110090

ESTIMATION OF INCARCERATION AND CRIMINAL CAREERS USING HIERARCHICAL MODELS

by

Mark J. Schervish

and

John P. Lehoczky

Progress

Final Report

Contract No. 85-BJ-CX-0004

Department of Statistics

Carnegie-Mellon University

June 11, 1986

1. Introduction

This report contains the results of a statistical analysis of the 1979 Survey of Inmates of State Correctional Facilities. The statistical analysis was undertaken to gain insight into the prison population, the criminal careers of prisoners and the nature of the offender population. The analysis is complicated for three major reasons. First, the survey does not represent a random sample from the offender population. This means that a variety of statistical adjustments are needed to overcome the sampling bias. Second, the survey was not adjusted for the length-biasing effect of sentence lengths. This problem, which is explained in detail later in the report, requires adjustments to properly interpret the nature of the criminal careers of those in prison. Third, both the prisoner and offender populations are very heterogeneous, exhibiting a wide range of offending behavior (crime types) and rates. It is crucial to use statistical models which can capture this heterogeneity rather than using traditional models and methods in which the populations are treated as homogeneous.

This report is organized as follows. Section HIERMODEL contains a discussion of hierarchical modelling, a simple approach that incorporates heterogeneity into the offender population. The associated estimation technique, empirical Bayes methodology, is described and illustrated by an example. Section LENGTHBIAS contains an introduction of the problem of length-biased sampling and its impact on proper statistical inference length-biased sampling is a particular problem for proper analysis of prison survey data. Section EXPLOR contains an exploratory analysis of the prison survey data. Section MODELFIT contains results of model fitting using the techniques outlined in earlier sections. Conclusions are also presented in Section MODELFIT. In Section FURTHER, we discuss further analyses which could be performed with this data set.

2. Hierarchical Models

The analysis of prison inmate survey data gives rise to many complex statistical problems. On the surface, one starts with a random sample of inmates in a collection of correctional institutions and administers an extensive survey instrument, for example the Survey of Prison Inmates of 1974 or 1979. Given that this is a random sample, it is tempting to generalize the results to a reference population. This is certainly valid if the reference population is taken to be all prisoners at the time of the survey, but it is not valid if the reference population is taken to be the population of all offenders.[1] The difficulty is that the collection of all offenders in all correctional institutions at any point in time is <u>not</u> a random sample of all offenders. In fact, the prisoners tend to be more frequent offenders, offenders with prior records and often commit more violent crimes. It follows that a random sample of prisoners at any particular time cannot provide a representative sample of the offender population at that time. Nevertheless, one wants to use such a survey to learn about typical offenders and to investigate the impact of variations of enforcement and imprisonment policies on the offender and prison populations. This goal can be realized, but a great deal of care is required.

Interestingly, there is a statistical approach which allows one to use a random sample of prisoners to learn about the offender population. The approach is based on the following ideas. We recognize that the offender population is very heterogeneous. It consists of individuals having a wide range of crime commission frequencies, distinct crime-type selections, different arrest rates, different career lengths, etc. The heterogeneity of the offender population can be built into a statistical model by introducing a hierarchical structure. There are two levels to the model. At the top level, each individual is assigned a set of parameter values which govern his criminal career. Each individual is assigned an independent set of parameters; however, these parameters are drawn from a common multivariate distribution. Each individual is at a lower level of the hierarchy. Once the parameter values have been assigned, the offender proceeds to embark on a criminal career.

A final issue concerns using hierarchical models to help to correct for the bias of surveying only prisoners instead of a sample of the general offender population.

---

[1]Even generalizing to the population of prisoners must be done with care, the phenomenon of length-biased sampling must be taken into account. This is discussed in detail in Section 3

There are two aspects to the inference. First, one would like to use the survey data set to learn about the overall offender population. For example, to learn the distribution of the individual offense rates or of the career lengths. This is especially useful in characterizing the offender population and understanding the impact of changes in criminal justice system policies. Second, one would like to make estimates of the parameters associated with each individual. There is a standard statistical approach to this called empirical Bayes analysis. This approach allows first for the estimation of the distribution of offender parameters and then for the estimation of the individual parameters. We will use the empirical Bayes approach in this study, but will restrict attention to the first aspect, namely estimating the distribution of the offender population parameters. We will shortly present an example to illustrate the empirical Bayes method.

If we focus on any particular offender and take note of his/her attributes (age, race, offense rate, career length, etc. - some of which are unobservable but can be represented as unknown parameters), then using a statistical model, we can compute the likelihood of his/her record at the time of the survey and the likelihood he/she will be in prison at the time of the survey. When dealing with prison survey data one must take note of the fact that only prisoners are eligible to be in the survey, and this conditioning event has a differential impact across the offender population, making it more likely that a high rate or violent offender will be included in the survey. By conditioning on this event, one corrects the likelihood function, and proper inferences can be made.

2.1. Example - Empirical Bayes Estimation

The survey of prison inmates consists of extensive data for a sample of prisoners. The individuals in the sample are very heterogeneous in that they exhibit wide variability in their parameter values. Given their histories, we would like to estimate the individual parameter values and the distribution of parameter values in the population. Consider the following simulation experiment. Suppose that we have a sample of 10 individuals. Each has a parameter λ chosen from an exponential distribution with unknown parameter β. That is, λ is a random variable with density function $f(\lambda) = \beta\exp(-\beta\lambda)$, $\lambda \geq 0$. Each of the individuals commits crimes and may or may not be arrested for each of them. We assume that the times between arrests are independent and exponentially distributed with mean $1/\lambda$ where λ is the parameter choosen by that individual. Assume that we have 5 interarrest times for each individual or a total of 50 data points. We wish to estimate β (which gives the population distribution) and $\lambda_1,\dots,\lambda_{10}$ for the 10 individuals. The following is a simulated data set for β=1. The 10 values of λ were drawn from an exponential (1) distribution. For any subject, 5 interarrest times were generated from an exponential distribution with the chosen parameter.

