Sexual Assault and the Doctrine of Chances

Ryan Wallentine
Utah State University

SEXUAL ASSAULT AND THE DOCTRINE OF CHANCES

by

Ryan Wallentine

Thesis submitted in partial fulfillment
of the requirements for the degree

of

DEPARTMENTAL HONORS

in

Mathematics
in the Department of Mathematics and Statistics

Approved:

Thesis/Project Advisor
Dr. David Brown

Departmental Honors Advisor
Dr. David Brown

Director of Honors Program
Dr. Kristine Miller

UTAH STATE UNIVERSITY
Logan, UT

Spring 2016

Abstract

Sexual assault is a crime whose offenders often commit multiple acts and its victims experience devastating effects. The doctrine of chances is a rule of evidence that may allow evidences of these past events or circumstances to be presented in a court case given they meet certain criteria. This research argues the probability of being innocently prosecuted for rape multiple times is sufficiently low to meet at least one of the criteria for the doctrine of chances to be used in a sexual assault case. Additional implications and related areas of research are included as well.

Acknowledgements

I would first like to acknowledge Dr. David Brown, the faculty member who has been very instrumental in helping me progress in this research. Throughout this project, he provided me with insight and opportunities that have helped give me both the means and motivation to push through the more difficult parts of this research. I would also like to thank Dr. Chris Corcoran for being willing to sign on as a committee member of this project.

I would like to also thank the Honors Program and Department of Mathematics and Statistics at Utah State University for giving me so many opportunities to learn how to think, how to research, and how to further apply these skills to my life. The support from members of both of these organizations have been very helpful to me both in and out of the classroom.

I would also like to acknowledge events such as Research on Capitol Hill and the Student Research Symposium that have given me opportunities to practice communicating my research and findings to different audiences.

TABLE OF CONTENTS

(Word count: 6023)

Introduction

Sexual assault and rape can have devastating and long-lasting effects on victims from which it is difficult to recover. In contrast to the punishment and shame that accompanies victims of the crime, many assailants walk free without even a prosecution, and the majority of them don't even get reported. Victims often have concerns that law enforcement won't believe them, that there is not enough proof, or that they don't want to be further humiliated. These concerns coupled with the already traumatizing experience of sexual assault are enough to prevent many victims from reporting the crime. Even when the crime is reported, if the case proceeds to court, it is difficult to prove that a rape occurred because it is difficult to find proof that sexual intercourse occurred or that this intercourse was not consensual.

It is in this court setting that the doctrine of chances (sometimes called the doctrine of objective chances) could be applied to help in the prosecution of such cases. The doctrine of chances (DOC) is a rule of evidence that "may allow evidences of other events and circumstances outside the charges in question, based on 'the objective improbability of the same rare misfortune befalling one individual over and over.'"[1] It is a non-character theory that accepts the fact that unfortunate or suspicious circumstances sometimes befall individuals. At the same time, it seeks to show the circumstance in question is unlikely to repeatedly befall an innocent individual.

[1] Smoland, Dain. "Keep Calm and Argue the Facts: A Pragmatic Approach to the Doctrine of Chances." Utah Bar Journal 26, no. 5 (September/October 2013): 45-49.

In order for the DOC to be triggered in a sexual assault case, the evidence must meet four requirements: materiality, similarity, independence, and frequency. First, materiality means the occurrence of a criminal act is disputed such that the prosecution needs to use the outside evidence to prove that it occurred.[2] Second, the evidence must be from an event that was similar to the charges in question. Third, the evidence must be from an accusation that was completely independent from the accusations at hand. Finally, the frequency of these similar and independent circumstances must be greater than the frequency that a typical person could be expected to experience these events.

This paper will focus on proving the qualification of the last requirement of the DOC in a rape case by finding the probability that an innocent person is prosecuted for rape more than once. In the first section, a summary of the statistics and data that formed a basis for the calculations in this research will be provided along with some useful graphics. The second section will introduce the mathematical concepts and models that were used to develop the main processes and equations. The third section will outline in detail these processes and equations leading up to the final conclusions along with the conclusions themselves. The fourth section will explain alternative methods that were explored throughout the duration of this research and why they were not used in the third section. The fifth and final section will sum up the final conclusions of this research and give an idea of the direction in which this research can be taken hereafter.

---

[2] Imwinkelried, Edward J., "The Use of Evidence of an Accused's Uncharged Misconduct to Prove Mens Rea: The Doctrines Which Threaten to Engulf the Character Evidence Prohibition," Ohio State Law Journal, vol. 51, no. 3 (1990), 575-604. https://kb.osu.edu/dspace/bitstream/handle/1811/64070/OSLJ_V51N3_0575.pdf.

Section 1: RAINN Statistics and the Lisak and Miller Study

RAINN (The Rape, Abuse and Incest National Network) is America's largest anti-sexual assault organization that has published statistics concerning sexual assault and rape. The statistics that have been most applicable to this research are those concerning reporting, prosecution, and imprisonment rates for rapists. On the RAINN website, it shows that out of 100 rapes, 32 get reported to the police, 7 lead to an arrest, 3 are referred to prosecutors, and 2 spend time in jail.[3]

In the study "Repeat Rape and Multiple Offending Among Undetected Rapists," David Lisak and Paul M. Miller surveyed 1,882 students attending a mid-sized, urban commuter university of diverse ages and ethnicity. Out of these students, 120 reported 483 acts that met legal definitions of rape. The distribution of these incidents is illustrated in Figure 1.1. Lisak and Miller also state in their introduction that around 6% to 14.9% (about 1/16 to 1/7) of men on college campuses report acts that meet the definition of rape or attempted rape.[4]

For the most part, the RAINN statistics were used as a basis for the models and equations, and the Lisak and Miller study was the object on which these models and equations were applied to obtain the results. The Lisak and Miller plays a larger role in the creation of some of the upcoming equations, but it is used more to supplement the RAINN statistics. Another statistic that is important in building the models and equations for this

---

[3] "Reporting Rates." Rape, Abuse and Incest National Network. Accessed December 28, 2015. https://rainn.org/get-information/statistics/reporting-rates.

[4] Lisak, David, and Paul Miller. "Repeat Rape and Multiple Offending Among Undetected Rapists." Violence and Victims 17, no. 1 (2002): 73-84. http://www.davidlisak.com/wp-content/uploads/pdf/RepeatRapeinUndetectedRapists.pdf.

study is the frequency of false accusations. According to the FBI, 8% of forcible rape complaints were unfounded meaning they were found to be either false or baseless. Though this does not necessitate that these accusations were false, this value will be used for the models and equations because it provides a high estimate.[5]

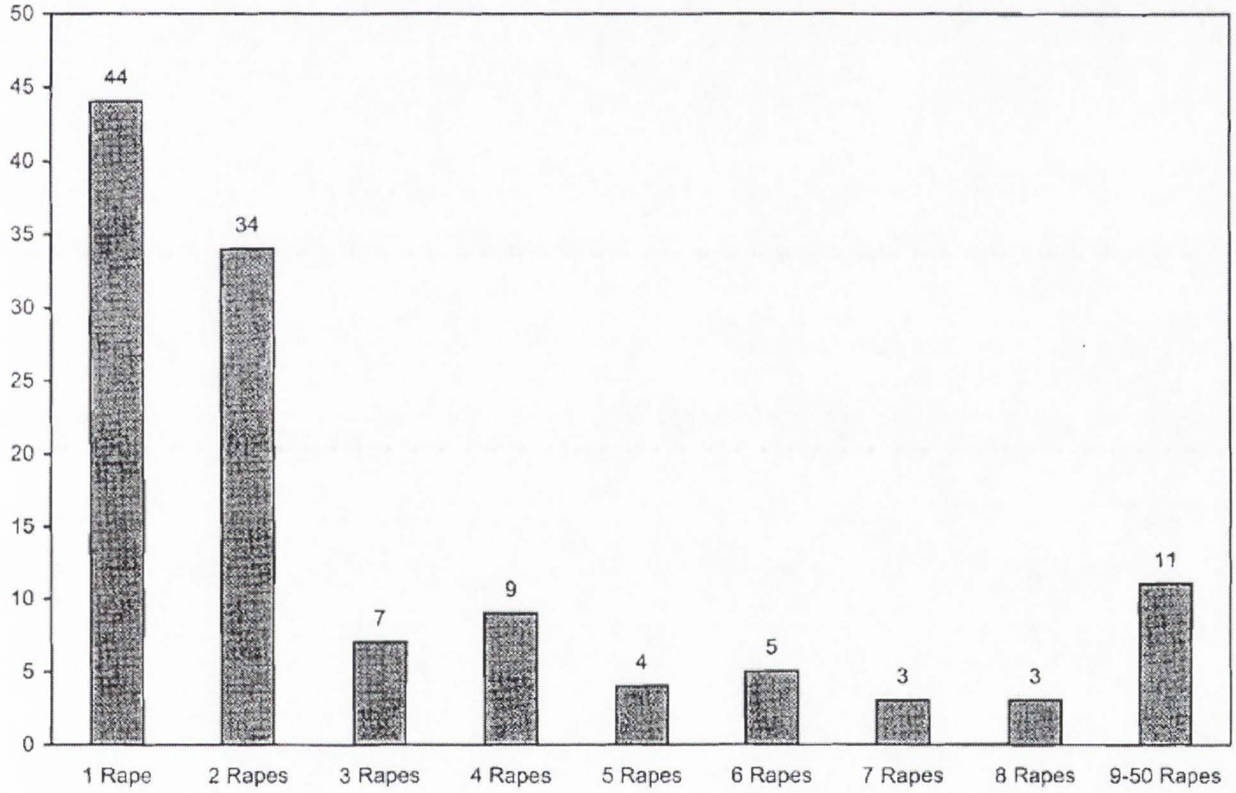

Figure 1.1: Number of rapists who committed single and multiple numbers of rape

Section 2: Mathematical Concepts and Models

Various concepts taken from Bayesian statistics have been implemented into this research. The most important of these concepts were *Bayes' Theorem* and *Bayesian networks*. Bayes' Theorem is as follows:

