transitioning from x_s to x_{s+1} given u_s and α

Adaptive Control

For each $\alpha \in A$ a SMP has been prespecified

$$U_t = g^{\alpha}(X_t)$$

At each time step, apply the SMP corresponding to the parameter estimate $\hat{\alpha}_t$

$$U_t = g^{\hat{\alpha}_t}(X_t)$$

Outline

- Identifiability
- Estimation
- Adaptive Control
- **Results With Identifiability** — Mandl (1973)
- Results W/O Identifiability — Borkar, Varaiya (1979,82)
- Conclusions

Main Result of Mandl, 1973

Theorem 1

For the model under consideration, with the assumptions previously outlined (indentifiability), there exists random time T such that for $t \geq T$

$$\hat{\alpha}_t = \alpha^*$$ Parameter estimate reaches limit point

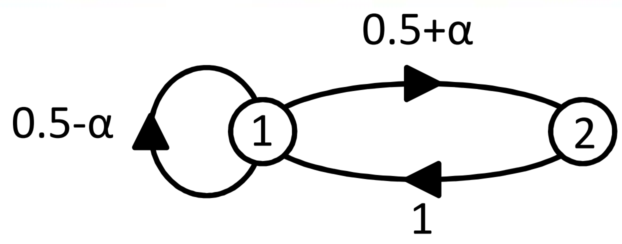
$$\alpha^* = \alpha^0$$ Limit point is true parameter value!

$$P\left(i, j; g^{\alpha^*}(i), \alpha^*\right) = P\left(i, j; g^{\alpha^0}(i), \alpha^0\right) \quad i, j \in S$$

Moreover, closed-loop transition probabilities are the same as if we had known the true parameter

Example 1: Identifiable System

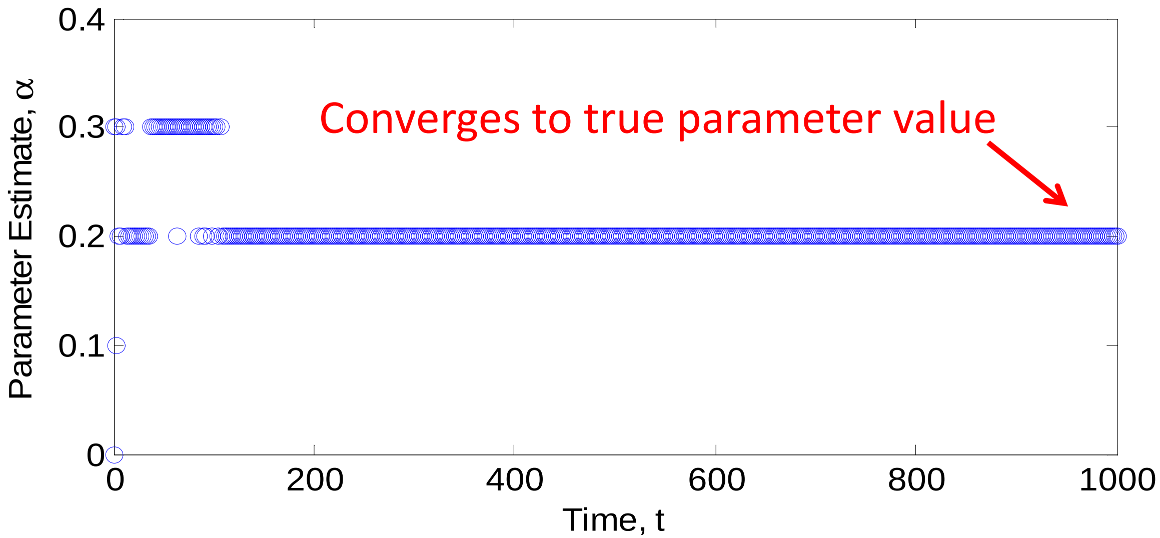
Markov Chain



$$S = \{1, 2\}$$
$$\boldsymbol{U} = \{1, 2\}$$
$$A = \{0.1, 0.2, 0.3\}$$

Transition Prob's $P_{ij} = \begin{bmatrix} 0.5-\alpha & 0.5+\alpha \\ 1 & 0 \end{bmatrix}$