Table 1: Interarrest Times

Subject Number	λ	1	2	3	4	5	Avg.($\bar{x}$)	($1/\bar{x}$)	Empirical Bayes
1	0.931	1.10	2.82	0.35	0.26	0.85	1.076	0.929	0.931
2	0.070	13.78	1.77	12.91	4.39	6.16	7.802	0.128	0.150
3	0.485	0.65	3.08	1.44	0.66	7.79	2.724	0.367	0.409
4	0.102	3.23	2.21	1.55	5.86	5.98	3.766	0.266	0.302
5	0.296	0.69	0.35	1.84	1.78	2.12	1.356	0.737	0.765
6	0.172	8.42	0.05	4.28	6.31	0.46	3.904	0.256	0.291
7	2.010	0.03	0.41	0.40	0.26	0.14	0.248	4.032	2.601
8	2.172	0.16	0.38	0.28	0.10	0.36	0.256	3.906	2.556
9	0.108	7.34	0.48	17.39	2.98	11.32	7.902	0.127	0.148
10	1.630	4.24	0.22	0.09	0.78	1.12	1.290	0.775	0.798

If one were to ignore the hierarchical structure in which the 10 individuals are related through the exponential (β) distribution, then an individual λ_i would usually be estimated by the maximum likelihood estimate $1/\bar{X}_i$. Consequently, subject 1 would be estimated by .929, virtually equal to the true value of .931. Most of the estimates are reasonable except for subjects 7, 8, 10 and perhaps 5. An overall measure of the accuracy of the 10 estimates is given by $\sum_{i=1}^{10}(\lambda_i - 1/\bar{X}_i)^2 = 8.072$.

We now use the empirical Bayes approach. This requires first estimating β, then estimating the individual λ values. The estimate of β is the unique solution of the fixed point equation:

$$\sum_{i=1}^{10} \left[1/(\bar{X}_i + \beta/5) \right] = 25/(3\beta).$$

In this case, β = 1.067, nearly equal to its actual value of 1.00. Next, an individual λ is estimated by the formula $6/(1.067 + 5\bar{X}_i)$. The individual estimates are given in the Empirical Bayes column of Table 1. Here the overall measure of accuracy becomes $\sum_{i=1}^{10} (\lambda_i$ - empirical Bayes estimate$)^2$ = 1.477, a major reduction from the 8.072 value using the MLE alone. It should be noted that major improvements occur for subjects 7 and 8. This simulation experiment is merely to illustrate the empirical Bayes method. In the example, it appears that the method offers very large improvements in mean square error over individual maximum likelihood estimates. This is true, but in this case is partly the product of assuming each λ is sampled from an exponential distribution. In practice we will not know the form of the distribution at the level of the hierarchy, and so will be forced to use a flexible family of distributions. The important point is that one can consider models in which individuals exhibit heterogeneity and can solve the estimation problems with empirical Bayes methods. These methods appear to be ideally suited to dealing with the large heterogeneity encountered in criminal justice data sets. The reader should consult Morris (1983), Deely and Lindley (1981) or Dempster, Rubin and Tsutakawa (1981) for an overview of the empirical Bayes estimation procedure, Rolph, Chaiken and Houchens (1983) have used hierarchical models as well.

3. Length-Biased Sampling Problems

In this section, we discuss a difficulty which arises in the analysis of prison survey data, namely length-biased sampling. The 1974 or 1979 Survey of Inmates of State Correctional Facilities are carried out essentially by taking a random sample of prisoners over a short period of time. This gives a random sample of prisoners in one sense, but not in another. Basically, it offers a picture of the occupancy of the prison cells but may not offer a clear picture of a "typical" prisoner. A brief discussion of length-biased sampling and its effects is given by Karlin and Taylor (1975, p. 175, p. 195). We illustrate with three simulation examples.

3.1. Example 1

Consider the following experiment. Suppose we consider a prison with 10 cells. At time 0, each cell is filled with a prisoner having a random sentence length (in the experiment we assume sentence lengths have an exponential distribution with a mean of 24 months). When a sentence is completed, a new prisoner enters that cell, again serving a random length sentence. The simulated sentences are presented in Table HYPOOCCUPAN. At any fixed time, we can examine the sentence lengths of the prisoners in the 10 cells.

Table 2: Hypothetical Cell Occupancies

Cell number	Sentences (in months)
1	60, 2, 4, 12, 28, 8, 13, 28, 14, 62, 40
2	50, 3, 18, 14, 13, 25, 4, 24, 6, 60, 20, 21
3	23, 7, 33, 65, 3, 7, 1, 4, 11, 24, 73
4	7, 6, 77, 8, 1, 9, 13, 58, 14, 30, 13, 22
5	2, 36, 1, 49, 14, 74, 17, 102
6	5, 23, 4, 19, 44, 48, 7, 6, 60, 37
7	19, 1, 27, 1, 4, 23, 18, 6, 1, 98, 3, 8, 6, 1, 2, 16, 22
8	22, 6, 9, 1, 59, 44, 10, 54, 8, 29
9	3, 9, 35, 18, 14, 20, 15, 19, 32, 4, 8, 46, 38
10	14, 4, 7, 60, 36, 31, 13, 18, 17, 4, 8, 63

There are a total of 116 sentences with an average length of 22.48 (close to the mean 24 of the distribution). Suppose we now look at the sentence lengths being served if we look at all the cells at time 10 years (120 months), 15 years (180 months) and 20 years (240 months). They are in Table SENSER.

It can easily be seen that the sentence lengths observed at the three time points are generally much longer than the "typical" sentence lengths. Indeed the three averages are 43.1 at 120 months, 46.3 at 180 months and 44.7 at 240 months.

Table 3: Sentences Served at Different Times

cell number	120 months	180 months	240 months
1	13	62	40
2	25	60	21
3	65	73	13
4	13	14	22
5	74	17	102
6	48	60	37
7	98	98	22
8	44	54	29
9	15	08	38
10	36	17	63

This is roughly twice as long as the average sentence length of 22.48. This is as expected and is an illustration of length-biasing. If we take a snapshot of sentence lengths at a fixed time, we are likely to see longer rather than shorter sentences.

3.2. Example 2

We now consider a second simulation example to illustrate that the sampling scheme used is more likely to capture more violent offenders or offenders with longer records. This is a product of the length-biased sampling and the fact that long sentences are highly correlated with violent offenses or long records.

Suppose there are two types of offenders, low rate (L) and high rate (H). Low rate offenders have sentences of length 1 while high rate offender have sentences of length 5. Assume that when a sentence is completed, a new offender appears to fill the cell and is equally likely to be L or H. As in example 1, there are 10 cells in the prison. Table CELLOCC shows the occupancy of each cell over the first 20 time units.