---

[5] "Crime Index Offenses Reported." The Federal Bureau of Investigation. 1996. Accessed December 28, 2015. https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/1996/96sec2.pdf.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where $P(A|B)$ is the probability that A is true given B, $P(B|A)$ is the probability that B is true given A, $P(A)$ is the probability of A happening regardless of B, and $P(B)$ is the probability of B happening regardless of A. This equation lays out the three main components required to be able to find the probability that an innocent person is prosecuted, or the probability that a person is prosecuted given that they are innocent, written $P(Pr|I)$. If we define Pr as the case where a person is prosecuted and I as the case where a person is innocent, then Bayes' Theorem can be applied and written as follows:

$$P(Pr|I) = \frac{P(I|Pr)P(Pr)}{P(I)}$$

The main objective is to determine the value of $P(Pr|I)$, so we must first find $P(I|Pr)$, $P(Pr)$, and $P(I)$. This gives the research a clear direction to follow to reach the desired conclusion.

A Bayesian network is a directed acyclic graph that represents a set of random variables and their conditional dependencies. The RAINN statistics and the FBI statistic concerning false reports were instrumental in developing the two main models used as a basis of every other network in this study. The first model, as shown in Figure 2.1, follows the process of what probabilities are involved when tracing cases of rape through the reporting and prosecution process. Each labeled box in Figure 2.1 shows a different stage of this process. The solid arrows show in which direction the process flows. The percentages to the right or above the solid arrows show what percent of the previous box's

contents move along the arrows' paths. The dotted arrow with the two dotted lines on either side of it shows that the "False accusations" box is contained in and makes up 8% of the "Reported to police" box.

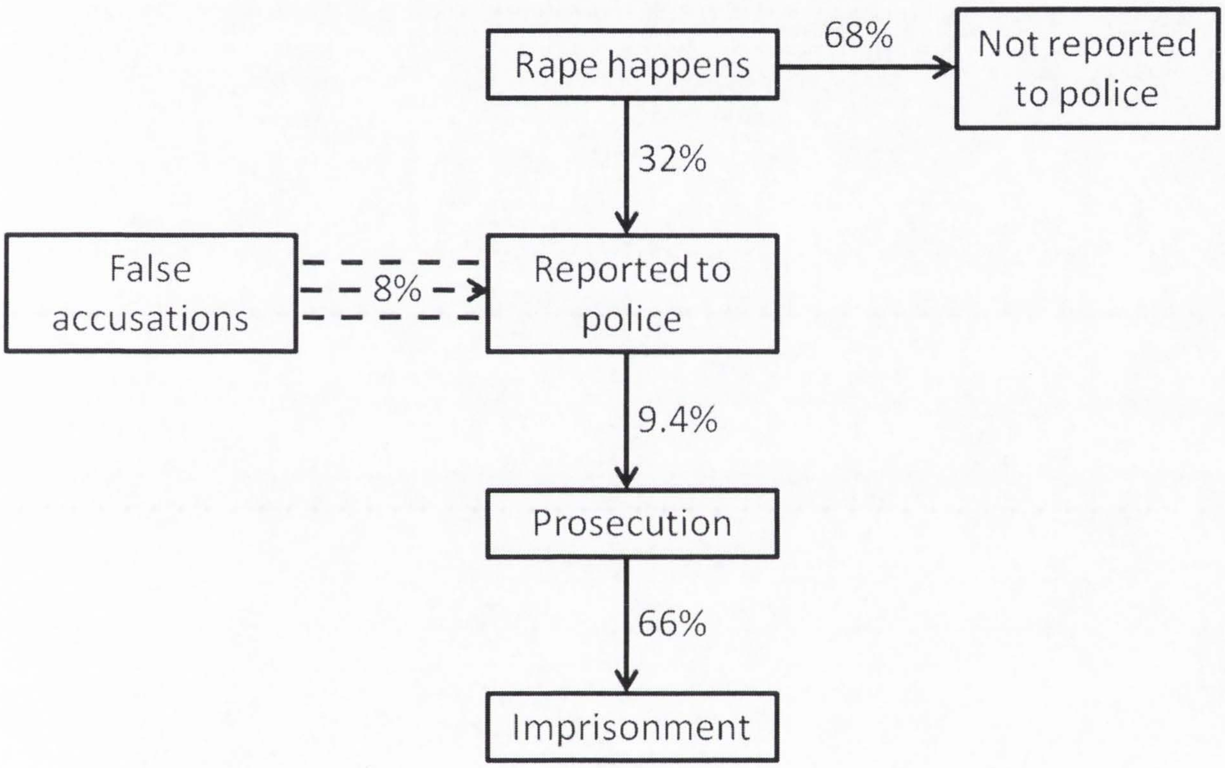

Figure 2.1: Model showing the probabilities surrounding the prosecution and imprisonment of rapes

The model in Figure 2.2 is similar to that represented in Figure 2.1, but it is more streamlined in that it cuts out the step concerning how many rapes are reported and jumps straight to how many are prosecuted. It also shows how many do not get imprisoned. This model is set up in order to be continuous in nature so it can track the probability one rapist has of being prosecuted or imprisoned after any number of rapes. Hence, the model is broken up into periods that start with the "Rape Happens" box and end with the "Freedom" box. The arrows, percentages, and labeled boxes each have the same general meaning as in

Figure 2.1. The dotted lined arrows connect the end of one period to the start of the next. These periods are incident dependent rather than time dependent, meaning for any given person, the period will begin if and only if they commit the crime.

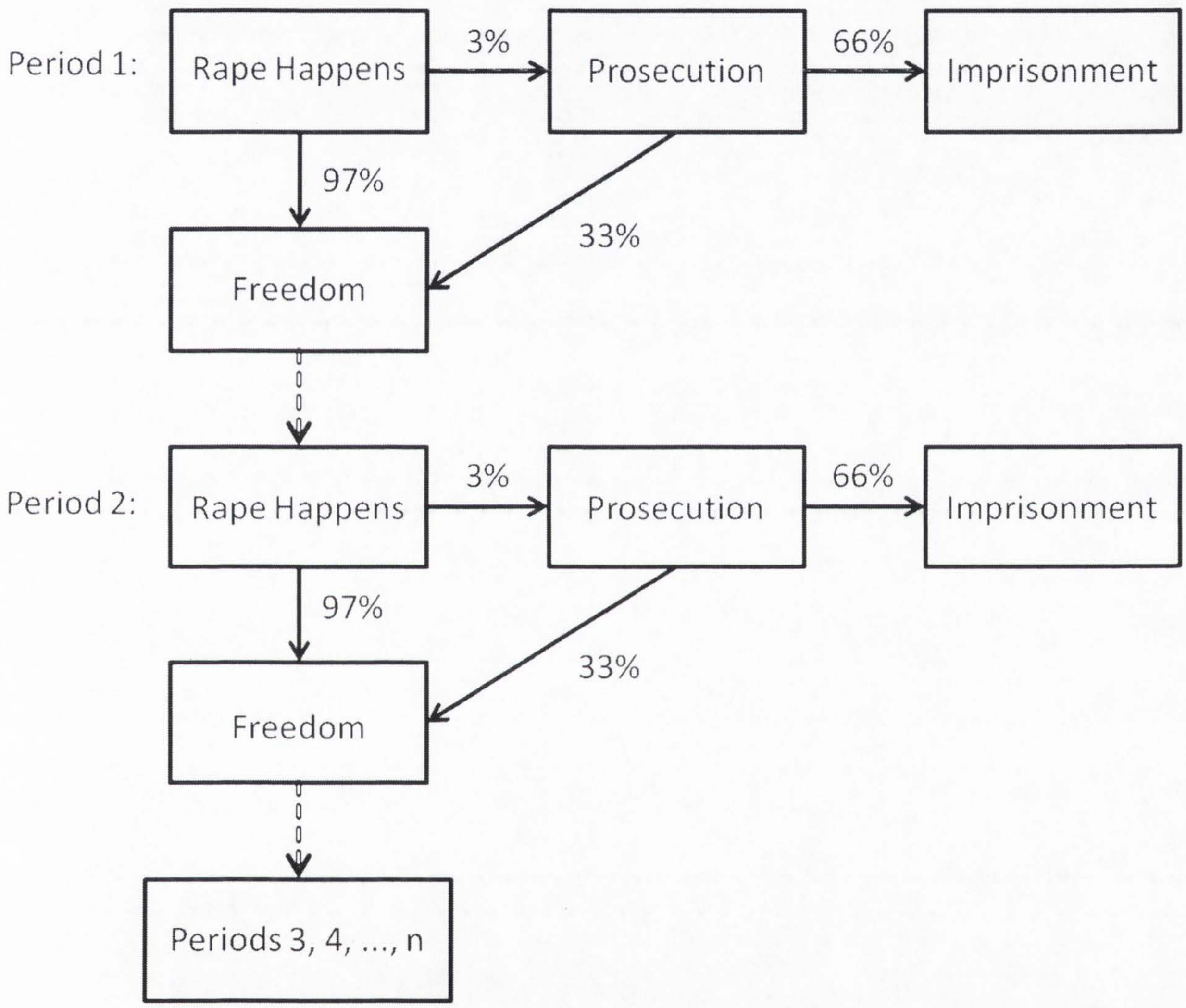

Figure 2.2: Model showing the probabilities surrounding the prosecution and imprisonment of multiple rapes committed by one person

