Statistical Methdology For Innocence Project Data

ASC

November 2012

Kobi Abayomi[1] Jessica Gabel[2] Otis Jennings[3]
1:SUNY at Binghamton
2:Georgia State University, College of Law
3: Columbia University, School of Business

Part 1

Part 1

Brief Background

Exoneration

Gross et al ([7]) define exoneration more broadly as:

> *An official act declaring a defendant not guilty of a crime for which he or she had been previously been convicted.*

Brief Background

Exoneration

Gross et al ([7]) define exoneration more broadly as:

> *An official act declaring a defendant not guilty of a crime for which he or she had been previously been convicted.*

Using this definition Gross et al count 340 total exonerated men and women between 1989 and 2003; 80% of whom had been imprisoned for five or more years; 73% of whom were exonerated on the basis of DNA evidence.

Brief Background

Exoneration

Gross et al ([7]) define exoneration more broadly as:

> *An official act declaring a defendant not guilty of a crime for which he or she had been previously been convicted.*

Using this definition Gross et al count 340 total exonerated men and women between 1989 and 2003; 80% of whom had been imprisoned for five or more years; 73% of whom were exonerated on the basis of DNA evidence.

The Innocence Project has focused on cases where exoneration = DNA exculpation.

These cases are just the observed positives

Brief Background

DNA Evidence

Kaye points out the recent recasting of DNA evidence, [9]:

> *It is important to note that DNA evidence has assumed an exculpatory role relatively recently...DNA testing for identification in criminal forensics was initially critiqued as too error prone to meet a legal evidentiary standard*

Brief Background

DNA Evidence

Kaye points out the recent recasting of DNA evidence, [9]:

> *It is important to note that DNA evidence has assumed an exculpatory role relatively recently...DNA testing for identification in criminal forensics was initially critiqued as too error prone to meet a legal evidentiary standard*

From the early to the late 1990s, the debates about DNA testing standards yielded to near-universal acceptance — partially due to technological advancement — of DNA testing as *the* definitive criminal identification tool, [10], [12] or [2]

Brief Background

DNA Evidence

Kaye points out the recent recasting of DNA evidence, [9]:

> *It is important to note that DNA evidence has assumed an exculpatory role relatively recently...DNA testing for identification in criminal forensics was initially critiqued as too error prone to meet a legal evidentiary standard*

From the early to the late 1990s, the debates about DNA testing standards yielded to near-universal acceptance — partially due to technological advancement — of DNA testing as *the* definitive criminal identification tool, [10], [12] or [2]
While DNA is vital to redress a wrongful conviction, its absence weakens cases — the vast majority of exoneration requests — where there simply is no DNA evidence available [13]

Brief Background

Wrongful Conviction

Brief Background

Wrongful Conviction

- Most cases lack DNA evidence.

Brief Background

Wrongful Conviction

- Most cases lack DNA evidence.
- Same fundamental errors?

Brief Background

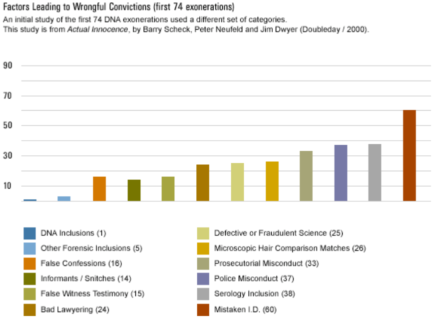
Wrongful Conviction

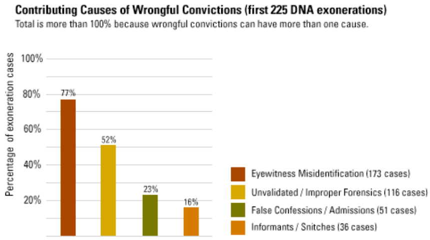
- Most cases lack DNA evidence.
- Same fundamental errors?
- Eyewitness Misidentifications, False Confessions, Jailhouse Snitches, and Flawed Forensics, [4]

Brief Background

Wrongful Conviction

- Most cases lack DNA evidence.
- Same fundamental errors?
- Eyewitness Misidentifications, False Confessions, Jailhouse Snitches, and Flawed Forensics, [4]





Brief Background

Factors in Wrongful Conviction
Current Work

- Examined through the framework of DNA testing: exclusion and non-identification, [6]

Brief Background

Factors in Wrongful Conviction
Current Work

- Examined through the framework of DNA testing: exclusion and non-identification, [6]
- Just a small fraction of the entreaties the IPs receive will ever have DNA evidence available...

Brief Background

Factors in Wrongful Conviction
Current Work

- Examined through the framework of DNA testing: exclusion and non-identification, [6]
- Just a small fraction of the entreaties the IPs receive will ever have DNA evidence available...
- ...Just a fraction of the potentially large numbers of wrongfully convicted [5]

Brief Background

- Examined through the framework of DNA testing: exclusion and non-identification, [6]
- Just a small fraction of the entreaties the IPs receive will ever have DNA evidence available...
- ...Just a fraction of the potentially large numbers of wrongfully convicted [5]

Part 1

Brief Background

DNA Evidence $\neq$ Exoneration $\neq$ Innocence

We have to assert, prior to building a statistical model

- Non-positive DNA evidence cannot perfectly identify Wrongful Convictions

Brief Background

DNA Evidence $\neq$ Exoneration $\neq$ Innocence

We have to assert, prior to building a statistical model

- Non-positive DNA evidence cannot perfectly identify Wrongful Convictions
- A Wrongful Conviction isn't functionally an Exoneration (using Gross' definition.)