Table 4: Hypothetical Cell Occupancies

All number	Offender Sequence	Occupancy at time 19.5
1	L,L,L,L,H,L,H,H	H
2	H,H,H,L,H	H
3	L,L,H,H,H,H	H
4	L,L,H,H,L,H,L,L	L
5	H,H,H,L,L,L,H	H
6	H,H,L,H,L,H	H
7	L,H,L,L,H,H,L,L	L
8	L,H,H,L,L,H,H	H
9	L,H,H,L,H,H	H
10	H,H,L,L,L,L,L,L,H	H

There are 34 L offenders and 36 H offenders, virtually identical as they should be. Nevertheless, at time 19.5, the 10 cells are occupied by 8 H and 2 L offenders, not representative of the 1/2 probability. The problem is again the length-biasing. We are more likely (probability 5/6) to find a cell occupied by an H than L because of the five to one sentence length ratio.

3.3. Example 3

Finally, consider a third example in which every offender receives a sentence with the same distribution (exponential with mean 5 months). However, the population consists of 50% high rate offenders (average street time between imprisonments of 1.25 months) and 50% low rate offenders (average street time between imprisonments of 5 months). We generated 1000 such offenders and tracked them for 20 months with each offender starting on the street. At that time we found 256 low rate offenders and 394 high rate offenders in prison. The average of the sentences being served by all of the 650 offenders in prison at that time was 9.6 months. The average of all sentences generated in the 20 month period for all 1000 offenders was 5.1 months. We see how the method of sampling offenders who are in prison at a fixed time can lead both to finding prisoners serving longer than average sentences and to finding a disproportionate number of high rate offenders compared to the population at large. For completeness, the theory of alternating renewal processes says that after a long time t, the probability is p/(p+s) that an offender will be in prison at time t. Here p is the average time spent in prison per sentence and s is the average time spent on the street. Given a 50-50 mixture of high and low rate offenders, the expected number of high rate offenders in prison at time t is $1/2 \times 1000 \times \frac{5}{5+1.25} = 400$ and for low rate offenders it is $1/2 \times 1000 \times \frac{5}{5+5} = 250$. The expected sentence length for those in prison is 2p = 10, in our example. One can see if the simulation results conform closely to the theory.

These three examples illustrate the problems which can arise when one tries to use prison survey data to make inferences about offenders in prison. Fortunately, it is rather easy to correct each of these effects, and this is done in Section CORRECTB.

4. Exploratory Analysis of Prison Inmate Data

In this section, we present an analysis of the 1979 Survey of Prison Inmates. The overall goal is to gain insight into the evolution of criminal careers. Consequently, we focused on the time sequence of events: when did the career begin (juvenile or adult - at what age). How many imprisonments has the offender had - of what length? - for what crimes? How much street time separated the imprisonments? Do covariates such as age or juvenile record matter? The goal, then, is to gain insight into criminal careers and to build models for them. This is a sequential process in which one begins with the simplest models and proceeds to more complicated models as necessary.

Before proceeding it is important to mention that there are problems with the survey data. Some, such as the length-biased sampling, can be suitably corrected with statistical adjustment. Others cannot be so simply corrected. First, there are certain internal inconsistencies in the data and cases where the interviewer did not follow the survey instrument instructions. This problem is discussed below. Second, there are problems associated with any self-report survey not based on official records. One can expect to find biases in the set of inmates who chose to participate in the study. One cannot be sure of the veracity of those who do participate. Furthermore, there are clear recollection biases for such a survey instrument. People tend to report in round numbers and accuracy decays strongly with time. All of these issues indicate that the results of our analysis must be treated as tentative until they can be replicated using different data sets.

We began by examining the data in detail to see how it might systematically depart from what would be predicted by the simplest stochastic models we have considered. We looked at the street times preceeding imprisonments as well as the times served for the various crime types. We looked at whether or not the offender had served time as a juvenile and the offender's race. We also compared first imprisonments to later ones. In all analyses of the data, we used only those inmates whose stated career histories were internally consistent. That is, we excluded all inmates who stated that they were imprisoned a certain number of times, but gave details on a different number of imprisonments. This left us with 7,033 inmates serving a total of 10,873 imprisonments. Using this imprisonment data, it is possible to study the sequence of street times and imprisonment lengths for each offender. We describe these next.

4.1. Street Times

Street times, measured in months, for any particular offender can be separated into the time before the first imprisonment and the times between imprisonments. First, consider the street times preceeding imprisonments for the various types of crime. We classified imprisonments according to the most serious charge listed by the offender. The first street time for an offender is defined to begin at the 18th birthday or after release from prison, if the offender was in prison on the 18th birthday.

4.1.1. Race and Offense Type

The first tabulation is according to race. Table OFFENSEBYRACE gives the median and average street times preceeding all imprisonments in the sample arranged by type of offense and race. The category "Other Violent" includes rape and aggravated assault. The category "Other Property" includes larceny and auto theft. The category "Other" includes all non-violent, non-property crimes such as drug offenses. For all offense types except the catch-all category "Other", blacks spent less time on the street before imprisonment. However, the differences are much smaller than the differences among the different offense types. The latter exhibits large variability ranging from 41.9 months for burglary to 118.9 months for murder for the overall average.

Table 5: Months of Street Time Tabulated by Offense and Race

	Black	Non Black	ALL	
	67.26	77.53	--	Median
Murder	115.11	122.62	118.91	Average
	514	525	1039	Count
	33.00	34.00	--	
Robbery	46.44	56.70	50.70	
	1090	773	1863	
Other	55.50	60.00	--	
Violent	90.99	99.22	95.16	
	984	1010	1994	
	27.00	27.00	--	
Buglary	40.93	42.59	41.92	
	730	1076	1806	
Other	35.00	43.00	--	
Property	56.59	70.11	63.89	
	1175	1378	2553	
	56.09	49.00	--	
Other	80.00	77.82	78.71	
	640	934	1574	
Not	40.00	46.00	--	
Specified	71.06	86.02	80.61	
	17	30	47	
	--	--	--	
ALL	67.65	74.39	71.20	
	5145	5728	10873	

4.1.2. Time Until First Imprisonment

Next, we investigated the times between imprisonments after the first and compared them with the street time before the first imprisonment. Since offenders who have been imprisoned as juveniles may be treated differently by the criminal justice system, we distinguish them from those not arrested as juveniles. Table FIRST summarizes this data.