In order to proceed from these models and equations, it was necessary to formulate several assumptions about factors for which there was no data. Most of these assumptions were made in such a way as to maximize $P(Pr|I)$. This is because if we find the maximum value to be sufficiently low, then it follows that all rape cases fulfill at least the fourth

requirement of the doctrine of chances. The first of these assumptions is that the RAINN statistics apply uniformly to both the innocent and guilty people against whom an accusation has been made. It is likely that there is less evidence available to convict an innocent person, so this should provide a reliable maximum value. This assumption also allows us to group all accusations together and perform calculations all at once.

The second assumption is that an innocent person is generally falsely accused of one rape. This helps maximize $P(Pr|I)$ by creating a one-to-one link between false accusations that are prosecuted and the innocent who are being prosecuted, thus increasing the innocent to guilty ratio among people who are prosecuted. A third assumption is that there is only one rapist per offense. There do exist cases where multiple people are involved in the offense, but by omitting these incidents, we increase the ratio of rapes to rapists, which thus increases the innocent to guilty ratio among people who are prosecuted. Finally, it is assumed that once a person is imprisoned for rape, they do not rape again. The reasoning for this assumption is that the average prison time for rape is 8-9 years, which is significantly longer than the duration of the studies being used in this research. This being the case, it will likely yield more accurate results if these cases are omitted.

Section 3: Research Report

As mentioned in section 2, Bayes' Theorem provides an outline of steps to follow to determine the probability that an innocent person is prosecuted. The three steps are to determine how many people that are prosecuted for rape are innocent, find the probability that a person is prosecuted for rape, and to find the probability that a person is innocent.

These three steps determine $P(I|Pr)$, $P(Pr)$, and $P(I)$ respectively, which make up all the values required to implement Bayes' Theorem.

3.1 Probability of Being Innocent

We will begin by finding the value of $P(I)$, because it is the easiest to find. Since about 1/16 to 1/7 of men were found to have executed acts that were legally defined as rape, then 6/7 to 15/16 of men did not. In other words, 6/7 to 15/16 of men are innocent. Since $P(I)$ is in the denominator of the fraction used in Bayes' Theorem, then in order to maximize our value for $P(Pr|I)$, we must use the minimum value of $P(I)$ in the above range which is 6/7. Therefore,

$$P(I) = \frac{6}{7} \qquad \blacksquare$$

3.2 Probability of Being Innocent When Being Prosecuted

3.2.1 Prosecution of those falsely accused

Next, we will determine the probability that a prosecuted person is innocent, $P(I|Pr)$. This is a complicated task being that the statistics concerning prosecution deal with instances of rape as opposed to individual rapists. In the Lisak and Miller study, about 63% of those surveyed had committed multiple offenses, with an average of 5.78 rapes per repeat offender (the average is 4.03 if one-time offenders are included). Subjecting every instance of rape to the same probabilities, it follows that the percentage of rapists that are prosecuted at some point is higher than that of prosecuted rapes.

We will begin by applying the model in Figure 2.1 to the rapists from the Lisak and Miller study. Out of the 483 rapes, 154.56 get reported. Since we are assuming 8% of all reports to be false, then the total reports of rape (both true and false reports) is given by

$$0.92x = 154.56 \rightarrow x = 168$$

where x is the total number of reports. Hence, we have 154.56 reported rapes and 13.44 false accusations. Assuming that all reports are equally likely to be prosecuted, then 14.49 rapes and 1.26 false accusations are prosecuted. From the assumptions laid out in section 2, we now have 1.26 people that are falsely accused. Now we must determine the number of rapists rather than rapes that get prosecuted. This is done by adjusting the model in Figure 2.2 in order to keep track of every individual rapist as shown in Figure 3.1.

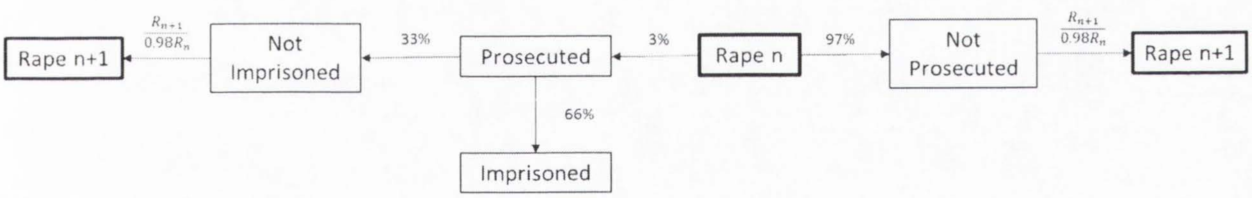

Figure 3.1: Adjusted model from Figure 2.2

In this model and for the remainder of this research, R_n denotes the number of rapists that were involved in at least n rapes. As seen in Figure 1.1, $R_1 = 120$, $R_2 = 76, R_3 = 42$, and so forth. If we expanded this model to include Rape $n + 2$, then we would expand from both Rape $n + 1$ boxes in a similar manner as with the Rape n box. By further expansion, we are able to find probabilities for how likely it is for a rapist to be prosecuted any number of times for any amount of rapes.

3.2.2 Finding the Total Number of Prosecuted Rapists

Let ${P_n}^*$ denote the expected number of people prosecuted after n rapes that were not prosecuted for $n - 1$ or fewer rapes, and consider R_n as previously defined. When $n = 1$, $R_1 = 120$, and the probability of prosecution is 3% since there is a one to one ratio of rapes to rapists at this point. Therefore, ${P_1}^* = 3.6$. Since 66% of those prosecuted are imprisoned, then the total number of rapists able to be involved in a second offense is

$$120 - (0.66)(0.03)(120) = 120 - (0.02)(120) = (0.98)(120) = 117.6.$$

From Figure 1.1, we know that $R_2 = 76$, meaning that 76 out of the remaining 117.6 end up raping again.

To find ${P_2}^*$, we are not interested in following those who have already been prosecuted, because they have already been prosecuted at least once. Since there is a 97% chance that a rapist is not prosecuted after the first rape, and since 76 out of 117.6 non-imprisoned rapists are involved in a second offense, then those involved in a second offense that have not yet been prosecuted is given by

$$(0.97R_1)\left(\frac{76}{117.6}\right) = 0.97(120)\left(\frac{76}{(120)(0.98)}\right) = \left(\frac{97}{98}\right)(76) = \left(\frac{97}{98}\right)R_2.$$

At this point, since 3% of rapes end up being prosecuted, then we multiply the above by 0.03 which gives us

$${P_2}^* = 0.03\left(\frac{97}{98}\right)R_2.$$

It is useful to calculate P_3^* in order to more fully illustrate the pattern present in this model. Using similar reasoning as for P_2^*, we get

$$P_3^* = (120)(0.97)\left(\frac{76}{(120)(0.98)}\right)(0.97)\left(\frac{42}{(76)(0.98)}\right)(0.03) = \left(\frac{97}{98}\right)^2 (42)(0.03)$$

$$= 0.03\left(\frac{97}{98}\right)^2 R_3.$$

At this point, it may be useful to see this process as a simpler cyclical model as shown in Figure 3.2. To find P_n^* for any n, we calculate $n-1$ complete cycles between the boxes labeled "Rape" and "Not Prosecuted" before following the path to the "Prosecuted" box. By following this model, we can find a formula for P_n^* for any n as follows

$$P_n^* = R_1(0.97)\left(\frac{R_2}{0.98R_1}\right)(0.97)\left(\frac{R_3}{0.98R_2}\right)\ldots(0.97)\left(\frac{R_n}{0.98R_{n-1}}\right)(0.03).$$