Brief Background

DNA Evidence $\neq$ Exoneration $\neq$ Innocence

We have to assert, prior to building a statistical model

- Non-positive DNA evidence cannot perfectly identify Wrongful Convictions
- A Wrongful Conviction isn't functionally an Exoneration (using Gross' definition.)
- The universe of Wrongful Convictions is (much) larger than the present or future universe of Non-positive DNA evidence

Brief Background

DNA Evidence $\neq$ Exoneration $\neq$ Innocence

We have to assert, prior to building a statistical model

- Non-positive DNA evidence cannot perfectly identify Wrongful Convictions
- A Wrongful Conviction isn't functionally an Exoneration (using Gross' definition.)
- The universe of Wrongful Convictions is (much) larger than the present or future universe of Non-positive DNA evidence

Brief Background

The Innocence Projects

- Endogenous difference: Different projects have different 'rules' and procedures

Brief Background

The Innocence Projects

- Endogenous difference: Different projects have different 'rules' and procedures
- Exogenous differences: Mechanisms that wrongfully convict may differ state by state, locale by locale, prosecutor by prosecutor...

Brief Background

The Innocence Projects

- Endogenous difference: Different projects have different ‘rules’ and procedures
- Exogenous differences: Mechanisms that wrongfully convict may differ state by state, locale by locale, prosecutor by prosecutor...
- Recordkeeping/Coding Data: Inconsistent state by state. National level ‘clearing-house’ (NY IP) vs. State-level records.

Brief Background

The Innocence Projects

- Endogenous difference: Different projects have different ‘rules’ and procedures
- Exogenous differences: Mechanisms that wrongfully convict may differ state by state, locale by locale, prosecutor by prosecutor...
- Recordkeeping/Coding Data: Inconsistent state by state. National level ‘clearing-house’ (NY IP) vs. State-level records.

Part 1

Brief Background

A Priori

Designing a study

- Define observably positive dependent variables.

Brief Background

A Priori

Designing a study

- Define observably positive dependent variables.
- Sampling methodology.

Brief Background

A Priori

Designing a study

- Define observably positive dependent variables.
- Sampling methodology.
- Rich statistical modeling framework

Brief Background

A Priori

Designing a study

- Define observably positive dependent variables.
- Sampling methodology.
- Rich statistical modeling framework

Part 2

The Data

Georgia Innocence Project

- Founded August 2002, assisted with 8 exonerations, 5 GIP alone.

The Data

Georgia Innocence Project

- Founded August 2002, assisted with 8 exonerations, 5 GIP alone.
- Have received over 5000 case requests from Georgia and Alabama

The Data

Georgia Innocence Project

- Founded August 2002, assisted with 8 exonerations, 5 GIP alone.
- Have received over 5000 case requests from Georgia and Alabama
- The GIP focuses on **DNA cases**

The Data

Georgia Innocence Project

- Founded August 2002, assisted with 8 exonerations, 5 GIP alone.
- Have received over 5000 case requests from Georgia and Alabama
- The GIP focuses on **DNA cases**

The Data

Georgia Innocence Project

- The Georgia Innocence Project (GIP) has a unique identifier, the GIP which is separate but *should* also be unique for the GADOC/Inmate number

The Data

Georgia Innocence Project

- The Georgia Innocence Project (GIP) has a unique identifier, the GIP which is separate but *should* also be unique for the GADOC/Inmate number
- Approximately 5000 cases, we sample approximately 10 percent

The Data

Georgia Innocence Project

- The Georgia Innocence Project (GIP) has a unique identifier, the GIP which is separate but *should* also be unique for the GADOC/Inmate number
- Approximately 5000 cases, we sample approximately 10 percent
- Exclude cases from Alabama, Florida.

The Data

Georgia Innocence Project

- The Georgia Innocence Project (GIP) has a unique identifier, the GIP which is separate but *should* also be unique for the GADOC/Inmate number
- Approximately 5000 cases, we sample approximately 10 percent
- Exclude cases from Alabama, Florida.

We want to exploit the temporality of the GIP record keeping

Part 2

Covariates 'in' the GIP Data

Illustration for GIP data is general enough to be applicable to most organizations

Let $\mathcal{X}$ be a collection of states for a Markov Process and let $\mathcal{Z}$ be associated covariates. *These data can easily be collected by any IP which keeps even minimal records for their internal files.*

1. $Z_1^j = 1$ False Confession?

Covariates 'in' the GIP Data

Illustration for GIP data is general enough to be applicable to most organizations

Let $\mathcal{X}$ be a collection of states for a Markov Process and let $\mathcal{Z}$ be associated covariates. *These data can easily be collected by any IP which keeps even minimal records for their internal files.*

1. $Z_1^j = 1$ False Confession?
2. $Z_2^j = 1$ Snitch?