Table 6: Comparison of Times Until First Adult Imprisonment With Times Between Later Imprisonments

	No Juvenile Imprisonments	With Juvenile Imprisonments	
First	64.0	28.0	Median
Adult	97.1	41.4	Average
Imprisonment	6032	935	Count
Later	25.2	18.0	
Adult	41.1	29.8	
Imprisonments	2650	1110	

It should be noted that 146 imprisonments are missing from this table, because the offender did not say whether or not he/she had been imprisoned as a juvenile. (This involves 66 different offenders.) The most striking feature of this table is the large difference between the time until first adult imprisonment and the times between later imprisonments. Also, there is a remarkable similarity between the later times between imprisonments for those with no juvenile record and the time until the first adult imprisonment for those with a juvenile record. It is as if a juvenile imprisonment history played the same role as the first adult imprisonment in shortening street time. Another feature of note is the fact that there are only 2657 later adult imprisonments for those 6034 offenders who had no juvenile imprisonments, whereas there are more later imprisonments (1099), for those with a juvenile imprisonment, than there are offenders (933). This suggests that those with a juvenile record are more likely to be repeat visitors to prison as adults. The large difference between time until first imprisonment and times between later imprisonments suggests that there may be some offenders who delay many months beyond age 18 before embarking on their criminal careers.

4.1.3. Sucessive Street Times

The previous subsection compared times to first adult arrest with the average of the subsequent street times. It is important to study the sequences of street times more carefully. Table JUVBYPRIOR below is used to see if the decrease in times between imprisonments continues beyond the second adult imprisonment. In this table we have treated the first adult imprisonment of those with a juvenile history as the second adult imprisonment, in the light of the results from Table 6 above.

Table 7: Street Times Classified By Juvenile History and Number of Prior Imprisonments

Priors	No Juvenile Imprisonments	With Juvenile Imprisonments	Not Stated	ALL	
	64.00	--	51.00	--	Median
0	97.13	--	72.14	96.86	Average
	6032	0	66	6098	Count
	28.00	28.00	30.84	--	
1	44.52	41.42	37.15	43.29	
	1721	935	60	2716	
	21.00	18.50	9.00	--	
2	34.90	30.51	14.91	32.46	
	615	643	14	1272	
	21.00	19.84	18.00	--	
3	36.69	30.59	48.03	33.23	
	205	298	5	508	
	24.00	16.00	78.00	--	
4	29.89	28.09	78.00	29.05	
	69	111	1	181	
5	22.10	11.00	--	--	
or	31.30	21.52	--	25.51	
More	40	58	0	98	
ALL	--	--	--	--	
	80.03	35.13	51.49	71.20	
	8682	2045	146	10873	

There appears to be a smaller decrease from one to two priors than from zero to one prior, and the street times are fairly stable after that point. Those offenders with imprisonments as juveniles continue to have consistently shorter street times between imprisonments throughout their careers.

4.2. Imprisonment Lengths

We now consider the other part of criminal careers, the imprisonment lengths. In all analyses described in this section, the current imprisonment is *not* considered, because it is censored, that is, we do not know how long it is, except that it is at least as long as the time already spent at the time the survey was taken.

4.2.1. Race and Offense Type

Table PRISONOFFBYRACE gives the median and average time spent in prison tabulated by race and offense type.

Table 8: Months of Prison Time Tabulated by Offense and Race

	Black	Non Black	ALL	
	42.000	47.500	--	Median
Murder	59.084	65.000	60.815	Average
	29	12	41	Count
	24.000	33.500	--	
Robbery	32.851	40.560	35.792	
	240	148	388	
Other	9.000	9.000	--	
Violent	19.025	20.413	19.715	
	246	243	489	
	12.570	14.000	--	
Burglary	17.008	18.141	17.657	
	348	466	814	
Other	7.250	9.000	--	
Property	10.723	14.056	12.432	
	648	682	1330	
	6.000	5.000	--	
Other	10.737	11.793	11.357	
	321	456	777	
ALL	--	--	--	
	16.690	17.519	17.123	
	1833	2007	3840	

It would appear that blacks spent slightly less time in prison per imprisonment for all offense types except those in the catch-all category "Other". This is surprising, since Table 5 indicates that blacks spent less time on the street leading up to imprisonment for all offense types except "Other". The differences between race, however, are not nearly as large as those between offense type.

4.2.2. First Imprisonments

One plausible line of reasoning suggests that the criminal justice system punishes repeat offenders more severely than first offenders, hence the first imprisonment would be expected to be shorter than later ones. Table FIRSTPRI illustrates

Table 9: Comparison of Lengths Of First Adult Imprisonment With Lengths Of Later Imprisonments

	No Juvenile Imprisonments	With Juvenile Imprisonments	
First	9.0	12.0	Median
Adult	15.3	17.9	Average
Imprisonment	1721	643	Count
Later	12.0	12.0	
Adult	18.5	20.1	
Imprisonments	929	467	

that this is indeed the case in the data set, although later imprisonment times for inmates who had been imprisoned as juveniles were not as much longer than the first one as for those with no juvenile imprisonments. Once again, this suggests that the existence of a juvenile imprisonment plays a similar role to the role played by the first adult imprisonment.

4.2.3. Successive Imprisonments

Next, we examined the sequence of times spent in prison categorized by offense type. In this analysis, an inmate with a juvenile record is assumed to have one adult imprisonment prior to the actual first adult imprisonment.

To clarify Table PRIPRIORBYOFFENSE, the 11 inmates in the murder category in the 1 prior imprisonment column are all people who served their second adult imprisonment for murder, regardless of the offense for which they served their first adult imprisonment. This table shows that the second adult imprisonment tends to be longer than the first and the third longer than the second, but the later ones tend to be shorter. This surprising observation could be explained by two different reasons. First, offenders who get four or more adult imprisonments must be getting shorter imprisonment times just to be able to fit so many into a lifetime. Secondly, these offenders may be committing less serious crimes, and, hence be getting shorter sentences.