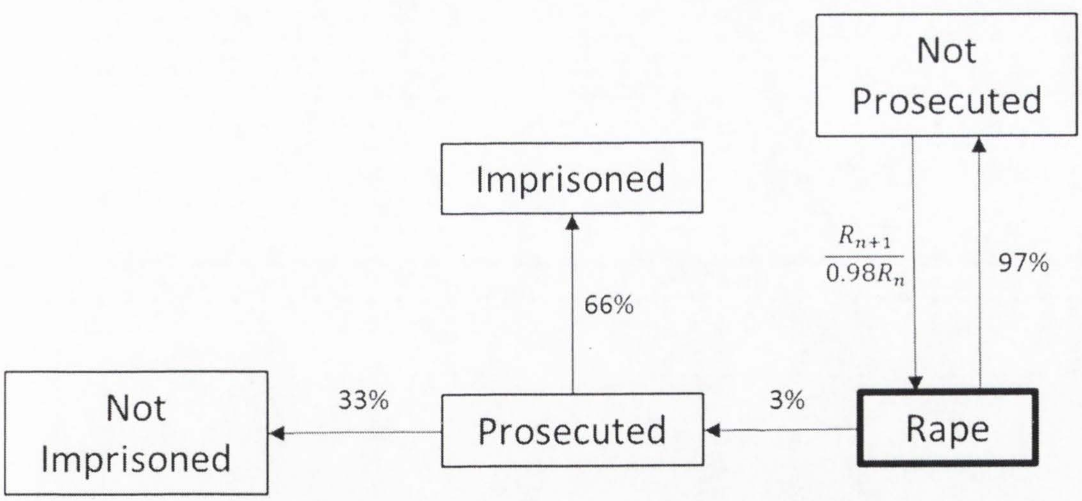

Figure 3.2: Cyclical model of Figure 3.1

There is a one-to-one correspondence between the number of times we multiply by R_k, for $k > 1$, to the number of times we multiply by $\frac{0.97}{0.98R_{k-1}}$ in the above formula. In

addition, for every $k < n$, we are both multiplying and dividing by R_k, so they cancel each other out. Therefore, the above formula can be simplified to

$$(1) \qquad P_n^{\ *} = 0.03\left(\frac{97}{98}\right)^{n-1} R_n.$$

Since this equation finds how many rapists are prosecuted after n rapes that were not prosecuted for their $n-1$ previous rapes, then by adding together $P_n^{\ *}$ for all n, we will find the total expected number of rapists that were prosecuted. Since our data only shows rapists that have been involved in 50 or less incidents, then we get the following equation where T stands for the total number of prosecuted rapists.

$$(2) \qquad T = \sum_{n=1}^{50} P_n^{\ *}$$

These two equations work as long as we have a value for R_n. However, as seen in Figure 1.1, we don't have precise values for R_n for $n > 9$. In order to approximate these values, we will let x be the average percentage of rapists who move on to commit $n+1$ rapes after their nth rape for $n > 9$. R_n for $n > 9$ can thus be estimated by

$$R_n \approx 11x^{n-9}.$$

By subtracting the number of offenses committed by those with 8 or less rapes from the total number of rapes, we see that the 11 rapists with 9 or more rapes committed 120 offenses. Also, after $n = 8$, there is at least one rapist that commits 41 additional offenses. With this information, we can say that

$$120 = \sum_{n=10}^{50} 11x^{n-9}.$$

We can find the value of x by applying the geometric sum as follows

$$120 = \sum_{n=10}^{50} 11x^{n-9} = \sum_{n=1}^{41} 11x^{n} = -11 + \sum_{n=0}^{41} 11x^{n} = -11 + 11\left(\frac{1-x^{42}}{1-x}\right) \to 120$$
$$= -11 + 11\left(\frac{1-x^{42}}{1-x}\right) \to \frac{131}{11} = \left(\frac{1-x^{42}}{1-x}\right) \to \frac{120}{11} - \frac{131}{11}x + x^{42} = 0 \to x$$
$$\approx 0.9184.$$

Hence, our estimate for R_n for $n > 9$ is

$$(3) \qquad R_n = 11(0.9184)^{n-9}$$

and thus our estimate for T is given by

$$T = \sum_{n=1}^{50} 0.03\left(\frac{97}{98}\right)^{n-1} R_n = \sum_{n=1}^{9} 0.03\left(\frac{97}{98}\right)^{n-1} R_n + \sum_{n=10}^{50} 0.03\left(\frac{97}{98}\right)^{n-1} (11(0.9184^{n-9}))$$
$$\approx 13.64.$$

Therefore, out of the given 120 rapists it is expected that 13.64 (11.36%) are prosecuted at least once. ■

We previously calculated that out of 483 rapes, the expected value for the total number of prosecutions is 15.75 where 14.49 were rapes and 1.26 were false accusations. We have now calculated that out of 120 rapists, 13.64 are expected to be prosecuted. Using these two numbers, we see a total value of 14.9 for people being prosecuted where about

91.54% are guilty and 8.46% are innocent. This latter percent is the probability that a person is innocent given that they are prosecuted, so $P(I|Pr) = 8.46\%$. ■

3.3 Probability of Being Prosecuted

The final step in this process is to find the probability that a person is prosecuted for rape, $P(Pr)$. Since a person has a 1/7 probability of being a rapist,[6] then

$$P(Pr) = \frac{6}{7}P(\mathrm{Pr}|I) + \frac{1}{7}P(\mathrm{Pr}|Not\ I).$$

Since $P(Pr|I)$ is the probability we are ultimately working towards and since it appears in the above equation, we will be unable to find a concrete value for $P(Pr)$ until we come to the final conclusion. If we are able to determine a value for $P(Pr|Not\ I)$, then we will be able to create a single variable equation from which we will be able to solve for $P(Pr|I)$.

In section 3.2.2, we found that 13.64 of the 120 rapists in question are expected to be prosecuted at least once. This gives us a probability of 11.37% that a rapist is prosecuted at least once. For the purposes of this research, this is the probability we will use for $P(Pr|Not\ I)$.[7] By replacing this value into the formula above, we get the following

$$(4) \qquad P(Pr) = \frac{6}{7}P(Pr|I) + \frac{1}{7}(0.1137).$$ ■

[6] Lisak, David, and Paul Miller. "Repeat Rape and Multiple Offending Among Undetected Rapists." Violence and Victims 17, no. 1 (2002): 73-84. Accessed December 28, 2015. http://www.davidlisak.com/wp-content/uploads/pdf/RepeatRapeinUndetectedRapists.pdf.

[7] See section 4 for alternative methods of finding values for $P(Pr|Not\ I)$.

3.4 Probability of Being Innocently Prosecuted

At this point, we now have all the information we need to use Bayes' Theorem to calculate $P(Pr|I)$. As a reminder, $P(I|Pr)$ is the probability of a person being innocent given that they are prosecuted, $P(Pr)$ is the probability of a person being prosecuted whether or not they are innocent, $P(Pr|I)$ is the probability of a person being prosecuted given that they are innocent, and $P(I)$ is the probability that a person is innocent. Following are all the pieces as they were found in the previous sections:

$$P(I|Pr) = 8.46\%$$

$$P(Pr) = \frac{6}{7}P(Pr|I) + \frac{1}{7}(0.1137)$$

$$P(I) = \frac{6}{7}.$$

Now, the final step will be to compile these values together into Bayes' Theorem and solve for $P(Pr|I)$:

$$P(Pr|I) = \frac{(0.0846)\left(\frac{6}{7}P(Pr|I) + \frac{1}{7}(0.1137)\right)}{6/7} \rightarrow P(Pr|I) = (0.0846)P(Pr|I) + \frac{1}{6}(0.1137)$$

$$\rightarrow (0.9154)P(Pr|I) = \frac{1}{6}(0.1137) \rightarrow P(Pr|I) = 0.00175133.$$

Therefore, there is a 0.175% probability that an innocent person is prosecuted. From here, we can further deduce the probability that an innocent person is prosecuted twice, which is most applicable in determining whether or not the DOC can be applied. In the introduction, the third criterion required for the DOC to take effect was for the two

incidents to be completely independent from one another. This being the case, and assuming that being innocently prosecuted for rape does not significantly affect the probability of being prosecuted a second time, then we can find the probability that an innocent person is prosecuted twice in the following manner

$$P(Pr = 2|I) = (P(Pr|I))^2 = 0.00175133^2 = 0.000003.$$

The probability that an innocent person is prosecuted twice for rape is then 0.0003%, which seems sufficiently low to support the claim that being prosecuted a second time for rape automatically meets the fourth criterion to allow the use of the DOC. The first three criteria still must be met before the DOC can be applied, but the fourth criterion is met almost by default based on the current statistics we have concerning rape and prosecution.

3.5 The Effects of Changing Statistics

There was recently a study completed by Julie Valentine, a BYU nursing professor, in which she examined 270 rape cases in Salt Lake County that transpired between 2003 and 2011.[8] This study focused mainly on the backlog of sexual assault kits collected by law enforcement when dealing with these cases. She also found that 6% of these cases led to prosecution. There are many reasons this number could be different than the 3% found by RAINN including but not limited to different sample sets, different sample sizes, or different

---

[8] McBride, Jon. "BYU Professor Works to Help Victims of Rape through In-depth Research and Training." BYU News. April 06, 2016. Accessed April 11, 2016. https://news.byu.edu/news/byu-professor-works-help-victims-rape-through-depth-research-and-training.

ways of collecting data. It does call to question how such changes could affect the results we've found in this section.