Covariates 'in' the GIP Data

Illustration for GIP data is general enough to be applicable to most organizations

Let $\mathcal{X}$ be a collection of states for a Markov Process and let $\mathcal{Z}$ be associated covariates. *These data can easily be collected by any IP which keeps even minimal records for their internal files.*

1. $Z_1^j = 1$ False Confession?
2. $Z_2^j = 1$ Snitch?
3. $Z_3^j = 1$ Race Black?

Covariates 'in' the GIP Data

Illustration for GIP data is general enough to be applicable to most organizations

Let $\mathcal{X}$ be a collection of states for a Markov Process and let $\mathcal{Z}$ be associated covariates. *These data can easily be collected by any IP which keeps even minimal records for their internal files.*

1. $Z_1^j = 1$ False Confession?
2. $Z_2^j = 1$ Snitch?
3. $Z_3^j = 1$ Race Black?
4. $Z_4^j = 1$ Victim White?

Covariates 'in' the GIP Data

Illustration for GIP data is general enough to be applicable to most organizations

Let $\mathcal{X}$ be a collection of states for a Markov Process and let $\mathcal{Z}$ be associated covariates. *These data can easily be collected by any IP which keeps even minimal records for their internal files.*

1. $Z_1^j = 1$ False Confession?
2. $Z_2^j = 1$ Snitch?
3. $Z_3^j = 1$ Race Black?
4. $Z_4^j = 1$ Victim White?
5. $Z_5^j =$ Duration in state

Covariates 'in' the GIP Data

Illustration for GIP data is general enough to be applicable to most organizations

Let $\mathcal{X}$ be a collection of states for a Markov Process and let $\mathcal{Z}$ be associated covariates. *These data can easily be collected by any IP which keeps even minimal records for their internal files.*

1. $Z_1^j = 1$ False Confession?
2. $Z_2^j = 1$ Snitch?
3. $Z_3^j = 1$ Race Black?
4. $Z_4^j = 1$ Victim White?
5. $Z_5^j =$ Duration in state
6. $Z_6^j =$ DNA available?

Covariates 'in' the GIP Data

Illustration for GIP data is general enough to be applicable to most organizations

Let $\mathcal{X}$ be a collection of states for a Markov Process and let $\mathcal{Z}$ be associated covariates. *These data can easily be collected by any IP which keeps even minimal records for their internal files.*

1. $Z_1^j = 1$ False Confession?
2. $Z_2^j = 1$ Snitch?
3. $Z_3^j = 1$ Race Black?
4. $Z_4^j = 1$ Victim White?
5. $Z_5^j =$ Duration in state
6. $Z_6^j =$ DNA available?

$Z_1^j, \ldots, Z_4^j$ are indicators for the presence of a characteristic *at state j* while Z_5^j is continuous. Notice that the covariates Z_j are *state dependent*; the value of each Z_j is recovered from the available case record for each state X_j.

This allows model to account for the amount of information available to the IP at each stage in case processing.

Part 2

Sampling the GIP Data

Data sheet for GIP

States:

List date of action from file:

X_W – Inmate Letter Received:___________

X_Q – GIP questionnaire sent__________

X_C – Case Closed:___________
(multiple dates ok)

X_I – Case Inculpated:____________

X_E – Case Exonerated:____________

Covariates:

List date of incidence from file (blank indicates no; multiple dates ok, indicates time varying)

Z_0 – Forensic Evidence?____________

Z_1 – False Confession?:________

Z_2 – Snitch?:________

Z_3 – Race Black?:___________

Z_4 – Victim White?:__________

Z_5 – Duration in Previous State:
(to be calculated from state data, ok to leave blank here)

Z_5_W:_________

Z_5_C:__________

Z_5_I:___________

Z_5_E:__________

Z_6 – DNA available?:_________

Z_7 – Eyewitness ID?:_________

Sampling the GIP Data

Georgia Innocence Project Data

1. Open active cases - in a filing cabinet near the entrance with open letters at the bottom of the filing cabinet

Sampling the GIP Data

Georgia Innocence Project Data

1. Open active cases - in a filing cabinet near the entrance with open letters at the bottom of the filing cabinet
2. Recently closed cases - near the rear right of the office. Cases remain in this area for three months

Sampling the GIP Data

Georgia Innocence Project Data

1. Open active cases - in a filing cabinet near the entrance with open letters at the bottom of the filing cabinet
2. Recently closed cases - near the rear right of the office. Cases remain in this area for three months
3. Closed, closed cases - at the rear of the office. Cases that have been inactive for more than three months