True parameter $\alpha^0 = 0.2$



Control Law

$$u = g^{0.1}(i) = 2$$
$$u = g^{0.2}(i) = 1$$
$$u = g^{0.3}(i) = 2$$

Outline

- Identifiability
- Estimation
- Adaptive Control
- Results With Identifiability
  Mandl (1973)
- Results W/O Identifiability
  Borkar, Varaiya (1979,82)
- Conclusions

Motivation of Borkar and Varaiya, 1979

Hypothesis

The identifiability assumption is too restrictive

$$P_{ij} = \begin{bmatrix} 0.5 - 2\alpha + \alpha u & 0.5 + 2\alpha - \alpha u \\ 1 & 0 \end{bmatrix} \quad \begin{array}{r} u \in \{1,2\} \\ \alpha \in \{0.1, 0.2, 0.3\} \end{array}$$

- [ ] Identifiable
- [x] Not Identifiable

Question

What happens if we relax this assumption?

Main Result of Borkar and Varaiya, 1979

Theorem 2

For the model under consideration, with the assumptions previously outlined except indentifiability, there exists random time T such that for $t \geq T$

$$\hat{\alpha}_t = \alpha^*$$

Parameter estimate reaches limit point

But this limit point is not necessarily equal to the true parameter!

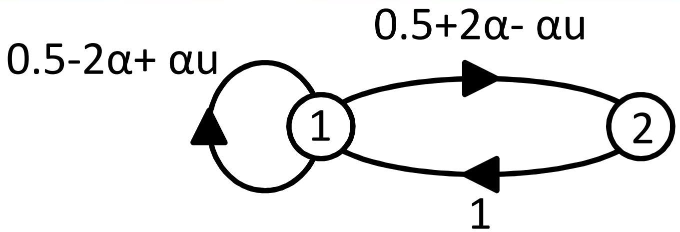
$$P\left(i, j; g^{\alpha^*}(i), \alpha^*\right) = P\left(i, j; g^{\alpha^*}(i), \alpha^0\right) \quad i, j \in S$$

However, closed-loop transition probabilities are the same

1. for true system and some imaginary system where the parameter really was α^*
2. control law corresponding to α^* is applied

Example 2: Not Identifiable System

Markov Chain



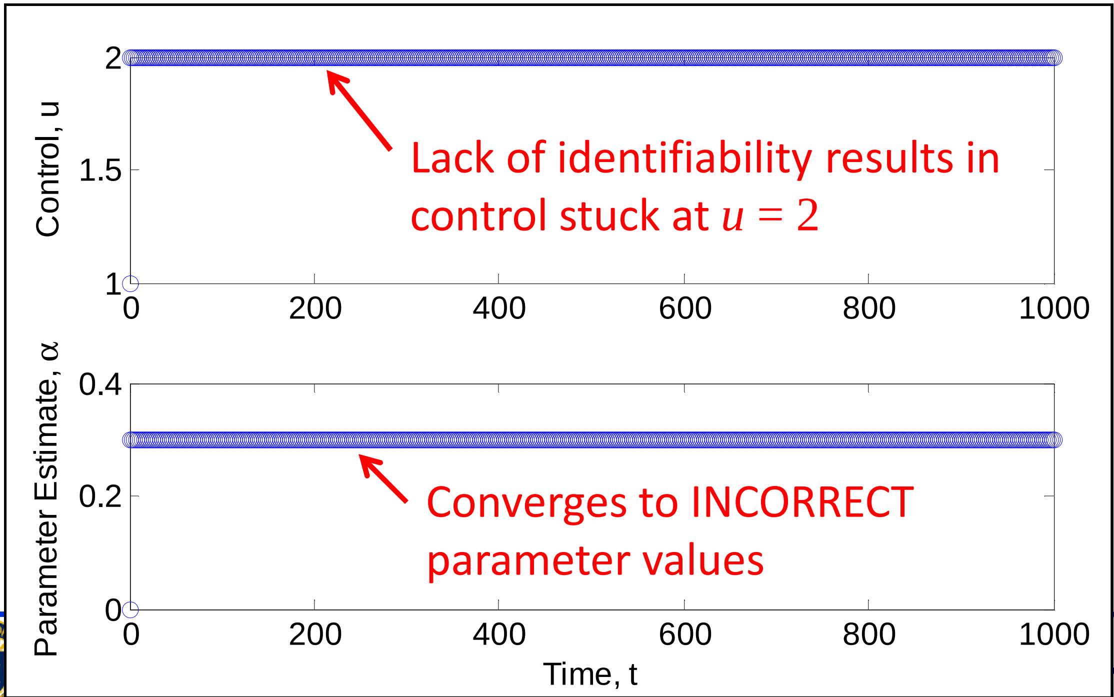
$$S = \{1,2\}$$

$$\boldsymbol{U} = \{1,2\}$$

$$A = \{0.1, 0.2, 0.3\}$$

Transition Prob's $P_{ij} = \begin{bmatrix} 0.5 - 2\alpha + \alpha u & 0.5 + 2\alpha - \alpha u \\ 1 & 0 \end{bmatrix}$