To answer this question, we must consider what will happen for varying percentages of prosecution other than 3%. If we reexamine the equations and models in this section replacing all references of this 3% prosecution rate with a variable x where $0 \leq x \leq 1$, then we will be able to determine how this could potentially affect the application of our results. The following equations (5) and (6) and Figure 3.3 show these adjustments.

$$(5) \qquad P_n^*(x) = x\left(\frac{1-x}{1-0.33x}\right)^{n-1} R_n$$

$$(6) \qquad T(x) = \sum_{n=1}^{50} P_n^*(x).$$

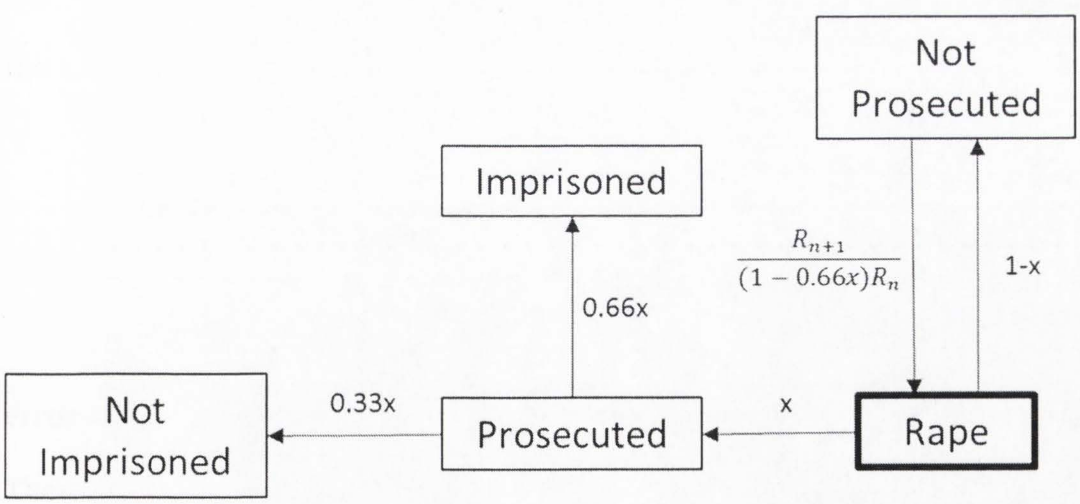

Figure 3.3: Cyclical model of Figure 3.1 with a variable prosecution rate.

If we plug in our calculated value for R_n into the above equations and follow the same processes outlined in this section, then we will be able to find the desired probabilities for all x. Being that values of R_n don't follow an obvious pattern or formula, it makes it difficult to graph the above equations to see how different values of x affect the end results. For this reason, we will create an estimate for R_n for all n, much like how we did before for $n > 9$ using the geometric sum.

$$483 = \sum_{n=0}^{49} 120y^n \rightarrow 483 = 120\left(\frac{1-y^{50}}{1-y}\right) \rightarrow \frac{483}{120} = \left(\frac{1-y^{50}}{1-y}\right) \rightarrow \frac{363}{120} - \frac{483}{120}y + y^{50} = 0$$

$$\rightarrow y \approx 0.751553.$$

Using our value for y, our estimate for R_n, hereafter given by R'_n, is

$$(7) \qquad R'_n = 120(0.751553)^{n-1}.$$

Now we will test the accuracy of R'_n by plugging it in to $T(0.03)$ and comparing it to the value of T found at the end of section 3.2.2.

$$T(0.03) = \sum_{n=1}^{50} P_n{}^*(0.03) = \sum_{n=1}^{50} 0.03\left(\frac{97}{98}\right)^{n-1}(120)(0.751553)^{n-1} \approx 14.06.$$

Our estimate is 14.06 total prosecuted rapists compared to 13.64 which gives us an error of 0.42 for 50 rapes, an accurate high estimate for the probability threshold. Therefore, we will use this estimate for R'_n along with equations (5) and (6) to develop a function $P(Pr|I)(x)$ from which we can calculate $P(Pr|I)$ for any x. Let $I(x)$ be the function of how many people are innocently prosecuted for an overall prosecution rate of x. If $x = 0.03$, then $I(0.03) = 1.26$ as found section 3.2.1. We will now find $I(x)$ for any x.

Since 3% of all rapes are prosecuted and 32% are reported, then the prosecution rate of the people reported is $\left(\frac{0.03}{0.32}\right)$. Also, since 8% of all reported rapes are considered false, then the total number of individuals that are reported for rape is $\frac{(0.32)(483)}{1-0.08} = 168$. By subtracting $(0.32)(483)$ from 168, we get 13.44 which is the expected number of people falsely accused of rape. By putting this all together, we can find $I(0.03)$ to be the following

$$I(0.03) = \left(\frac{0.03}{0.32}\right)\left(\frac{(0.32)(483)}{1-0.08} - (0.32)(483)\right) = \left(\frac{0.03}{0.32}\right)(0.32)(483)\left(\frac{1}{0.92} - 1\right)$$

$$= (0.03)(483)\left(\frac{0.08}{0.92}\right) = 0.03(42).$$

Since 0.03 isn't a factor in any value other than the 0.03 included above, then

$$I(x) = 42x.$$

This means that the equation for $P(I|Pr)$, given by $P(I|Pr)(x)$ is

$$(8) \qquad P(I|\mathrm{Pr})(x) = \frac{I(x)}{T(x) + I(x)}.$$

Recall from section 3.3, $P(Pr|Not\ I)$ was calculated by taking the ratio on $\frac{T(0.03)}{120}$. Also, the value for $P(I)$ is not dependent on the rate of prosecution, so it will remain at 6/7. Now, all that is left is to put this all together and solve for the expression $P(\mathrm{Pr}|I)(x)$ which is the probability that an innocent person is prosecuted with a prosecution rate of x.

$$P(Pr|I)(x) = \frac{[P(I|Pr)(x)][P(Pr)(x)]}{6/_7} = \frac{[P(I|Pr)(x)]\left[\frac{6}{7}P(\mathrm{Pr}|I)(x) + \frac{1}{7}P(Pr|Not\ I)(x)\right]}{6/_7}$$

$$= \frac{\left(\frac{I(x)}{T(x)+I(x)}\right)\left[\frac{6}{7}P(Pr|I)(x) + \frac{1}{7}\left(\frac{T(x)}{120}\right)\right]}{6/_7}$$

$$= \left(\frac{I(x)}{T(x)+I(x)}\right)P(Pr|I)(x) + \frac{1}{720}\left(\frac{I(x)T(x)}{T(x)+I(x)}\right).$$

Solving for $P(Pr|I)$, we then get

$$P(Pr|I)(x) = \left(\frac{I(x)}{T(x)+I(x)}\right)P(Pr|I)(x) + \frac{1}{720}\left(\frac{I(x)T(x)}{T(x)+I(x)}\right)$$

$$\to \left(1 - \frac{I(x)}{T(x)+I(x)}\right)P(Pr|I)(x) = \frac{1}{720}\left(\frac{I(x)T(x)}{T(x)+I(x)}\right)$$

$$\to \left(\frac{T(x)}{T(x)+I(x)}\right)P(Pr|I)(x) = \frac{1}{720}\left(\frac{I(x)T(x)}{T(x)+I(x)}\right) \to P(Pr|I)(x)$$

$$= \frac{1}{720}\left(\frac{I(x)T(x)}{T(x)+I(x)}\right)\left(\frac{T(x)+I(x)}{T(x)}\right) \to P(Pr|I)(x) = \frac{I(x)}{720}.$$

When we plug in 0.03 for x, then we get

$$P(Pr|I)(0.03) = \frac{I(0.03)}{720} = \frac{42(0.03)}{720} = 0.00175$$

which is almost exactly what we got for $P(Pr|I)$ in section 3.4.1. Therefore, our equation for $P(Pr|I)(x)$ *is*

$$(9) \qquad P(Pr|I)(x) = \frac{I(x)}{720} = \frac{7x}{120} \ \textit{where}\ 0 \le x \le 1.$$

Thus we see that the range of $P(Pr|I)(x)$ is [0%, 5.8%]. For the 6% report rate found by Julie Valentine, we see that $P(Pr|I)(0.06) = 0.35\%$, which is still a very small probability. This equation should be a good estimate for any small changes in x because it does not take into consideration any other factors that could have been altered by the change in the prosecution rate. For example, it is possible that if the prosecution rate rose too much, people would be more likely to make false accusations. It could also be the case that as prosecution rates rise, the number of rapists also declines. Many other factors go into this that are not considered in this calculation, but for minor changes in x that have little to no effect on these outside factors, this equation is an estimate.