Table 10: Imprisonment Times by Number of Prior Imprisonments and Offense Type

	Number Of Prior Imprisonments 0	1	2	3	4	5 or More	ALL	
	39.000	55.000	67.500	56.000	--	--	--	Median
Murder	52.656	72.366	81.250	56.000	--	--	60.815	Average
	25	11	4	1	0	0	41	Count
	23.000	31.500	42.000	31.000	22.000	--	--	
Robbery	32.510	35.499	46.737	45.500	27.367	--	35.792	
	177	140	48	16	7	0	388	
Other	9.000	12.000	10.000	5.000	9.000	3.500	--	
Violent	17.792	21.368	28.397	12.662	15.500	12.650	19.715	
	245	145	58	23	6	12	489	
	12.000	13.000	17.500	12.000	13.500	28.500	--	
Burglary	16.019	18.309	20.140	17.277	17.133	27.667	17.657	
	331	300	124	43	10	6	814	
Other	6.470	9.250	9.600	12.500	5.800	8.000	--	
Property	10.015	13.051	14.709	20.246	14.968	7.503	12.432	
	637	446	165	52	19	11	1330	
	6.000	5.330	6.000	4.500	2.000	3.000	--	
Other	10.015	13.127	14.197	8.743	5.326	11.922	11.357	
	366	229	109	46	18	9	777	
ALL	--	--	--	--	--	--	--	
	15.309	18.227	21.038	18.083	13.936	13.359	17.123	
	1781	1272	508	181	60	38	3840	

4.3. Ages of Inmates

There is concern that the imprisonment process may not be stationary over time. That is, there may be local, regional or national trends toward longer or shorter street times or toward longer of shorter imprisonment times. It also might be the case that an offender may terminate his career or simply stop being imprisoned after some age. In the detailed comparisons, we look only at ages 30 and below, because the survey data become very thin (i.e. very few observations in each category) beyond that age. This tends to reduce the impact of long-term trends as well.

4.3.1. Comparison of Age Distributions

Figure AGECOMPARE shows the empirical distribution of the ages of inmates in the 1979 prison inmate sample together with the distribution of ages of people in the United States population at large in 1979. The difference between the two distributions is striking, and suggests that prisoners are an age-biased sample from the population at large. Offenders of ages 19-24 are over represented in the survey compared with the relative size of this group in the population.

4.3.2. Relation Between Age and Street Time

We also considered the relationship between the age at release from prison and the time spent on the street until the next imprisonment. (Times until the first imprisonment are not considered here.) Figures 1 through 6 display boxplots of the natural logarithms of street times between imprisonments arranged by the type of offense at the end of the street time and the age at the beginning of the street time. To read the plots, the I's appear at the sample quartiles and the + appears at the sample median. The --'s extend to 1.5 times the distance between the two I's or until the most extreme observation, within the 1.5 interval, whichever comes first. If there are more extreme observations outside the interval, they are indicated by *'s. If any observation is more than 3 times the distance between the I's above or below an I, it is indicated by an O, to denote "outlier."

To convert the logarithm of street time back to street time, it is necessary to exponentiate. We recall that 1 corresponds to 3 months, 2 to about 7 months, 3 to 20 months, etc. To illustrate with an example, consider the first row of Figure 2, consisting of 42 observations corresponding to age 18. The median is around 3.6 or about 36 months. Here the 1.5 times interquartile range distance is approximately 1.8, and our interval extends from about 1.8 (6 months) to 5.4 (221 months). There is an observation outside this interval at 0.0 (1 month).

Figure 1: Comparison of Population and Sample Age Distritutions

The purpose of the plots was to use them to look for some sort of trend in street time with age, indicating a non-stationarity of the street time process. The plots are of the natural logarithm of street times rather than the actual street times, because the distributions of street times are so skewed that visual comparison is virtually impossible in normal units units of time (months). There appears to be no visually noticeable trend in any of the plots, although there appear to be some fluctuations in some of them. The boxplots for Robbery, Other Violent, Burglary and Other Property show a surprising amount of stability over age, in spite of the small sample size. Murder and other show somewhat less stability but still do not exhibit any trends. Apparently this group is quite homogeneous (within offense type). This indicates that trends can be largely ignored for this range of ages. There isn't enough data in each offense type on offenders above age 30 to make with confidence any statements about trends.

Figure 2: Boxplots of Log(Street Times) Before Imprisonment For Murder

```
AGE                                                  COUNT
 18     *            -----I     +   I--------          42
 19 *   *        --------I     +   I---------          24
 20 *   *           -------I   +  I------              26
 21                     ----I    +     I----           18
 22          ------------I     +   I--------------     22
 23                       ----I +I---   * O            24
 24                 **       I+   I---    *            17
 25  O               *     --I + I-------*             13
 26                  ---------I  +    I------          20
 27 ---------------I              +      I------        8
 28                 ---------I   +    I------          10
 29                 -------I          +  I----         11
 30                       ------I+  I--                10
    ----+---------+---------+---------+---------
       0.0       1.5       3.0       4.5
```

Figure 3: Boxplots of Log(Street Times) Before Imprisonment For Robbery

```
AGE                                                    COUNT
 18   *    *    -----------I     +   I----------        170
 19   *       ------------I    +   I------               86
 20        ---------------I    +    I--------            68
 21   *    ---------------I    +     I--------           76
 22   **   *      --------I     +  I-------              71
 23       *---------------I    +    I---------           60
 24 O      *    * ----------I   +   I----------          38
 25   **        -----------I +      I---------           44
 26            * *        ---I +     I------------       27
 27   *             --------I   +    I---------          24
 28   O                   -----I +    I-------       *   25
 29               ----------I    +   I---------          20
 30   *            ------I     +     I--                 15
    -------+---------+---------+---------+---------
          0.0       1.5       3.0       4.5
```

Figure 4: Boxplots of Log(Street Times) Before Imprisonment For Other Violent

```
AGE                                                 COUNT
18 *   * *  ------------I     +   I---------           149
19           --------I     +     I----------            50
20 *   *      ---------I   +     I------                68
21     *      ----------I      +   I---------           51
22 *   ----------------I     +     I-----------         48
23 *                ------I    +   I---                 44
24 *   ----------------I    +     I---------            43
25       -------------I    +      I----------           40
26 O        *     --------I    +  I---------            39
27            ------------I    +   I--------            29
28                      ---I      +   I-----            19
29     *      -----------I   +       I-----             24
30 -------------------I        +         I---           16
   ----+---------+---------+---------+---------
      0.0       1.5       3.0       4.5
```