Sampling the GIP Data

Georgia Innocence Project Data

1. Open active cases - in a filing cabinet near the entrance with open letters at the bottom of the filing cabinet
2. Recently closed cases - near the rear right of the office. Cases remain in this area for three months
3. Closed, closed cases - at the rear of the office. Cases that have been inactive for more than three months

$$\hat{p}_1 = .03285714$$

$$\hat{p}_2 = .03163265$$

$$\hat{p}_3 = .9355102$$

Sampling

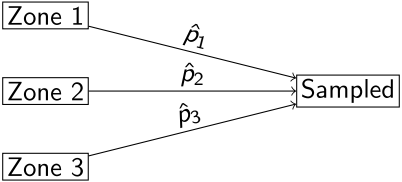


Figure : Sampling design for GIP data

Sampling the GIP Data

Two moving parts to consider in devising a sampling frame for the GIP data

1. The Zone (case classification): overrepresent Zone 1, underrepresent Zone 3.

Sampling the GIP Data

Two moving parts to consider in devising a sampling frame for the GIP data

1. The Zone (case classification): overrepresent Zone 1, underrepresent Zone 3.
2. The GIP number: The number is a (loose) proxy for time in the GIP system. Cases are worked for different lengths of time. **Are the birth/death, state transitions stationary given the covariates?**. This is to say that we may have to assume that the policies (and thus the effect of covariates on state transitions and time in system) of the GIP have remained relatively stable no matter when the case has arrived.

Sampling the GIP Data

Two moving parts to consider in devising a sampling frame for the GIP data

1. The Zone (case classification): overrepresent Zone 1, underrepresent Zone 3.
2. The GIP number: The number is a (loose) proxy for time in the GIP system. Cases are worked for different lengths of time. **Are the birth/death, state transitions stationary given the covariates?**. This is to say that we may have to assume that the policies (and thus the effect of covariates on state transitions and time in system) of the GIP have remained relatively stable no matter when the case has arrived.

Eventual model needs to be rich enough to account for possible non-stationarity in the sampled observations

Sampling the GIP Data

We sample a GIP number and then a Bernoulli Choice

1. Zone 3: Sample 5 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.

Sampling the GIP Data

We sample a GIP number and then a Bernoulli Choice

1. Zone 3: Sample 5 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
2. Zone 2: Sample 3 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.

Sampling the GIP Data

We sample a GIP number and then a Bernoulli Choice

1. Zone 3: Sample 5 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
2. Zone 2: Sample 3 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
3. Zone 1: Sample 2 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.

Sampling the GIP Data

We sample a GIP number and then a Bernoulli Choice

1. Zone 3: Sample 5 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
2. Zone 2: Sample 3 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
3. Zone 1: Sample 2 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
4. Repeat until youve exhausted the list of random numbers

Sampling the GIP Data

We sample a GIP number and then a Bernoulli Choice

1. Zone 3: Sample 5 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
2. Zone 2: Sample 3 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
3. Zone 1: Sample 2 cases. Draw numbers in order from list in attached file, if the number doesnt appear in the files exactly, take the next nearest up or down depending upon up or down column.
4. Repeat until youve exhausted the list of random numbers

With 'burn-in' (sample 1st 100 from) Zone 3 cases

Sampling the GIP Data

GIP#	Up?	Up?
1605	1	Up
1400	-1	Down
1865	1	Up
4020	1	Up
1288	1	Up
4289	1	Up
1909	1	Up
3058	1	Up
1313	1	Up
3247	1	Up
646	1	Up
2326	-1	Down
1420	1	Up
3160	1	Up
4115	-1	Down
3440	1	Up
2827	-1	Down
4012	-1	Down
1384	1	Up
3123	-1	Down
1552	-1	Down
1984	1	Up
916	1	Up
4173	1	Up
1714	-1	Down
295	1	Up
1172	1	Up
368	1	Up
3128	-1	Down
4409	1	Up
2206	1	Up
4303	-1	Down
2605	1	Up
2119	-1	Down
3710	-1	Down
1541	1	Up
2863	-1	Down
4888	-1	Down
2900	-1	Down
491	-1	Down
3978	1	Up
3778	-1	Down
1459	-1	Down
2652	-1	Down
3360	1	Up
204	-1	Down

Part 3

Methodology

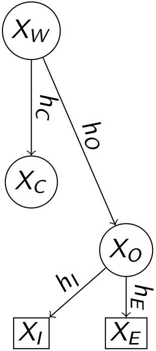


Figure : Multistate Hazard Model for Exoneration Data: X_W - Letter received; X_C - Case Closed; X_I - Case Inculpated; X_E - Case Exonerated.

The Data

State	No. Ever in State	Entries to State				
		X_W	X_C	X_O	X_I	X_E
X_W	3717		2491	558	-	-
X_C	2490	-	-	-	-	-
X_O	558	-	-	-	95	7
X_I	95	-	-	-	-	-
X_E	7	-	-	-	-	-

Figure : 2009-2010 Snapshot of The GIP (pseudo-data)

The Model

Let the *hazard rates* for transition between states be:

$$h_j(t) = \mathbb{P}(X_{\mathbf{Z}}(t+\epsilon) = x_{j,\mathbf{z}_j} | X_{\mathbf{Z}}(t) = x_{j*,\mathbf{z}_{j}*}) \quad (1)$$

In the Figure the hazard rates are labeled with h and the 'states' are labeled by X.
The 'covariate information', $\mathbf{Z}_j$ are the demographic, case, *state duration* information unique to each record.

