Copula Based Multistate Hazard Model: An Inferential Methodology for Innocence Project

Kobi Abayomi[*] Jessica Gabel[†] Otis Jennings[‡]

Abstract

Since 1992, the Innocence Projects (IP) have helped over 270 wrongly convicted persons prove their innocence and gain freedom. Despite the important successes of the IP's these exonerations may be a miniscule sample of the number of wrongly convicted persons who languish in prison. Statistical research on possible identifiers of likely 'exonerable' cases does not exist: current work is merely descriptive and non-inferential. We demonstrate an inferential methodology designed for the Innocence Projects: a multi-state hazard model with an augmentation via the parametric copula. Our approach is designed to exploit records that are readily available to exoneration workers. This approach offers a coherent, statistical framework for identification of significant and important factors on data that are readily available to the IP's

Key Words: Hazard Model, Exoneration Data, Parametric Copula, Markov Processes

1. Introduction

On April 9th, 2009 Timothy Cole was posthumously exonerated after serving thirteen years in prison, of a twenty-five year sentence, for a rape he did not commit. Mr. Cole maintained his innocence throughout his trial, conviction and imprisonment but died before DNA testing exonerated him 2008. The Innocence Project of Texas was able to obtain DNA analysis on items from the crime scene. The tests conclusively identified another man as the real perpetrator, Jerry Wayne Johnson, who had confessed to the crime in 1995 four years before Cole died in prison. Timothy received a full pardon on March 1, 2010 [4].

On September 21, 2011 Troy Davis was executed by the Georgia State Department of Corrections after a protracted battle to introduce possibly exculpatory evidence [6]. Unfortunately for Mr. Davis his claim to innocence did not include any DNA evidence; this the case for the majority of people who seek post-conviction relief.

News reports of recently exonerated men and women — often after long periods of incarceration — are no longer infrequent [7]. In some cases, violent criminals remain free and the wrongly convicted are punished instead [20]. In addition, forensic procedures have been exposed as inadequate and inaccurate [16]. The methods by which the convicted can petition for relief are few; the procedures legal advocates use for identification and redress of likely candidates are *ad hoc* [7].

1.0.1 The Innocence Networks and...

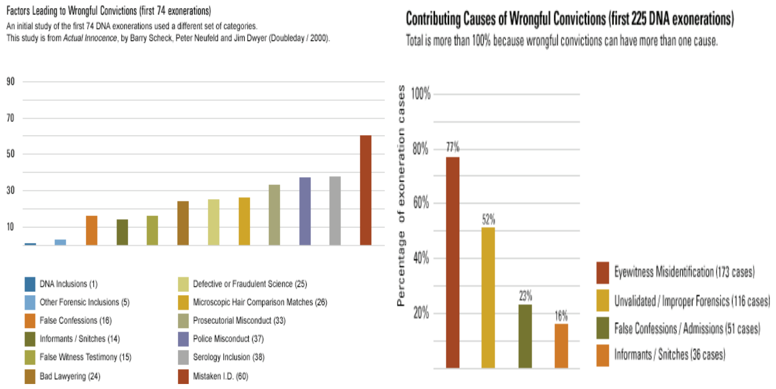
In 2009, the work of Innocence Network member organizations led to the exoneration of 27 people in the United States. Since 1992, the Innocence Projects have helped over 270 (at this writing) wrongly convicted persons prove their innocence and gain freedom. Unfortunately, these exonerations are likely a miniscule sample of the number of wrongly convicted persons. Recent research suggests that the number innocents languishing in prison may be greater than 28,500, in non-death penalty cases alone [10]. Many of these cases slip through

[*] Georgia Tech, ISyE, 765 Ferst Dr, Atlanta, GA 30331

[†] Georgia State University, College of Law, Atlanta, GA

[‡] Duke University, Fuqua School of Business, Durham, NC

the cracks: DNA evidence is unavailable, non-existent or insufficient to merit review for exoneration.