Figure 5: Boxplots of Log(Street Times) Before Imprisonment For Burglary

```
AGE                                               COUNT
18 *   ---------------I    +     I------------        259
19     --------------I   +    I--------               108
20 *   --------------I   +     I------------           81
21 ---------------I     +     I-------                 84
22 *      ----------I     +    I----------             76
23 --------------I      +    I-------                  56
24            ----------I   +    I--------             49
25 *        ----------I     +  I----------             46
26 *        ---------I    +     I--------              57
27 *          ---------I   +      I-----               27
28          ----------I       +   I-------             34
29 *   --------------I        +  I----                 14
30              ----------I   +     I---               21
   ----+---------+---------+---------+---------
      0.0       1.5       3.0       4.5
```

Figure 6: Boxplots of Log(Street Times) Before Imprisonment For Other Property

```
AGE                                                                   COUNT
 18           *-------------------I      +     I----------             262
 19           ----------------I      +     I------------               106
 20 *         *  *----------------I      +  I---------                 102
 21           ----------------I       +    I------------               106
 22           *   *    -----------I   +    I-----------------          108
 23           *   ---------------I      +    I-------                   78
 24         * *   **   ----------I   +   I----------                    71
 25           *   *    -----------I    +   I------------                59
 26           *  ---------------I    +    I-----------                  44
 27                  -------------I   +    I------                      49
 28               -------------I     +    I------------                 32
 29           *               -------I     +    I------                 32
 30           *               -----I     +  I----------                 28
     -+---------+---------+---------+---------+---------
            -1.5       0.0       1.5       3.0
```

Figure 7: Boxplots of Log(Street Times) Before Imprisonment For Other

```
AGE                                                                COUNT
 18                 *  * *  ---------I    +  I-----                 138
 19                 ------------I     +     I-----                   54
 20                  -----------I     +   I-----                     56
 21                 * *     -------I    +     I------                54
 22                 *       -----I    +   I--------                  41
 23 O                 ------------I   +     I-----                   55
 24                 *            --I    +   I------                  45
 25                 *          --I    +  I---------                  30
 26                 ------------I     +    I-------                  40
 27                 O         **  ----I + I-----                     39
 28                                ---I + I-------                   17
 29                          ------I    +  I-------                  28
 30                           -------I    +   I----                  26
     -------+---------+---------+---------+-------
          -2.0       0.0       2.0       4.0
```

4.3.3. Relation Between Age and Lengths of Imprisonment

Next we look to see if there is any noticeable trend in imprisonment length with age. Once again, we have made boxplots of the natural logarithm of imprisonment lengths, because the lengths in months are so heavily skewed.

None of the plots shows a noticeable trend in either direction as age increases, although the Property and "Other" crimes have a hint of upward trend starting in the mid to late 20's. Indeed, the Figure 10 for Burglary exhibits remarkable stationarity. These plots together with those in Section 4.3.2 suggest that the stationarity assumption of the stochastic models we intend to fit may be a good approximation when adjusted for offense type. However, the analyses in Sections 4.1.3 and 4.2.3 suggest that stationarity does not hold for overall street times and imprisonment lengths. This might be due to a trend toward more serious offenses later in life for those who continue to be imprisoned.

Figure 8: Boxplots of Log(Imprisonment Time) for Robbery

```
AGE                                              COUNT
 18           **      ------I   +  I------        48
 19 *       *  -----------I     +   I---          59
 20              *        ----I   +  I---         41
 21           *     *   ------I   + I-----        39
 22                   -------I  +   I---          29
 23           *     ---------I    + I-------      41
 24            -------------I      + I-----       28
 25                   -----I      +  I-----       14
 26                      *    --I  +I----         24
 27                 *          --I  + I-----      16
 28                  ---------I   +   I----       12
 29                         -----I +I----- *      11
 30                            ---I+I--            9
    ---+---------+---------+---------+------
     -2.0       0.0       2.0       4.0
```

Figure 9: Boxplots of Log(Imprisonment Time) for Other Violent

```
AGE                                                COUNT
 18     -----------------I    +      I---------      51
 19 O   *           ---------I  +   I---------       54
 20 *          -------------I     +    I------       51
 21     *      -------------I  +    I-------         51
 22       ----------------I    +      I----------    37
 23          -------------I       +  I---------      35
 24           -----------I      +     I------        27
 25                * -----------I  +   I-------      31
 26 *                   -----I +       I----         25
 27                      ----I  +       I-----       15
 28 O                       ----I  +I----- *         17
 29                  -----I     +     I---           12
 30                 -----------I     + I-------      19
    --------+---------+---------+---------+-----
         -2.0       0.0       2.0       4.0
```

Figure 10: Boxplots of Log(Imprisonment Time) for Burglary

```
AGE                                                  COUNT
 18            * ------------I    +  I---------       124
 19        O       **----------I   +  I-------        121
 20        O         ----------I  +  I---------       112
 21         O  *  *   *---------I +   I---             85
 22        O           ---------I  +  I-----           63
 23                    *   ------I  + I-----           64
 24 O                    * ------I  + I---    *        51
 25                  *      ----I   + I-----           38
 26                       ------I  +   I-----          31
 27                          -----I +  I-------        19
 28                       --------I +   I-             21
 29            O               ---I + I-----           15
 30                            ---I  + I-----          10
    --------+---------+---------+---------+-----
         -2.0       0.0       2.0       4.0
```

Figure 11: Boxplots of Log(Imprisonment Time) for Other Property

```
AGE                                                COUNT
 18       -----------I       +   I----------        158
 19       -------------I      +  I----------        166
 20       * * ------------I   +   I--------         144
 21 *     * -------------I      + I---------        130
 22 *   * * * -------------I  +    I----------      124
 23 *       *    ----------I    +  I--------         91
 24       * * -------------I   +    I------------    77
 25 O          **  ----------I   +  I----------       74
 26            ------------I     + I-------          64
 27       *      ----------I     +  I-----           49
 28           -------------I     +  I------          49
 29         -------------I     +   I-----            36
 30            *    ----------I   +  I----           30
    -------+---------+---------+---------+------
         -2.0       0.0       2.0       4.0
```

Figure 12: Boxplots of Log(Imprisonment Time) for Other

```
AGE                                                 COUNT
 18 *   ----------I   +     I-----------             72
 19     ------------I       +  I--------------       92
 20 *     -------------I      +    I--------         79
 21 *     **------------I    +   I-------            69
 22     * * -----------I      +  I------------       76
 23       ------------I    +     I----------         63
 24 *       -----------I       +   I---------        49
 25 *     -------------I       +   I--------         49
 26 *       ------------I     +    I------           42
 27 -------------------I         +    I---------     29
 28           *   *       ---I   +   I-              13
 29                    -------I    +I---------       19
 30       -------------I       +    I----            34
    -------+---------+---------+---------+------
         -2.0       0.0       2.0       4.0
```

4.4. Conclusions From the Exploratory Analysis

The features which we discovered in the exploratory analysis of the prison inmate data, can be summarized as follows:

- There are differences between blacks and non-blacks in street times and imprisonment lengths; however, these differences are small when compared with the differences between the different offense types or between successive imprisonments.