The Model

The simplest version of the model is to fit proportional hazards

$$h_j(t) = h_0^j exp\{\beta^T \mathbf{Z}^j\} \tag{2}$$

between each pair of adjacent or communicating states X_{j*}, X_j. *This is to treat the state transitions, via the estimated hazard rates, as conditionally independent.*
This is a useful first approach as methods for fitting proportional hazards are straightforward and ubiquitous.

The Model

Consider though that we desire inference on the probability of a case being worthwhile of review. In the context of the model this is the probability, hazard, or survival rate of a case to a time t, or state X_j, given covariates. Let

$$H(t) = H_j(t) = \{Z^j; x_1, ..., x_t\} \tag{3}$$

be the 'history' of a case at time t — the state history and record of time dependent covariates at time t.

The Model

Consider

$$\pi(s|H(t)) = \mathbb{P}(\mathcal{X} = X_E \textit{ in } s > t|H(t)) \tag{4}$$

then to be the probability a case makes it through to exoneration, given its history, by time s. *In the simple model this is to ascertain the (assumed) conditionally independent hazard rates* h_j, evaluate them at the observed covariates and multiply them together.

The Augmented Model

The augmentation is to relax the conditional independence assumption, i.e. the conditionally independent separately estimated hazard functions h_j, by concatenating the entire process across states using a Copula representation of the Chapman-Kolmogorov equations. This elucidation follows the method in [1].

Markov Process

The Chapman-Kolmogorov equations

$$f_{X_{t_1},\ldots,X_{t_n}} = \int_{-\infty}^{\infty} f_{X_{t_n}|X_{t_{n-1}}}(X_{t_n}|X_{t_{n-1}}) \cdots f_{X_{t_2}|X_{t_1}}(X_{t_2}|X_{t_1}) dX_{t_2} \cdots dX_{t_{n-1}}$$

hold that the progression of the random process X_{t_i} is governed by these transition probabilities, 'averaging' probability mass over the conditionally independent states.

'Tunable' Markov Process

Calling $C_{t_i t_j}$ the copula of the random variables X_{t_i}, X_{t_j}, then, for $t_i < t_j < t_k$

$$C_{t_i t_k} = C_{t_i t_j} * C_{t_j t_k} \tag{5}$$

is an equivalent representation of the CK equations, and

$$\mathbb{P}(X_t \in A | X_s = x) = \frac{\partial C_{st}(F_s(x), F_t(a))}{\partial X_s}$$

is the copula version of the CK transition probability.

‘Tunable’ Markov Process

The estimation problem here is to fit the copulae, i.e. the *transition dependence* between states, from data. This is just to write (5) as

$$C_{t_i t_k;\theta_1,\theta_2} = C_{t_i t_j;\theta_2} * C_{t_j t_k;\theta_1}. \tag{6}$$

This yields a likelihood type method

$$(\hat{\theta}_1, \hat{\theta}_2) = \arg\max_{\theta_1,\theta_2} C_{t_i t_k;\theta_1,\theta_2} = C_{t_i t_j;\theta_2} * C_{t_j t_k;\theta_1}. \tag{7}$$

The tunable MSS model

This is just to concatenate the hazard functions at each state h_j by parametric copula via equation (6) by an $m-$fold operation of $*$, m the number of total states, or number of states by desired time t in

$$H(t) = H_j(t) = \{Z^j; x_1, ..., x_t\}$$

.

The parameters of the Copulae (in 7) are fitted via maximum likelihood or sieve method, say, and the proportional hazard model is used for the marginal distribution at each state X_j.

Using this approach we can obtain estimates for the effects of the covariates in $H(t)$, equation (3), using (the Gumbel-Hougard) Copulae to concatenate the state-by-state transitions into a full Markovian process.

Methodology

Exploit length of time in ‘state’...

Methodology

Exploit length of time in 'state'...

Methodology

Exploit length of time in 'state'...

- Proxy for the IP's 'prior' or *ad hoc* model for likeliness of exoneration

Methodology

Exploit length of time in 'state'...

- Proxy for the IP's 'prior' or *ad hoc* model for likeliness of exoneration
- 'Inflate' data via survival curve

Methodology

Exploit length of time in 'state'...

- Proxy for the IP's 'prior' or *ad hoc* model for likeliness of exoneration
- 'Inflate' data via survival curve

...in the presence of covariates

- False Confession?

Methodology

Exploit length of time in ‘state’...

- Proxy for the IP’s ‘prior’ or *ad hoc* model for likeliness of exoneration
- ‘Inflate’ data via survival curve

...in the presence of covariates

- False Confession?
- Snitch?

Methodology

Exploit length of time in 'state'...

- Proxy for the IP's 'prior' or *ad hoc* model for likeliness of exoneration
- 'Inflate' data via survival curve

...in the presence of covariates

- False Confession?
- Snitch?
- Race Black?

Methodology

Exploit length of time in 'state'...

- Proxy for the IP's 'prior' or *ad hoc* model for likeliness of exoneration
- 'Inflate' data via survival curve

...in the presence of covariates

- False Confession?
- Snitch?
- Race Black?
- Victim White?