(a) 74 cases **(b)** 225 cases

Figure 1 (a): Factors counted in the first 74 Innocence Project DNA exonerations, see [18] (b): Factors counted in the first 225 DNA exonerations, see [17]. Notice that errors due to improper forensics are 52% of the first 225 cases (right hand panel) but only 30% of the first 74 cases (left hand panel). Some of this difference may be researcher discretion - different coders for each panel; some of the difference may be an increased focus on forensic sophistication. These panels - and their apparent differences - likely obscure valuable, latent information.

1.0.2 ...Barriers to Post-Conviction Relief

Current statutes that provide for post-conviction DNA testing are a good starting point for redressing wrongful convictions. Nonetheless, they are a fragmented and often unrealistic approach as the thresholds for post-conviction DNA testing are set almost impossibly high. Moreover, post conviction DNA testing statutes assume that DNA is available, testable, and capable of demonstrating innocence. These assumptions only apply to a handful of cases. In most jurisdictions, the ability to directly challenge a conviction evaporates after three years [7]. Moreover, while every state has at least one postconviction remedy by which a prisoner can assert innocence for a crime, these remedies have exacting stardards that are nearly impossible to meet [7].

1.1 Background: Exoneration Research

The Innocence Network member organizations are the primary, but not the sole, actors for post-conviction relief [17]. The IP network specializes in 'DNA cases' - cases where the primary exculpatory evidence is a negative DNA match - and count 258 post conviction DNA exonerees.

Gross et al ([10]) 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. It is important to note that DNA evidence has assumed an exculpatory role relatively recently [11]. DNA testing for identification in criminal forensics was initially critiqued as too error prone to meet a legal evidentiary standard: see [3], [13] or [2]. 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 [16]. DNA identification (or non-identification) has become indispensable to exoneration work. At the same time, however, while DNA is vital to redress a wrongful conviction, its absence weakens the cases — the vast majority of exoneration requests — where there simply is no DNA evidence available [16].

Unfortunately, these are the cases that are most often and easily forgotten. While these cases are different in that they lack DNA evidence, the errors are fundamentally the same: eyewitness misidentifications, false confessions, jailhouse snitches, and flawed forensics — regardless of whether DNA could be used to exonerate the accused [7].

1.2 Current Work

The current research has examined the problem — what is statistically unique and identifiable about the wrongfully convicted: in their demographics, prior criminal records, and the facts and location of their conviction — through the framework of DNA testing: exclusion and non-identification [8]. Yet, such cases are only a portion of the entreaties the Innocence Projects receive and just a fraction of the potentially large numbers of wrongfully convicted [9]. Further, statistically unique and identifiable characteristics in the data — demographics, prior criminal records, the facts and locations of conviction — have not been addressed at all [10].

1.3 Our Framework

We offer a methodology to address factors that may predict or be associated with wrongful convictions by exploiting the *internal* case records that the IP's collect in the course of their exoneration work. We present a version of a Multi-State Hazard/Survival (MSS) model, augmented with parametric Copulae functions, as a model for the processing the IP's conduct on cases they receive for exoneration review. This approach exploits the length of time a case — a request for exoneration — persists in operational 'states' as a proxy for the IP's 'prior' or expert knowledge. Inference on a model for the length of time and probability of transition between 'states' — level and amount of work at an IP — yields an *ad hoc* model for the determinants, and likelihood, of exoneration in the presence of case level covariates.

Below we discuss the necessary *internal* data, in context with the model, its augmentation, and (simulated) results on an inflated sample of pseudo-data.

2. Methodology

Our method is to fit an augmented version of the multi-state (Markovian) hazard model well illustrated on clinical data by Silverstein et. al [19]. The ordinary version of the model is a special case of a random truncated model on a Markov process [12]. In [19], for instance, the 'states' are classifications of levels of illness and recovery from Crohn's disease; here the 'states' will be waypoints in the procedural investigation of a case at an Innocence Project. The model is a 'hazard' model as the interstate transition probabilities are be expressed as hazard rates via (semi-parametric) proportional hazards, conditional on covariates.