- Those who have juvenile records appear to behave, at the beginning of their careers, similarly to the way those with no juvenile record behave after release from their first imprisonment.

- The time spent on the street before the first imprisonment seems to be longer than the time spent between imprisonments.

- There are sizeable differences in imprisonment lengths between different offense types. Specifically, violent offenses lead to shorter imprisonments.

- For those who have completed their first imprisonment, first imprisonment tends to be shorter than the later ones, although not dramatically so.

- The distribution of the ages of inmates is dramatically different from that of the ages of people in the general population. The difference between the two population age distributions is greatest in the 19-24 age range which is greatly over-represented in the sample of inmates.

- Controlling for offense type, there appears to be no noticeable trend in either imprisonment length or street time with the age of the offender, over the 18-30 age group.

We will make use of these findings in constructing statistical models in Section MODELFIT.

5. Stochastic Model Fitting

It is clear from the exploratory analyses of Section 4 that a simple model of imprisonment would not be adequate to describe the many features we see in the 1979 prison inmate survey data. In this section, we describe those models which we have fit so far to the data, and give an indication of what further work should be carried out with this data set in the future. Despite its obvious inadequacies, we begin with a very simple model, and then build upon it.

5.1. The Simplest Model

The simplest stochastic model which one could fit to the survey data would be one in which the street time had the same probability distribution for every inmate before every imprisonment. In addition, the model would assign the same distribution to every imprisonment length of every individual. If we let these distributions be exponential with the mean street time equal to $1/\theta$ and the mean imprisonment length equal to $1/\delta$, then the likelihood function for the imprisonment survey data set would be

$$\theta^n \delta^{n-m} \exp(-\{x\theta + s\delta\}), \tag{5.1}$$

where n is the total number of imprisonments, m is the number of inmates in the survey, x is the total of all street times for all inmates, and s is the total of all imprisonment lengths for all inmates. The reason δ has exponent n-m instead of n is that the current imprisonment time for each inmate is censored and has likelihood $\exp(-t\delta)$ rather than $\delta\exp(-t\delta)$. The maximum likelihood estimates (MLE's) of θ and δ are easily calculated to be $\hat{\theta}$ = n/x and $\hat{\delta}$ = (n-m)/s. For the survey data we have, n = 10,873, m = 7,033, x = 779,356, and s = 242,336. Hence $\hat{\theta}$ = 0.013951, and $\hat{\delta}$ = 0.015846.

This would lead to the conclusion that offenders have an averge street time of 71.7 months and an average sentence length of 63.1 months. This model is a starting point for our analyses. The average street time and sentence length so derived from this model are, however, not appropriate estimates for the offender population. The most obvious problem is that the individual offenders in the survey do not provide a random sample of the offender population. Rather offenders in the sample tend to overrepresent prisoners who have relatively long sentences (the length-biasing problem). Moreover, prisoner populations in general are not representative of the offender population in that they tend to be more frequent offenders and possibly ones who commit more violent crimes.

5.2. Correcting the Sampling Bias

In order to use the data from the survey to make inferences about offenders at large, we need to calculate the likelihood conditional on the offender being in prison at the time of the survey. This correction is necessary, because the only way an offender can be included in the survey is to be in prison at the time of the survey In the simple model introduced in Section 5.1, the probability of being in prison at the time of the survey can be calculated to be

$$q(\theta,\ \delta) = \Sigma_{a=0}^{804}\ p_a \theta \{1-\exp(-[\theta+\delta]a)\}/(\theta+\delta), \qquad (5.2)$$

where p_a is the proportion of people in the population whose age is 18+(a/12). We stop at age 85 (804 months), because the prison population is almost entirely age 85 or below. This calculation is made using well-known results for two state continuous time Markov chains assuming that each individual is free at age 18. We can estimate θ and δ by maximum likelihood in this case, but there is no closed form solution, because the likelihood function is (5.1) divided by $q(\theta,\ \delta)^m$. The maximum likelihood estimates[2] are given by

$$\hat{\theta} = 0.00602$$

$$\hat{\delta} = 0.03955.$$

(For purposes of comparison with other models, the log-likelihood of this model is -67683. An increase of 2 points in the log-likelihood from the addition of one parameter can be considered a marginal improvement.)

Notice what the effect is of correcting the selection bias. The estimate of the imprisonment rate θ gets much smaller (less than half of what it is in the simplest model). This is due to the fact that high rate inmates are more likely to be in the sample, hence, the estimated imprisonment rate for the population of offenders at large should be smaller than the observed rate for inmates in the sample. Similarly, the estimate of the mean imprisonment length $1/\delta$ is now 25.28 months as opposed to the 63.1 month estimate of the simple model. Once again, we estimate the average imprisonment length for an offender at large to be smaller than the average for those in the survey, because long sentences are likely to be over-represented in the survey.

---

[2]Throughout this report, we present only the results, not the details, of the numerical analysis needed to determine the maximum likelihood estimates and the model log-likelihood.