True parameter $\alpha^0 = 0.2$



Control Law

$$u = g^{0.1}(i) = 2$$

$$u = g^{0.2}(i) = 1$$

$$u = g^{0.3}(i) = 2$$

Interpretation of Theorem 2

Recall the objectives:

- Estimate the unknown parameter α ✗
- Satisfactorily control the Markov chain ?

Source of Difficulty:

- Only uses control $U_k = g^{\alpha^*}(X_k)$
- Can only identify subset $\{P_{ij}(u) \mid u = g^{\alpha^*}(i)\}$ of all transition probs $\{P_{ij}(u) \mid u \in \boldsymbol{U}\}$
- Adaptive controller should “probe” outside subset

Control Randomization
Borkar and Varaiya (1982)

Main Idea

Probe outside the subset of transition probabilities by perturbing control signals randomly

Key Assumption: "Partial" Identifiability

If $\alpha \neq \beta$ in A, then the MC is partially identifiable if in every open set $O \in U$, there exist $u \in O$ s.t.

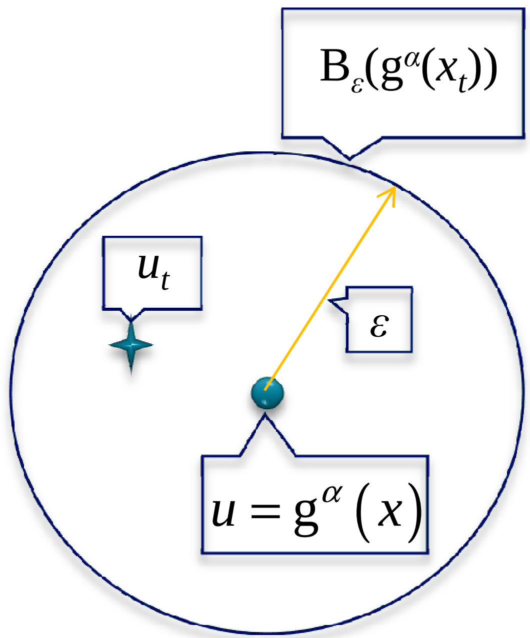
$$P_{ij}(u,\alpha) \neq P_{ij}(u,\beta)$$

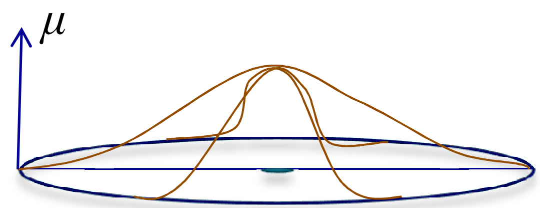
Control Randomization

Borkar and Varaiya (1982)

<u>Construction of Randomized Controls</u>

- Current parameter estimate = α
- Control $u = \mathrm{g}^{\alpha}(x)$
- Construct an open ball $\mathrm{B}_{\varepsilon}(u_t)$ for each u_t
  - Radius $\varepsilon > 0$ small centered at u
- Construct a probability measure μ on $\boldsymbol{U}$ which assigns probabilities to open balls
- Perform independent experiment to generate random control u_t from $\mathrm{B}_{\varepsilon}(\mathrm{g}^{\alpha}(x_t))$ using μ





Main Result of Borkar and Varaiya, 1982

Theorem 3

For the model under consideration, with the assumptions previously outlined including partial indentifiability, then under any ε-randomization of g

$$\lim_{t \to \infty} \hat{\alpha}_t = \alpha^*$$ Parameter estimate reaches limit point