Our augmentation is to concatenate the conditional, intra-state, hazards using parametric Copulae functions as versions of the ordinary Chapman-Kolmogorov equations. This approach was outlined by Darsow et. al [5], an example is found in [1]

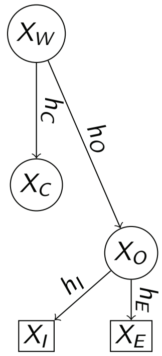


Figure 2 Multistate Hazard Model for Exoneration Data: X_W - Letter received; X_C - Case Closed; X_l - Case Inculpated; X_E - Case Exonerated.

2.1 Data

In the MSS model we gain inference by measuring the effect of a covariate on the hazard rates for state transition. In layman's terms: this is how much time cases with particular covariates take to transition from one state to another, if at all. In analog to clinical trials, this time is often how long a patient lives before progressing to a different stage of illness/recovery. The 'hazard rate' then is the effective rate of transition per time unit. The MSS model allows us to calculate this rate for particular values of the covariates, to access significance and comparative effect.

Our illustration for IP data is general enough to be applicable to most organizations: Let $\mathcal{X}$ be a collection of states for the Markov Process and let $\mathcal{Z}$ be associated covariates. In this illustration we let: $\mathcal{X}$ = (X_W - Letter received, X_C - Case Closed, X_l - Case Inculpated, X_E - Case Exonerated); and the covariates $\mathcal{Z}$ = ($Z_1^j = 1$ False Confession?, $Z_2^j = 1$ Snitch?, $Z_3^j = 1$ Race Black?, $Z_4^j = 1$ Victim White?, $Z_5^j =$ Duration in previous state). $Z_1^j, ..., Z_4^j$ 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 is to allow the model to account for the amount of information available to the IP at each stage in their case processing.

These data can easily be collected by any IP which keeps even minimal records for their internal files.

	No. Ever	Entries to State				
State	in State	X_W	X_C	X_O	X_I	X_E
X_W						
X_C						
X_O						
X_I						
X_E						

Figure 3 Sample template for collecting 'state' data $\mathcal{X}$

In this illustration we have 'inflated' a severely reduced subset ($n = 223$) of Georgia Innocence Project (GIP) data by resampling with replacement. We have substituted covariate information from the GIP files with publicly available information from the Georgia Department of Corrections. This yielded the $n = 3717$ pseudo-observations in Figure 4.

	No. Ever		Entries to State			
State	in 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 4 GIP pseudo-data

2.2 Model

Let the hazard rates for transition between states be:

$$h_j(t) = \mathbb{P}(X(t+\epsilon) = x_j | X(t) = x_{j*}) \tag{1}$$

In Figure 2 the hazard rates are labeled with h and the 'states' are labeled by X. The 'covariate information' are the demographic, case, *state duration* information unique to each record.

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.

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. Consider

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

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.

3. 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].

3.1 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 ([14]).

3.2 Copula approach

Take $X_1 \sim F_{X_1},\ X_2 \sim F_{X_2}$ and set $U = F_{X_1}$ and $V = F_{X_2}$; the pair (U, V) are the '*grades*' of (X_1, X_2) i.e. the mapping of (X_1, X_2) in F_{X_1}, F_{X_2} space. A copula is a function that takes the 'grades' as arguments and returns a joint distribution function, with marginals F_{X_1}, F_{X_2}.

$$C(U, V) = F_{X_1,X_2}$$

Any multivariate distribution function can yield a copula function,

$$F_{X_1,X_2}(F^{-1}_{X_1}(U), F^{-1}_{X_2}(V)) = C^{'}(U, V)$$

that it: the correspondence which assigns the value of the joint distribution function to each ordered pair of values (F_{X_1}, F_{X_2}) for each X_1, X_2 is a distribution function called a copula ([15]).