Following [14] ([11]) approximate this with 'conditional' proportional hazard curves, on 'left-truncated' data

Part 3

Methodology: Conditionally Independent Model

Cox proportional hazards (Conditionally Independent Model)

Methodology: Conditionally Independent Model

Cox proportional hazards (Conditionally Independent Model)

$$h_j(t) = h_0^j exp\{\beta^T \mathbf{Z}^j\} \tag{8}$$

...in the presence of covariates $\mathbf{Z}$

- $Z_1^j = 1$ False Confession? Yes.

Methodology: Conditionally Independent Model

Cox proportional hazards (Conditionally Independent Model)

$$h_j(t) = h_0^j exp\{\beta^T \mathbf{Z}^j\} \tag{8}$$

...in the presence of covariates $\mathbf{Z}$

- $Z_1^j = 1$ False Confession? Yes.
- $Z_2^j = 1$ Snitch? Yes.

Methodology: Conditionally Independent Model

Cox proportional hazards (Conditionally Independent Model)

$$h_j(t) = h_0^j exp\{\beta^T \mathbf{Z}^j\} \tag{8}$$

...in the presence of covariates $\mathbf{Z}$

- $Z_1^j = 1$ False Confession? Yes.
- $Z_2^j = 1$ Snitch? Yes.
- $Z_3^j = 1$ Race Black? Yes.

Methodology: Conditionally Independent Model

Cox proportional hazards (Conditionally Independent Model)

$$h_j(t) = h_0^j exp\{\beta^T \mathbf{Z}^j\} \tag{8}$$

...in the presence of covariates $\mathbf{Z}$

- $Z_1^j = 1$ False Confession? Yes.
- $Z_2^j = 1$ Snitch? Yes.
- $Z_3^j = 1$ Race Black? Yes.
- $Z_4^j = 1$ Victim White? Yes

Methodology: Conditionally Independent Model

Cox proportional hazards (Conditionally Independent Model)

$$h_j(t) = h_0^j exp\{\beta^T \mathbf{Z}^j\} \tag{8}$$

...in the presence of covariates **Z**

- $Z_1^j = 1$ False Confession? Yes.
- $Z_2^j = 1$ Snitch? Yes.
- $Z_3^j = 1$ Race Black? Yes.
- $Z_4^j = 1$ Victim White? Yes
- $Z_5^j =$ Duration in previous state

Only Z_5^j is really 'time-varying'

Methodology

$$h_0^j \equiv h_0$$

Methodology

$h_0^j \equiv h_0$

Z	coef	exp(coef)	sig?
Confess?	0.36	1.03	*
Snitch?	-.59	.55	
Black?	-.093	.91	
Victim White?	-.16	.85	**
Duration in Prev. State	1.02	2.76	

Methodology

$h_0^j \equiv h_0$

Z	coef	exp(coef)	sig?
Confess?	0.36	1.03	*
Snitch?	-.59	.55	
Black?	-.093	.91	
Victim White?	-.16	.85	**
Duration in Prev. State	1.02	2.76	

Interpretation? Initial review process? Unclear interpretation since 'hazard' (prob. of exiting state) means something different in between different states.

Methodology

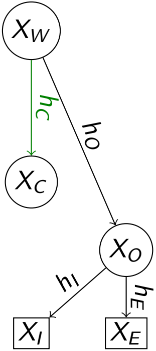


Figure : Multistate Hazard Model for Exoneration Data: X_W - Letter received; X_C - Case Closed; X_I - Case Inculpated; X_E - Case Exonerated.

Methodology

$$h_0^j,\ j = C$$

Methodology

$h_0^j,\ j = C$

Z	coef	exp(coef)	sig?
Confess?	-0.49	0.61	**
Snitch?	0.0053	1.005	
Black?	-.081	.92	*
Victim White?	-.003	.99	
Duration in Prev. State	0.609	1.83	

Methodology

$h_0^j,\ j = C$

Z	coef	exp(coef)	sig?
Confess?	-0.49	0.61	**
Snitch?	0.0053	1.005	
Black?	-.081	.92	*
Victim White?	-.003	.99	
Duration in Prev. State	0.609	1.83	

Interpretation? Cases selected because of 'false confession' claim in intake are not quickly dispensed of. Some cases may 'linger' but then closed anyway.

Methodology

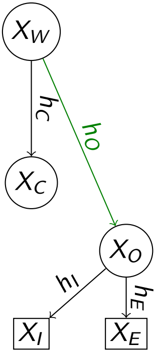


Figure : Multistate Hazard Model for Exoneration Data: X_W - Letter received; X_C - Case Closed; X_I - Case Inculpated; X_E - Case Exonerated.

Methodology

$h_0^j,\ j = O$

Methodology

$h_0^j,\ j = O$

Z	coef	exp(coef)	sig?
Confess?	0.0181	1.02	
Snitch?	0.418	1.51	
Black?	-.1809	.83	
Victim White?	0.06	1.06	
Duration in Prev. State	0.522	1.685	***

Methodology

$h_0^j,\ j = O$

Z	coef	exp(coef)	sig?
Confess?	0.0181	1.02	
Snitch?	0.418	1.51	
Black?	-.1809	.83	
Victim White?	0.06	1.06	
Duration in Prev. State	0.522	1.685	***

Interpretation? Cases actually worked. Duration in $X_W = \textit{letter received}$ significant implies cases wait awhile?