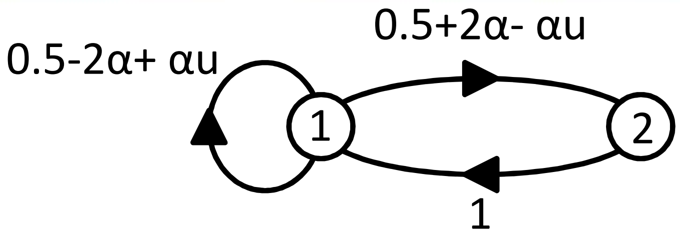
$$\alpha^* = \alpha^0$$ Limit point is true parameter value!

$$P\left(i, j; g^{\alpha^*}(i), \alpha^*\right) = P\left(i, j; g^{\alpha^0}(i), \alpha^0\right) \quad i, j \in S$$

Moreover, closed-loop transition probabilities are the same as if we had known the true parameter

Example 3: Control Randomization

Markov Chain



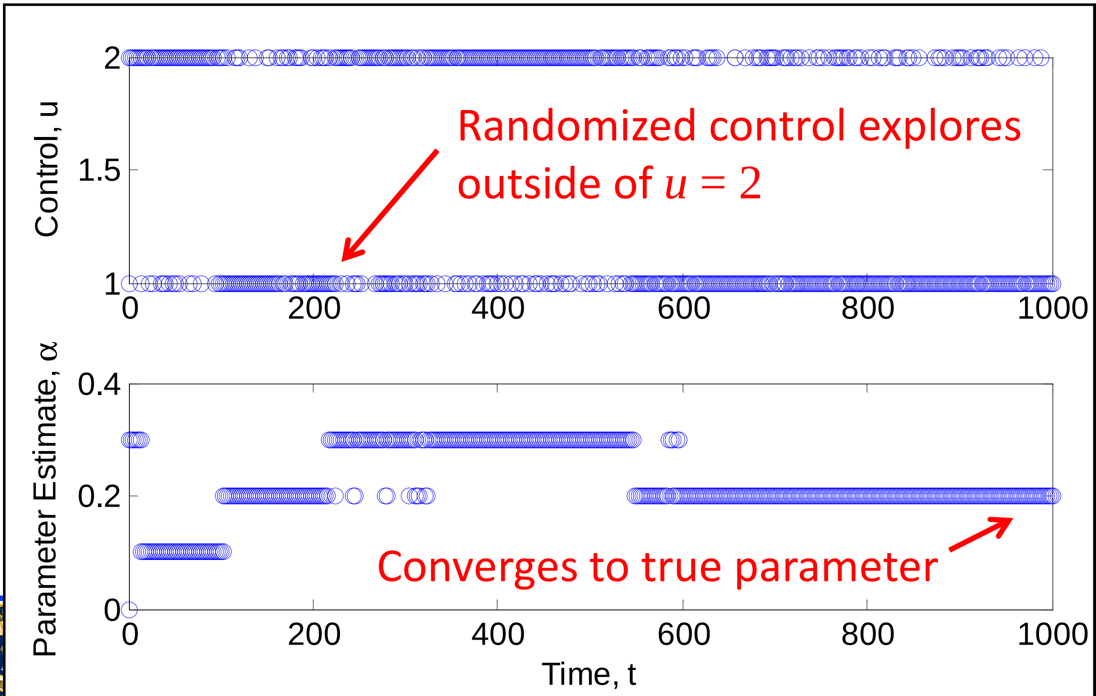
$S = \{1,2\}$

$\boldsymbol{U} = \{1,2\}$

$A = \{0.1, 0.2, 0.3\}$

Transition Prob's $P_{ij} = \begin{bmatrix} 0.5-2\alpha+\alpha u & 0.5+2\alpha-\alpha u \\ 1 & 0 \end{bmatrix}$

True parameter $\alpha^0 = 0.2$



Randomized Control Law

$\mathrm{P}(u=2 \mid \alpha \in \{0.1, 0.3\}) = 0.75$ Control

$\mathrm{P}(u=1 \mid \alpha \in \{0.1, 0.3\}) = 0.25$ Probe

$\mathrm{P}(u=1 \mid \alpha \in \{0.2\}) = 0.75$ Control

$\mathrm{P}(u=2 \mid \alpha \in \{0.2\}) = 0.25$ Probe

Interpretation of Theorem 3

Recall the objectives:

- Estimate the unknown parameter α ✓
- Satisfactorily control the Markov chain ✓

Summary:

- Relaxed restrictive identifiability condition through <u>local probing</u>

Outline

Identifiability Estimation Adaptive Control	Results With Identifiability Mandl (1973)	Results W/O Identifiability Borkar, Varaiya (1979,82)	**Conclusions**

Concluding Remarks

- Simultaneously identify and control a Markov chain with unknown parameters
- MLE converges to true parameter in closed-loop, under identifiability condition
- Without identifiability condition, MLE may not converge to true parameter
- Through “probing” control only need “partial” identifiability to converge to the true parameter

Potential Future Directions

- Theoretical Questions
  - What if $\alpha^0 \notin A$?
  - Can control randomization be stopped eventually?
  - Alternative estimation methods?
- Application Areas
  - Adaptive control of hybrid vehicles through drive cycle identification
  - Algorithms for calculating the maximum likelihood estimate

Thank you for your attention!

Questions?
Comments?
Suggestions?

Appendix Slides

Parameter Estimation

Maximum Likelihood Estimator

$$L_t(\alpha) = \prod_{s=0}^{t-1} P\left(X_t, X_{t+1}; U_t, \alpha\right)$$

$$\hat{\alpha}_t = \arg\max\left\{L_t(\alpha)\right\}$$

Likelihood Ratio

$$\Lambda_t(\alpha) = \frac{L_t(\alpha)}{L_t\left(\alpha^0\right)}$$

Control Randomization
Borkar and Varaiya (1982)

<u>Construction of Randomized Controls</u>

- Suppose current parameter estimate is α and corresponding control is $u = \mathrm{g}^{\alpha}(x)$
- Construct an open ball with radius $\varepsilon > 0$ small centered at u, denoted by $\mathrm{B}_{\varepsilon}(u_t)$, for each $u \in \boldsymbol{U}$
- Construct a probability measure μ on $\boldsymbol{U}$ which assigns probabilities to open balls
- Perform independent experiment to generate random control u_t from $\mathrm{B}_{\varepsilon}(\mathrm{g}^{\alpha}(x_t))$ using μ

Sequence $\{u_t, t = 0,1,\ldots\}$ is called the ε-randomization of g

Presentation Outline

- Model and Problem Formulation
- Main Results under Identifiability (1st paper)
- Results without Identifiability (2nd paper)
- Control Randomization (3rd paper)
- Concluding Remarks

Motivation of Borkar and Varaiya, 1979

Motivation

The identifiability assumption is too restrictive

For every $\alpha \neq \beta$ in A, there exist at least one $i \in S$

$$\begin{bmatrix} P(i,1;u,\alpha) & P(i,2;u,\alpha) & \cdots & P(i,I;u,\alpha) \end{bmatrix}$$
$$\neq \begin{bmatrix} P(i,1;u,\beta) & P(i,2;u,\beta) & \cdots & P(i,I;u,\beta) \end{bmatrix}$$
$$\forall u \in \boldsymbol{U}$$

Question

What happens if we relax this assumption?

Presentation Outline

- Model and Problem Formulation
- Main Results under Identifiability (1st paper)
- Results without Identifiability (2nd paper)
- Control Randomization (3rd paper)
- Concluding Remarks

Model

- Controlled Markov chain (MC)
- Finite state space $S = \{1, 2, \ldots, I\}$
- Finite action space $\boldsymbol{U}$
- Finite parameter space A
- Transition probabilities

  $\{P(i, j; u, \alpha) \quad i, j \in S \quad u \in U \quad \alpha \in A\}$
- Stationary Markov policy (SMP) $U_t = \mathrm{g}(X_t)$
- Perfect Recall

Objective

- Estimate the unknown parameter α
- Adaptively Control the Markov chain
- Analyze the asymptotic convergence properties
  - Convergence or not?
  - What assumptions are necessary?
  - Quantify performance of adaptive MC

Assumptions

1. Controlled MC is irreducible, $\forall u \in \boldsymbol{U}, \alpha \in A$
2. True parameter $\alpha^0 \in A$
3. The SMP corresponding to α, g^α, precomputed
4. "Identifiability" condition:

   If $\alpha \neq \beta$ in A, then

$$P_{ij}(u,\alpha) \neq P_{ij}(u,\beta) \qquad \text{For } \underline{\text{any}}\ u \in U$$

Outline

- Model and Problem Formulation
- Main Results under Identifiability (1st paper)
- Results without Identifiability (2nd paper)
- Control Randomization (3rd paper)
- Concluding Remarks