Joint distributions are specified by *marginal* and *dependence* parameters; for example a bivariate exponential distribution

$$H_\theta(x_1, x_2) = 1 - e^{-\lambda_1 x_1} - e^{-\lambda_2 x_2} + e^{-(\lambda_1 x_1 + \lambda_2 x_2 + \theta x_1 x_2)}$$

has marginal parameters λ_1, λ_2 and dependence parameter θ. The copula version for this joint distribution is

$$C_\theta(u, v) = H(-ln(1-u), -ln(1-v)) =$$
$$= (u + v - 1) + (1-u)(1-v) * e^{-\theta \ln(1-u)\ln(1-v)}$$

and the marginal parameters, still extant, are sublimated in the probability integral transformation of $U = F_{X_1;\lambda_1}, V = F_{X_2;\lambda_2}$

3.3 Markov Processes via Copula: Darsow, Nguyen and Olsen

Following Darsow, Nguyen, Olsen ([5]), define

$$(A * B)(x_1, x_2) = \int_0^1 \frac{\partial A(x_1, t)}{\partial x_2} \cdot \frac{\partial B(t, x_2)}{\partial x_1} dt$$

for A, B copulas and x_1, x_2 in I. Since, for $X_1, X_2 \sim F_{X_1}, F_{X_2}, C$

$$\mathbb{P}(X_1 < x_1 | X_2 = x_2) = \frac{\partial C(F_{X_1}, F_{X_2})}{\partial X_2}$$

and

$$\mathbb{P}(X_2 < x_2 | X_1 = x_1) = \frac{\partial C(F_{X_1}, F_{X_2})}{\partial X_1}$$

then, for any three random variables X_1, X_2, X_3, where $(X_1 \perp X_3)|X_2$

$$C_{X_1X_3} = C_{X_1X_2} * C_{X_2X_3}$$

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

$$C_{t_it_k} = C_{t_it_j} * C_{t_jt_k} \tag{5}$$

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.

3.4 'Tunable' Markovian models via Parametric Copula

A markov process is 'conventionally' specified by a set of initial distributions F_0 and a family of transition probabilities $f_{X_i|X_j}(X_i|X_j)$; as an estimation problem, the goal is to estimate these transition probabilities from data.

In this copula based approach we assign the marginal distributions for each state F_{X_1},..., F_{X_m}, and specify family of copulas satisfying (5). 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

$$C_{t_it_k;\theta_1,\theta_2} = C_{t_it_j;\theta_2} * C_{t_jt_k;\theta_1}. \tag{6}$$

This yields a likelihood type method

$$(\hat{\theta}_1, \hat{\theta}_2) = \arg\max_{\theta_1,\theta_2} C_{t_it_k;\theta_1,\theta_2} = C_{t_it_j;\theta_2} * C_{t_jt_k;\theta_1}. \tag{7}$$

for fitting copula as transition probabilities, and an interpretation of the estimated parameters as the *transitional dependence* for the fitted Markov process. The copula dimensions match that of the transition probabilities: the simplest and special case is for 2-copula for pairwise conditional transitions.

This method is especially useful in Markov process estimation problems where: marginal distributions are available for each state; where the initial distribution of the process is non-informative; and where parametric models for *transition dependence* are desirable.

3.5 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 and the inflated pseudo-data in Figure 4 we can obtain estimates for the effects of the covariates in $H(t)$, equation (3), using the Gumbel-Hougard copula to concatenate the state-by-state transitions into a full Markovian process.

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	

Figure 5 Estimated ‘effects’ from pseudo-data for $H(t)$

There is no real interpretation to generate here, as these are pseudo-data resampled from a small portion of the full GIP data. Note that interpretation of the mock effects is heuristically equivalent to those in an ordinary proportional hazards model: larger coefficients indicate a greater hazard, i.e. a quicker time to the end ‘state’. Note as well that $\pi(s|H(t))$ (equation 4) is a function of the transitional dependence parameters $\Theta = (\theta_1, ..., \theta_{|\mathcal{X}|})$, as such a function of the particular copulae used in the model. Parameter significance is generated via bootstrap resampling here.