Another value for which it is interesting to explore changes is the rate at which rapes are reported. As seen in Figure 1, 32% of rapes are reported and about 9.4% (exactly 9.375%) of those that are reported end up being prosecuted. Recall that equation (5) estimates the number of people prosecuted after n rapes that were not prosecuted for $n-1$ rapes, and equation (6) estimates the total number of prosecuted rapists. If the prosecution rate remains constant with respect to the number of rapes that get reported, then we can determine how many people could be prosecuted by adjusting equation (5) as follows

$$P_n^*(z) = 0.09375z\left(\frac{1-0.09375z}{1-0.33(0.09375z)}\right)^{n-1} R_n,$$

where z is the reporting rate. We finish by applying this change to equation (6)

$$T(z) = \sum_{n=1}^{50} 0.09375z\left(\frac{1-0.09375z}{1-0.33(0.09375z)}\right)^{n-1} R_n.$$

From this point, if we divide by 120, we can find the prosecution rate for any reporting rate by inputting any number from 0 to 1 for z. Using the same estimate for R_n as was used previously in this section, we obtain the graph in Figure 3.4. The graph being depicted on the following page is exactly what is described above and can be written as the following equation

$$\frac{T(z)}{120} = \sum_{n=1}^{50} 0.09375z\left(\frac{1 - 0.09375z}{1 - 0.33(0.09375z)}\right)^{n-1}(0.751553)^{n-1}.$$

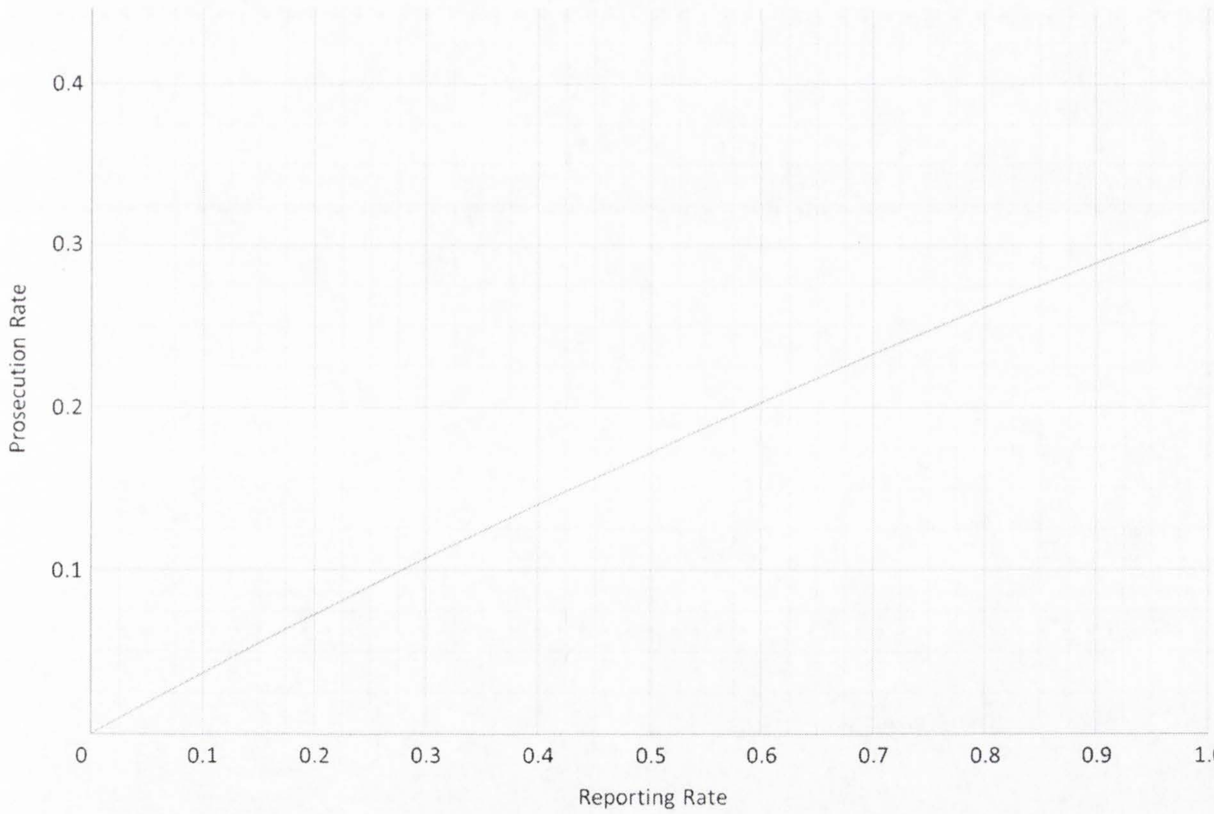

Figure 3.4: Graph showing the prosecution rate for different reporting rates.

One of the strongest implications from exploring various reporting rates is the effect it has on the ratio of false reports to legitimate ones. If the number of reports remains the

same and the number of rapes that are reported increases, then there will be a smaller ratio of false reports to legitimate claims. For example, if the reporting rate increases to 50%, then the expected number of rapists prosecuted increases to about 20.65 whereas the number of false accusations that proceed to prosecution remains at 1.26. Hence the probability of a person being innocent given that they are being prosecuted decreases from 8.46% to 5.75%.

Section 4: Alternative Methods

The methods used throughout section 3 are far from being the only ways to go about these calculations. Attempts were made to choose the most accurate methods of calculation throughout the entire process, though in some instances it was difficult to discern which method would yield the most accurate results. In such instances, the method that yielded a higher end value for $P(Pr|I)$ was chosen in order to produce a safe maximum in order to find out whether or not the DOC could be implemented. This research could potentially be applied to a lot more than just the DOC, so in order to make this research as complete as possible, the alternative methods not included in section 3 will be included below.

4.1 Dividing Prosecutions by Average Offenses per Rapist

The first alternative method we will examine explores another way to find $P(I|Pr)$. This method follows the model in Figure 2.1 exactly as was done in section 3, but the final value was calculated in a different way. After determining that 14.49 out of 15.75

prosecuted cases were not false accusations, 14.49 was divided by the average number of rapes per rapist. This gave us 1.26 innocent people to 3.6 guilty people being prosecuted which resulted in $P(I|Pr)$ being 26%. This value is much greater than the one that was actually used in section 3, but it was not used because the method left too many factors out of consideration. The biggest weakness to this method is it doesn't consider the many different probabilities of being prosecuted and not being prosecuted for each rape up to n for each individual. Therefore, even though it would increase our value for $P(Pr|I)$, it was not accurate enough to be used.

4.2 Continuous Probability Calculations

This method followed the model in Figure 2.2 in order to determine the probability that a single offender is prosecuted at least once after n rapes. This probability is denoted by P_n, not to be confused with P_n^* as defined in section 3.2.2. By following Figure 2.2, once a person was prosecuted, we no longer followed that person because they had been prosecuted at least once. So $P_1 = 0.03, P_2 = 0.03 + 0.97(0.03), P_3 = 0.03 + 0.97(0.03) + (0.97)^2(0.03)$, and so on. Since there is a 97% chance for a person not to be prosecuted, then when $n = 2$, the person had a 3% chance of being prosecuted at the first offense, and a 3% chance of being prosecuted for the first time at the second offense after having a 97% chance of not being prosecuted the first time. This continues indefinitely until the person no longer commits offenses. Since we add an additional $(0.97)^{n-1}(0.03)$ for each offense, then the value for P_n is

$$(10) \qquad P_n = \sum_{k=0}^{n-1} 0.03(0.97)^k.$$

This method was further applied to the Lisak and Miller Study in order to determine a value for $P(Pr|Not\ I)$. This was done by summing together the products of $P_n(R_n - R_{n+1})$ for each n from 1 to 50 and then dividing it all by 120. This is essentially taking the probability that a person is prosecuted at least once for n rapes and multiplying it by the number of people who committed exactly n rapes. Written out, this gives us

$$\sum_{n=1}^{50} \frac{P_n(R_n - R_{n+1})}{120}$$

which in turn gives us

$$P(Pr|Not\ I) = \sum_{n=1}^{50}\left[\left(\frac{(R_n - R_{n+1})}{120}\right)\left(\sum_{k=0}^{n-1} 0.03(0.97)^k\right)\right] \approx 0.1004.$$

When this method was used, it seemed relatively accurate. The main reason this method was not implemented in section 3 was because it yielded a lower value for $P(Pr|Not\ I)$ which in turn gave us a lower value for $P(Pr|I)$ in the end. Since we were seeking an upper bound and could not see an issue with the method used in section 3, the method that yielded the higher result was used.

Figure 2.2 could also be used to find values for ${P_n}^m$ which is the probability that a person who has committed n rapes is prosecuted m times. For example, we were able to determine from Figure 2.2 that ${P_n}^2$ was the following

$$(11) \qquad {P_n}^2 = \sum_{k=0}^{n-2} \binom{n}{k} (0.03)^{n-k}(0.97)^k.$$

In section 4.3, we will explore another alternative method that yields an equation that makes it much easier to calculate ${P_n}^m$ for any n and m.

4.3 The Grid Game

This final method was used to determine $P(Pr|Not\ I)$ by adding together the probability of a person being prosecuted exactly m times for n rapes for all m and n. This method used a very visual way of representing how these probabilities were calculated. Referring to Figure 4.1, this method starts us automatically in the dot located at the top right corner of the grid. For each offense that gets prosecuted, we move to the dot directly below our current location. For each offense that does not get prosecuted, we move to the dot directly to the left. For any given values n and m, this method will produce a grid with $M = m + 1$ rows and $N = n - m + 1$ columns such that we end up counting the number of shortest possible paths that go from the top right corner to the bottom left corner. The grid shown in Figure 4.1 shows a 7x11 grid where $m = 6$ and $n = 16$.