Methodology

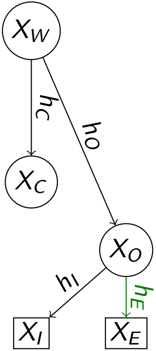


Figure : Multistate Hazard Model for Exoneration Data: X_W - Letter received; X_C - Case Closed; X_I - Case Inculpated; X_E - Case Exonerated.

Methodology

$h_0^j, \ j = E$

Methodology

$h_0^j,\ j = E$

Z	coef	exp(coef)	sig?
Confess?	0.917	1.02	*
Snitch?	-0.037	0.963	
Black?	-0.326	0.722	***
Victim White?	0.053	1.065	
Duration in Prev. State	0.00323	1.003	

Methodology

$h_0^j,\ j = E$

Z	coef	exp(coef)	sig?
Confess?	0.917	1.02	*
Snitch?	-0.037	0.963	
Black?	-0.326	0.722	***
Victim White?	0.053	1.065	
Duration in Prev. State	0.00323	1.003	

Interpretation? All the GIP exonerees are black thus far.

Methodology

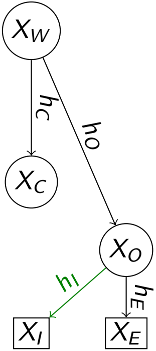


Figure : Multistate Hazard Model for Exoneration Data: X_W - Letter received; X_C - Case Closed; X_I - Case Inculpated; X_E - Case Exonerated.

Methodology

$h_0^j,\ j = I$

Methodology

$h_0^j,\ j = I$

Z	coef	exp(coef)	sig?
Confess?	0.024	1.024	
Snitch?	-0.066	1.069	
Black?	-0.0571	1.058	
Victim White?	0.0573	1.059	
Duration in Prev. State	-0.973	.907	**

Methodology

$h_0^j,\ j = I$

Z	coef	exp(coef)	sig?
Confess?	0.024	1.024	
Snitch?	-0.066	1.069	
Black?	-0.0571	1.058	
Victim White?	0.0573	1.059	
Duration in Prev. State	-0.973	.907	**

Interpretation? The longer cases waited in the previous state, the longer it took to inculpate. Problems processing cases efficiently?

Part 3

Augmented Model

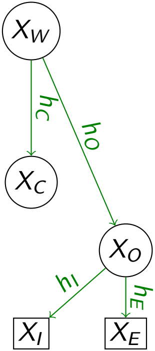


Figure : Multistate Hazard Model for Exoneration Data: X_W - Letter received; X_C - Case Closed; X_I - Case Inculpated; X_E - Case Exonerated., $\mathbf{h}_\Theta = (h_C, h_O, h_E, h_I)$

Augmented Model

$$\Theta = (\theta_C, \theta_O, \theta_I, \theta_E)$$

are the dependence parameters for the augmented model and can be interpreted as how far away from conditional independence the states are. MLE estimates, (with s.e's) are:

$$\hat{\Theta}_{MLE} = (\hat{\theta_C} = .721(.366), \hat{\theta_O} = .226(.490), \hat{\theta}_I = .293(.158), \hat{\theta}_E = .910(.478))$$

Augmented Model

Z	coef	exp(coef)	sig?
Confess?	-0.39	0.677	*
Snitch?	-.52	.594	*
Black?	.092	1.09	*
Victim White?	-.163	.849	**
Duration in Prev. State	-1.72	0.179	

Figure : Estimated ‘effects’ from Augmented Model for $H(t)$

Augmented Model

Consider again

$$\pi(s|H(t),\mathbf{Z}) = \mathbb{P}(\mathcal{X} = X_E \text{ } in \text{ } s > t|H(t),\mathbf{Z}) \tag{9}$$

the probability a case makes it through to exoneration, given its history, by time s. Pick covariates $\mathbf{Z}$.

- For the conditionally independent model this is just a multiplication of hazard curves evaluated at $\mathbf{Z}$...
- For the augmented model, we concatenate the curves using the estimators $\hat{\Theta}$

Part 3

Augmented Model

Set $\mathbf{Z}$ = (No Confess, No Snitch, Non White Victim, Mean Time in each previous state).

- Conditionally Independent $\hat{\pi}(s = 100, \mathit{Black}) = 0.001833239$

Augmented Model

Set $\mathbf{Z}$ = (No Confess, No Snitch, Non White Victim, Mean Time in each previous state).

- Conditionally Independent $\hat{\pi}(s = 100, Black) = 0.001833239$
- Conditionally Independent $\hat{\pi}(s = 100, White) = 0.00036903$

Augmented Model

Set $\mathbf{Z}$ = (No Confess, No Snitch, Non White Victim, Mean Time in each previous state).

- Conditionally Independent $\hat{\pi}(s = 100, Black) = 0.001833239$
- Conditionally Independent $\hat{\pi}(s = 100, White) = 0.00036903$
- Augmented $\hat{\pi}(s = 100, Black) = 0.01076137$

Augmented Model

Set $\mathbf{Z}$ = (No Confess, No Snitch, Non White Victim, Mean Time in each previous state).