4. Conclusion

We offer this model as a coherent approach to modeling the effect of covariates on the probability of transversing the stages of case-work by an Innocence Project. We do note that there are strong endogenous and exogenous differences across IP’s, even those working with the same state: Different projects have different ‘rules’ and procedures; the mechanisms available to redress wrongful convicts differ state by state; record-keeping and coding of data may be inconsistent between IP’s.

We believe even the naive version of this model — conditionally independent hazards between pairs of states — can be immensely useful to exoneration workers and researchers. First: this model can be generated from data the IP’s already have available and serves as an proxy for a ‘case-control’ approach where internal records are compared to *external* ones. Instances of exoneration not handled by the IP’s as well as cases of ‘true-positive’ guilt to serve as matched controls are necessary though difficult, if not impossible, to collect. Second: fitting this model requires the IP data to be organized in a way that highlights the time dependency of the case-processing. This may usefully generate further case review.

The augmented model is (more fully) semi-parametric: the marginal distributions are (semi-parametric) proportional hazards while the joint distribution over the process is fit with parametric copula. For models with few states and covariates the parameters can be estimated by brute-force versions of maximum likelihood; for larger models we expect this method of estimation to be problematic. We specifically consider estimation, and the properties of estimators — most importantly the effect coefficients —- in an upcoming full version of the paper on complete data.

References

[1] Kobi Abayomi and Lee Hawkins. Copulas for tunable markov processes: Heuristics for extreme-valued labor market outcomes: Fortune 500 ceo’s vs. professional athletes. *Proceedings of the 2009 Joint Mathematics Meetings*, 16, 2010.

[2] D. Berry. Dna fingerprinting: What does it prove? *Chance*, 3:15–36, 1990.

[3] R. Chakraborty and K. Kidd. The utility of dna typing in forensic work. *Science*, 254:1735–1739, 1991.

[4] The Houston Chronicle. Perry pardons exonerated convict posthumously. March 1, 2010, 2010.

[5] W. Darsow, B. Nguyen, and E. Olsen. Copulas and markov processes. *Illinois Journal of Mathematics*, 36:600–642, 1990.

[6] Troy Davis. Free troy davis. `www.troydavis.org`, 2011.

[7] J. Gabel and M. Wilkinson. 'good science gone bad: How the criminal justice system can redress the impact of flawed forensics. *Hastings Law Journal*, 2008.

[8] B. Garrett. Judging innocence. *Columbia Law Review*, 108:1–71, 2010.

[9] B. Garrett. The substance of false confession. *Stanford Law Review*, 62:1051–1119, 2010.

[10] S. Gross. Exonerations in the united states: 1989 through 2003. *The Journal of Criminal Law & Criminology*, 95, 2005.

[11] D. Kaye. The science of dna identification: From the laboratory to the courtroom (and beyond). *Minn. J.L. Sci & Tech.*, 8:409–427, 2007.

[12] Keiding. Random truncation models and markov processes. *Annals of Statistics*, 19:582–602, 1992.

[13] E. Lander. Dna fingerprinting on trial. *Nature*, 339:501–505, 1989.

[14] G. Lawler. *Introduction to Stochastic Processes.* Chapman and Hall/CRC Press, New York, 2000.

[15] R. Nelsen. *An Introduction to Copulas*. Springer, New York, 2006.

[16] Committee on Identifying the Needs of the Forensic Sciences Community; Committee on Applied and National Research Council. Theoretical Statistics. Strengthening forensic science in the united states: A path forward. February 2009, 2009.

[17] The Innocence Project. Innocence project reports. `http://www.innocenceproject.org`, last visited August, 30, 2011, 2011.

[18] B. Sheck, P. Neufeld, and J. Dwyer. *Actual Innocence*. Doubleday, New York, 2000.

[19] D. Silverstein. Clinical course and costs of care for crohn's disease: Markov model analysis of a population based cohort. *Gastroenterology*, 117:49–57, 1999.

[20] New York Times. Unyielding in his innocence, now a free man. October 29, 2009, 2009.