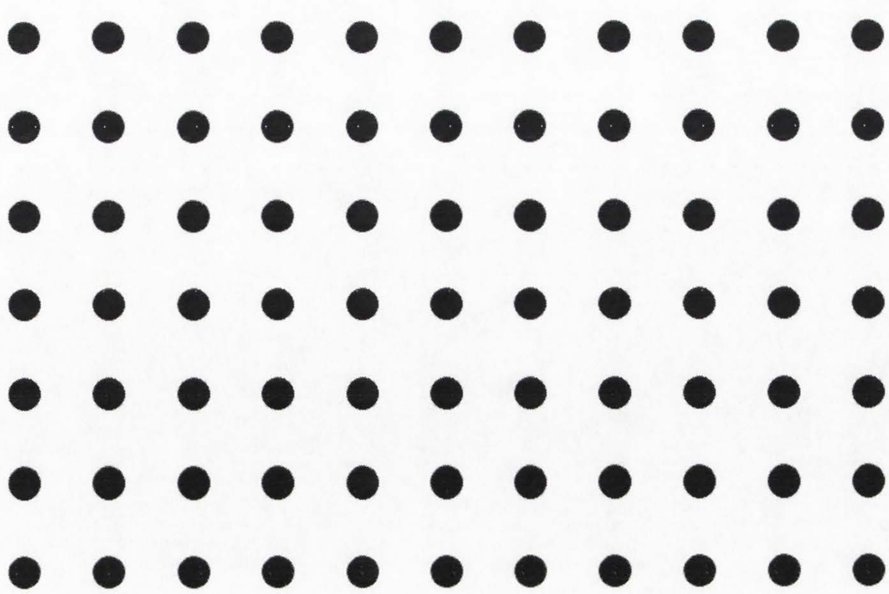

Figure 4.1: MxN grid where $M = 7$ and $N = 11$.

Figure 4.2 uses the 7x11 grid and shows one possible path that can be taken. The red arrows are cases where the rapist was prosecuted, and the green arrows are cases where they were not.

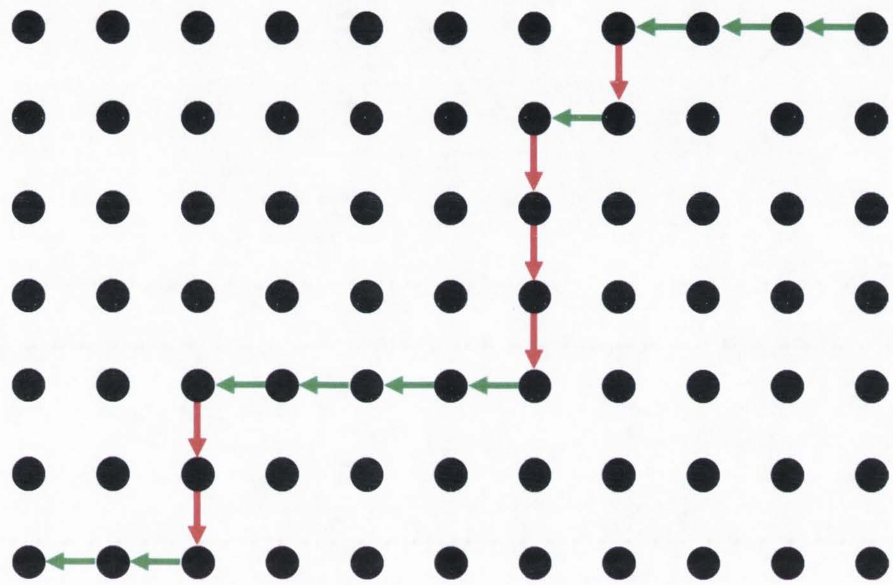

Figure 4.2: Example path on the MxN grid showing a person who committed 16 total rapes and was prosecuted for the 4th, 6th, 7th, 8th, 13th, and 14th offenses.

Using the above system, we can count all the possible combinations that a person who has committed n rapes can be prosecuted m times. By turning the grid diagonally, we can see that the number of paths we can take from the starting point to any other point follows a familiar pattern as shown in Figure 4.3.

This grid ends up becoming a recreation of Pascal's triangle which gives us the polynomial coefficients. Hence for any n and m, the number of paths to the corresponding point in the grid is $\binom{n}{m}$, or the binomial coefficient. By summing up all possible values of m for each n, we can find the total number of possible outcomes for a person who has committed n offenses. We can also find the probability of any one of these paths by

multiplying $\binom{n}{m}$ by 0.97 for every time we follow the green arrows and 0.03 for every time we follow the red ones. Since we follow the red arrows m times, then we follow the green arrows $n - m$ times, so our equation to find the probability of any path is

$$\binom{n}{m}(0.97)^{n-m}(0.03)^{m}.$$

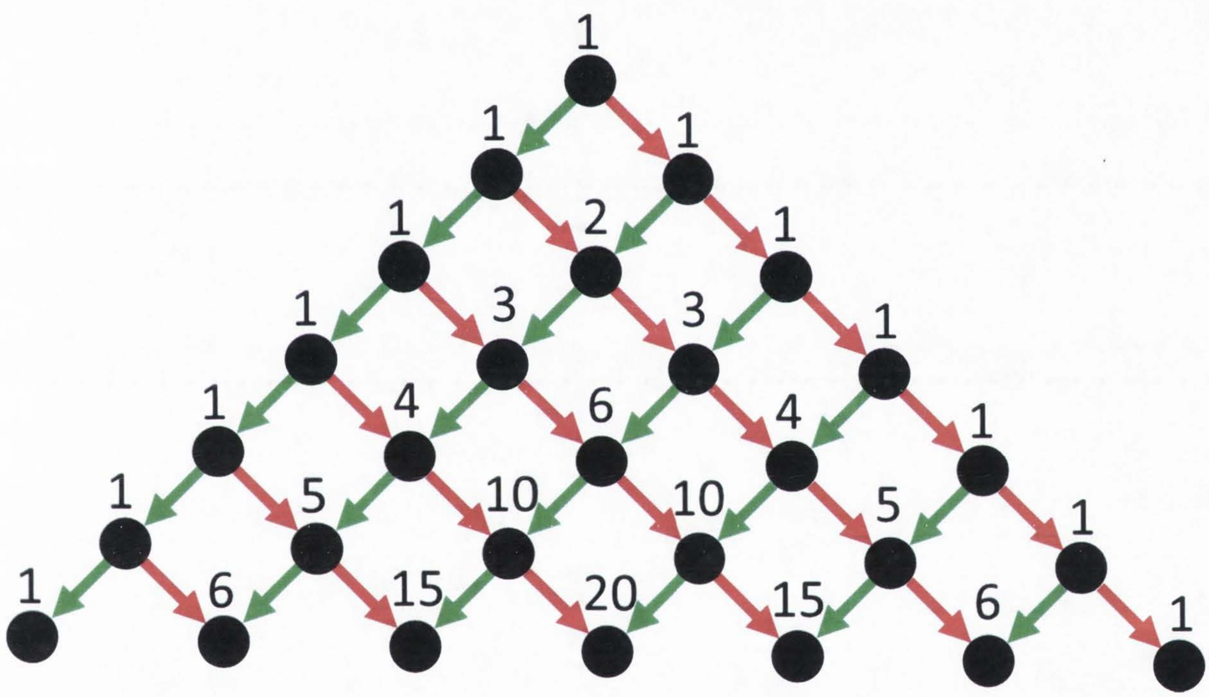

Figure 4.3: MxN grid turned 45 degrees counter-clockwise where each number represents the number of valid paths to that point.

By applying the binomial theorem to this, we find that

$$\sum_{m=0}^{\infty}\binom{n}{m}(0.97)^{n-m}(0.03)^{m} = (0.97 + 0.03)^{n} = 1.$$

This is somewhat intuitive since there is a 100% chance of somebody following one of the paths for committing n rapes after committing that many rapes. Also, since $\binom{n}{m} = 0$ for all

$m > n$, then the above equation does not consider any scenario where a person is prosecuted more than the number of times they commit an offense.

One might notice a resemblance between the above summation and equation 11 from section 4.2. By adjusting the m value in the above summation to only count the paths where somebody is prosecuted at least 2 times, we get

$$\sum_{m=2}^{\infty} \binom{n}{m} (0.97)^{n-m}(0.03)^{m}.$$

For ease of comparison, here is equation 11

$$P_n{}^2 = \sum_{k=0}^{n-2} \binom{n}{k} (0.03)^{n-k}(0.97)^{k}.$$

In equation 11, k represents the number of times that a person commits an offense and is not prosecuted whereas in the above equation, m represents the number of offenses that are prosecuted. Hence, $m = n - k$. We can therefore adjust $P_n{}^2$ as follows

$$P_n{}^2 = \sum_{k=0}^{n-2} \binom{n}{k} (0.03)^{n-k}(0.97)^{k} = \sum_{m=2}^{n} \binom{n}{n-m} (0.03)^{m}(0.97)^{n-m}$$

$$= \sum_{m=2}^{n} \binom{n}{m} (0.03)^{m}(0.97)^{n-m}.$$

Hence we see that these two summations are exactly the same. We can now give an expression for $P_n{}^m$ to be

$$(12) \qquad P_n{}^m = \sum_{q=m}^{n} \binom{n}{m} (0.03)^{q}(0.97)^{n-q}$$

where m represents the number of times a rapist is prosecuted, and n is the number of offenses the rapist commits.