- Conditionally Independent $\hat{\pi}(s = 100, \mathit{Black}) = 0.001833239$
- Conditionally Independent $\hat{\pi}(s = 100, \mathit{White}) = 0.00036903$
- Augmented $\hat{\pi}(s = 100, \mathit{Black}) = 0.01076137$
- Augmented $\hat{\pi}(s = 100, \mathit{White}) = 0.01237557$

The conditionally independent estimates are much lower and the magnitudes of the probabilities reverse.

Part 4

Null	A pre-requisite for statistical modeling and inference
Case-Control	More expensive, requires good matching: *care must be taken to aggregate cases across States, IPs, etc.*
Bayesian	$\beta_j \sim ?$
'Better' Conditionally Independent MP	$(\sum_j \alpha_j h_0^j)\ exp\{\beta_j \mathbf{Z}^j\}; \alpha_j \sim ?$

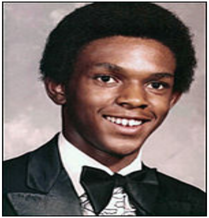
Good further methods need to account for: data collection, sampling, and estimation under non-stationarity and non-independence

Part 4

Motivation

The *Sine Qua Non* of Wrongful Conviction

Timothy Brian Cole: 7.1.1960 - 12.2.1999

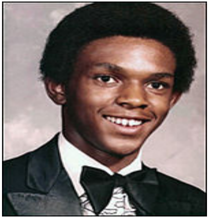


- Texas Tech Student Convicted of Rape in 1985

Motivation

The *Sine Qua Non* of Wrongful Conviction

Timothy Brian Cole: 7.1.1960 - 12.2.1999

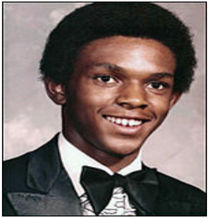


- Texas Tech Student Convicted of Rape in 1985
- Died in prison from asthma attack in 1999, refused parole $\leftarrow$ refused to admit guilt

Motivation

The *Sine Qua Non* of Wrongful Conviction

Timothy Brian Cole: 7.1.1960 - 12.2.1999

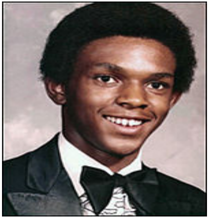


- Texas Tech Student Convicted of Rape in 1985
- Died in prison from asthma attack in 1999, refused parole $\leftarrow$ refused to admit guilt
- Jerry Johnson had confessed to rape in 1995; Court exoneration (posthumous) 2007-2007

Motivation

The *Sine Qua Non* of Wrongful Conviction

Timothy Brian Cole: 7.1.1960 - 12.2.1999



- Texas Tech Student Convicted of Rape in 1985
- Died in prison from asthma attack in 1999, refused parole $\leftarrow$ refused to admit guilt
- Jerry Johnson had confessed to rape in 1995; Court exoneration (posthumous) 2007-2007

GOAL: Assist (in particular) the Innocence Network's Exoneration Work

References I

- Abayomi, K and Hawkins, L. "Copulas for tunable markov processes: Heuristics for extreme-valued labor market outcomes: Fortune 500 ceos vs. professional athletes." *Proceedings of the 2009 Joint Mathematics Meetings*, **16**, 2010.
- Berry, D. (1990) "DNA Fingerprinting: What Does it Prove?" *Chance*, **3**. 15-36.
- Darsow, W., Nguyen, B., and Olsen, E. (1992) "Copulas and Markov Processes." *Illinois Journal of Mathematics.* **36**, 4. pp. 600-642.
- Gabel, J. Wilkinson, M. (2008) "Good Science Gone Bad: How the criminal justice system can redress the impact of flawed forensics." *Hastings Law Journal.* May 2008.
- Garrett, B. (2010) "The Substance of False Confessions." *Stanford Law Review*, **62**, 1051-1119.
- Garrett, B. (2008) "Judging Innocence." *Columbia Law Review.*, **108**, 1-71.

References II

- Gross, S., et al. (2005) Exonerations in the United States: 1989 Through 2003. *The Journal of Criminal Law & Criminology.* **95**, 2.
- The Innocence Project, `http://www.innocenceproject.org`, last visited August 1, 2011.
- Kaye, D. (2007) "The Science of DNA Identification: From the Laboratory to the Courtroom (and Beyond)." *Minn. J.L. Sci & Tech.* **8**, 2. 409-427.
- Chakraborty, R., Kidd, K. (1991) "The Utility of DNA Typing in Forensic Work." *Science.*, **254**, 5039. 1735-1739.
- (1992) "Random Truncation Models and Markov Processes" *Annals of Statistics*, **19**, 582-602.
- Lander, E. (1989) "DNA fingerprinting on trial." *Nature*, **339**. 501-505.

References III

"Strengthening Forensic Science in the United States: A Path Forward. February 2009. *Committee on Identifying the Needs of the Forensic Sciences Community; Committee on Applied and Theoretical Statistics, National Research Council.*

Silverstein et. al (1999) "Clinical course and costs of care for Crohn's disease: Markov model analysis of a population based cohort" *Gastroenterology* **117**, 49-57.