Section 5: Conclusion

The main goal of this research was to determine the probability that a person was innocently prosecuted twice in order to satisfy the fourth criterion for use of the doctrine of chances in a court case dealing with rape. The calculations in section 3 found this probability to be 0.0003%, which we believe is sufficiently low to meet this criterion. This does not necessarily mean that the doctrine of chances should always be used in the court case, because there remain the three criteria of materiality, similarity, and independence. This research does however imply that unless there are major changes in the statistics concerning rape prosecutions and false accusations that this fourth criterion will most likely be met in a rape case.

Though this conclusion has obvious implications within the courtroom, much of what we have examined could be applied in many other ways as well. We saw one of these implications at the end of section 3 when we examined different reporting rates. If there is a way for more people to report the crime when it happens, then the ratio of false accusations to legitimate reports will decrease. Helping victims of rape to report the crime to law enforcement is not always one of the main focuses because there are so many traumatizing emotional experiences tied to each instance that it is often more important to support and comfort the victim rather than use the victim to prosecute the offender. Considering that most cases of rape are carried out by somebody who is known to the victim, the feelings of betrayal and hurt only exacerbate the trauma of an already

traumatizing experience. In addition, law enforcement officials are often used to interrogating suspects of crimes rather than victims, and many are unequipped to properly interview a victim of sexual assault without causing additional harm.[9]

With every hurdle society places in the way of victims reporting rapes, the more rapes will go unreported. The more unreported rapes, the more rapists walk free without punishment. The more unpunished rapists, the more people we put in danger of being raped. There are many potential areas of study that could help us find ways to create a safer world from rape and sexual abuse, but finding ways to remove these hurdles from the process of reporting rape is one additional area of study that can and should be further examined.

[9] "Reports and Studies." Sexual Assault & Anti-Violence Information. Accessed May 3, 2016. http://www.usu.edu/saavi/info/reports.cfm.

(Word Count: 1200)

Reflective Writing

Before I began working on this honors thesis, I looked at it with dread, but I now look back on my overall experience with fondness. Throughout the entire process, my feelings concerning the project seemed to be in a constant state of flux between frustration and excitement. This is by far the largest Mathematics project I've worked on to date, and as a result it has given me a large degree of satisfaction upon completion. This project has given me opportunities to present it to legislators on Capitol Hill and to my peers at the student research symposium. Though this project accurately represents a large part of the progress I have made during my university experience, it also proved to be a process through which I was able to learn and become more familiar with how to not only do research but to communicate it to others as well.

One of the first struggles I had with this project was starting it. Before beginning work on anything, it looked like such a huge and insurmountable project, which makes sense since I had never before done a project quite of this magnitude. Even after changing my topic once, I had a hard time building up the motivation to be able to do any substantial work on it for a long time. There was even a period of time where I contemplated not doing this capstone project so I could lessen the burden of my senior year. It wasn't until my thesis advisor talked to me one day about an opportunity to present this research at the Research on Capitol Hill event when I finally committed to this project 100%. When I decided to take part in this event, it solidified my commitment to both finish and start this project.

Due to a large capstone history paper I wrote during the Fall 2015 semester, the first substantial work I did on this project was during the winter break. I spent the whole last week of the break working all day everyday on getting my calculations to the point where I could turn in a poster detailing my findings by the beginning of the Spring 2016 semester. Once I had these tentatively finalized conclusions ready for my presentation, I stopped working on it for a while in order to finish other obligations.

My second big push on this project was during Spring Break where I once again spent the majority of my time working on this thesis. Due to the little amount I'd spent working on it since December, I had to spend quite a bit of time looking through the 8 pages of calculations and scribbles I had made to find my conclusions before I could get back to work. In order to prevent this from happening again, I made an additional 8 pages of notes where I wrote interpretations of everything and explained what the notes meant. I also highlighted all the most relevant findings. After making this document, I didn't anticipate writing the paper being too difficult because I had already completed the bulk of the calculations, or so I thought.

This is where I ran into one of the greatest difficulties of the project. When I sat down to write about all the research and calculations I had made, I kept finding out that my calculations had not been 100% correct. At some point I ran into a fraction and had no documentation about where it came from. There were many inferences and concepts that made perfect sense in my head, but when I went to explain them in detail, the logic simply wasn't there. In order to make further progress on the project, I had to reexamine almost

everything I had done up to that point in order to both rework the parts that were incorrect and to more fully understand why the calculations worked the way they did.

This process of writing the paper is where I believe I made the most progress on this project. I have had plenty of experience doing math and making calculations, but most of the time this entailed proving a mathematical statement and then moving on to prove another unrelated statement. The mindset I allowed myself to develop prior to this project, where I would finish a problem and then almost immediately forget about it and move on to something new, was one of the biggest factors that caused my struggle from the previous paragraph. Trying to write a paper detailing my results made me realize that I did not actually understand what I had been doing as well as I thought I did.

One practice that would have helped mitigate my struggles would have been to make a mini write-up of everything I did as I did it. For example, whenever I find an equation or calculate a certain probability, if I would have immediately written down the processes, my reasoning behind the processes, and what the final result meant, I would have been able to not only not get lost after spending time away from the research, but I also could have possibly copied these descriptions into the final paper. In this way, I could have been writing my paper and doing the research simultaneously.

Another practice I should have done was have a more consistent schedule of working on the project instead of three or four periods where I pounded out as much as I could in as little time as possible. This would have reduced the amount of time I would have had in between sessions to forget about what I had previously done. Similarly, it would have been good to get over my intimidation for the project and just started working

on it sooner. When beginning a project like this, it's normal not to know what one is doing because the work has not yet been put into it. Throughout this project, I have repeatedly found the following concept taught by my thesis advisor to be true: if you put enough time and work into the research, you will make progress. Sometimes I have found this progress means figuring out that you previously made a mistake in your research and have to go back to fix it. In the end, I have found that it doesn't matter how many times the research starts going in the wrong direction because if it is good research, it will typically be directed back in the direction it is supposed to go.

I am very pleased with the experiences I have been given through working on this project. Throughout my college career, I have never felt like I was working on something that was as applicable to the real world as this project. This alone I find to be very exciting. Through this research, I feel like I have answered the age old question of high school students who, in the drudgery of some math class they don't like, ask, "When am I ever going to use this?" The true answer here, is whether or not you ever do use it, the most important takeaway is that you now know how to use it for whenever you may need it.

Bibliography

"Crime Index Offenses Reported." The Federal Bureau of Investigation. 1996. Accessed December 28, 2015. https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/1996/96sec2.pdf.

Imwinkelried, Edward J., "The Use of Evidence of an Accused's Uncharged Misconduct to Prove Mens Rea: The Doctrines Which Threaten to Engulf the Character Evidence Prohibition," Ohio State Law Journal, vol. 51, no. 3 (1990), 575-604. https://kb.osu.edu/dspace/bitstream/handle/1811/64070/OSLJ_V51N3_0575.pdf.

Lisak, David, and Paul Miller. "Repeat Rape and Multiple Offending Among Undetected Rapists." Violence and Victims 17, no. 1 (2002): 73-84. Accessed December 28, 2015. http://www.davidlisak.com/wp-content/uploads/pdf/RepeatRapeinUndetectedRapists.pdf.

McBride, Jon. "BYU Professor Works to Help Victims of Rape through In-depth Research and Training." BYU News. April 06, 2016. Accessed April 11, 2016. https://news.byu.edu/news/byu-professor-works-help-victims-rape-through-depth-research-and-training.

"Reporting Rates." Rape, Abuse and Incest National Network. Accessed December 28, 2015. https://rainn.org/get-information/statistics/reporting-rates.

Smoland, Dain. "Keep Calm and Argue the Facts: A Pragmatic Approach to the Doctrine of Chances." Utah Bar Journal 26, no. 5 (September/October 2013): 45-49.

Author Biography

Ryan Wallentine is a senior at Utah State University that will be graduating in May 2016 with a dual major in Mathematics and History with minors in Japanese and Asian Studies. Near the end of his junior year of college, Ryan learned enough Japanese through self-study to be able to test into the second year Japanese classes in order to add a Japanese minor by adding just one year onto his university experience. Ryan has been the recipient of the A-pin for having a 4.0 GPA for two consecutive semesters with at least 15 graded credits, the Lilywhite Scholarship through the college of science, and recently received the Outstanding Undergraduate Mathematics Award. After graduation, Ryan will spend at least a year teaching English abroad in Japan after which he plans on pursuing a law degree. Ryan hopes these experiences along with those from his undergraduate years will help him in pursuing a career that will allow him to travel and live in many different places and countries.