5.3. A Model With Dropout

Since the distribution of the ages of inmates differs so dramatically from that of the population at large (see Figure 1), we felt compelled to deal with this fact in our models. One plausible way to do so is to assume that offenders eventually drop out of their careers. We have chosen to model dropout as a decison made after each imprisonment. With a certain probability, the offender will decide to end his/her career. Call that probability p, and make it another parameter in the model. The probability of being available for the survey is now equal to

$$q(\theta,\ \delta,\ p) = \Sigma_{a=0}^{804}\ p_a \theta\{\exp(-.5[\theta+\delta-Q]a)-\exp(-[\theta+\delta+Q]a)\}/Q, \tag{5.3}$$

where $Q = ([\theta+\delta]^2-4p\theta\delta)^{1/2}$, and everything else is as in (5.2). Notice that $q(\theta,\ \delta,\ 0) = q(\theta,\ \delta)$, so that (5.2) is the special case of (PDROP) with p = 0. The likelihood function for this model is

$$\theta^n([1-p]\delta)^{n-m}\exp(-\{x\theta+s\delta\})/q(\theta,\ \delta,\ p)^m, \tag{5.4}$$

where the extra factor of (1-p) for each completed imprisonment arises from the fact that the inmate has, of course, not ended his/her career after each of the previous imprisonments. We do not know whether the career will end after the current imprisonment. Once again, there is no closed form for the MLE's, but we can calculate them numerically using iterative methods. We find:

$$\hat{\theta} = 0.01192,$$

$$\hat{\delta} = 0.04288,$$

$$\hat{p} = 0.56742.$$

The log-likelihood for this model is -64678, which is 3005 higher than the log-likelihood for the model without dropout, and only one parameter has been added.

Notice that the estimated probability of dropout is quite high. This dropout probability forces many people to end their careers early and leads to a distribution of inmates' ages which more closely resembles the survey data. In fact, Figure NEWAGECOMPARE shows how remarkably close the predicted age distribution of inmates is to the observed data. The predicted age distribution is calculated as

$$\hat{p}_a = c\ p_a\ q(\hat{\theta},\ \hat{\delta},\ \hat{p}),$$

where c is the constant required to make the sum of $\hat{p}_a$ equal to 1.

All of the remaining models we will fit include a parameter for dropout. There was no point in including such a parameter in the simplest model of Section 5.1, because no one in the sample has dropped out of his/her career. It is only when we introduce the correction for sampling bias and compare the age distribution of

Figure 13: Comparison of Observed and Predicted Age Distributions

inmates to that of the population at large that it make sense to introduce dropout into a stochastic model based on inmate data.

5.4. Hierarchical Models

In this section, we consider the possibility of heterogeneity in the offender sample. The average street time before an imprisonment in the sample is 71.199 months with a standard deviation of 84.9 months. This standard deviation is about 20% higher than what would be expected if the sample were actually from an exponential distribution with mean 71.199. One possibility, of course, is that the exponential distribution does not fit well. Another possibility is that not all street times have the same exponential distribution. We saw in Section 4.1.2 that those with juvenile records had shorter street times than those without. This suggests that we consider θ to be a random variable with a distribution in the offender population. The model would be that the ith offender selects a value of θ, say θ_i, according to this distribution, and then proceeds to be imprisoned at rate θ_i.

5.4.1. A Simple Hierarchical Model

The simplest distribution for a random variable is one which is concentrated on only two values θ_1 and θ_2 and has probabilities p_1 and p_2 for the two values. Let $\theta_1 < \theta_2$. Then those offenders who "choose" θ_1 will be imprisoned at a lower rate than those who "choose" θ_2. To fit such a model we need not estimate the value of θ for every individual, but rather, only the two values θ_1 and θ_2 together with the probability p_1 (since $p_2 = 1 - p_1$). The likelihood function becomes more complicated now, because we must distinguish the street times of different inmates. If inmate i has n_i imprisonments with total length s_i and the total street time is x_i, then the likelihood function is

$$\Pi_{i=1}^{m}[\Sigma_{j=1}^{2} p_j \theta_j^{n_i}([1-p]\delta)^{n_i-1}\exp(-\{x_i\theta_j+s_i\delta\})/q(\theta_j,\ \delta,\ p)]. \tag{5.5}$$

The MLE's are

$$\hat{\theta}_1 = 0.004541,$$

$$\hat{\theta}_2 = 0.023017,$$

$$\hat{p}_1 = 0.431435,$$

$$\hat{p}_2 = 0.568565,$$

$$\hat{\delta} = 0.042763,$$

$$\hat{p} = 0.543828.$$

The log-likelihood is -64070, which is 608 larger than the log-likelihood of the model of Section 5.3, with two extra parameters. Note that the high rate is estimated to be about 5 times as high as the low rate.

5.4.2. Some Variations on the Simple Hierarchical Model

If a distribution over two values for θ provides a better fitting model than a single θ, it is possible that three values could be even better. In fact, we can use any distribution over the possible θ values we might like. To fit a distribution with three θ values requires changing the upper limit of summation in (5.5) from 2 to 3 and assuring that $\theta_1 < \theta_2 < \theta_3$. Also, we must add a parameter p_3 and assure that $p_1 + p_2 + p_3 = 1$. Doing these things allows us to obtain MLE's as follows:

$$\hat{\theta}_1 = 0.002877,$$

$$\hat{\theta}_2 = 0.017559,$$

$$\hat{\theta}_3 = 0.066545,$$

$$\hat{p}_1 = 0.294479,$$

$$\hat{p}_2 = 0.655751,$$

$$\hat{p}_3 = 0.049779,$$

$$\hat{\delta} = 0.042777,$$

$$\hat{p} = 0.537653.$$

The log-likelihood is -64020, which is only 50 higher than for the two-θ model, with two extra parameters. We are approaching the point of diminishing returns with models allowing a finite number of values of θ.

To carry the procedure one last step, in order to be sure that no great improvement is still possible, we fit a hierarchical model with a distribution for θ concentrated on four values. The log-likelihood only increased to -64018, or 2 higher than the previous model with two extra parameters. Adding one value of θ at a time appears to have reached its limit of usefulness.

5.4.3. A Continuous Parameter Model

Instead of requiring that θ_i, the imprisonment rate for offender i, be one of finitely many values, we can allow θ_i to be a random variable with a continuous distribution. A choice for that distribution which will make certain calculations easy is an element of the Gamma family, due to the close connection it has with the family of exponential distributions. Moreover, the Gamma family allows a wide variety of shapes and it is a reasonable choice.

If we replace the discrete distribution of θ with a Gamma distribution, the likelihood function becomes

$$\Pi_{i=1}^{m}\left[\int_0^{\infty}\theta^{\alpha+n_i}([1-p]\delta)^{n_i-1}\exp(-\{(x_i+\beta)\theta+s_i\delta\})\beta^{\alpha}/\{\Gamma(\alpha)q(\theta,\ \delta,\ p)\}d\theta\right], \tag{5.6}$$

where α and β are the parameters of the Gamma distribution, which need to be estimated. The MLE's are

$$\hat{\alpha} = 1.9888,$$

$$\hat{\beta} = 134.55,$$

$$\hat{\delta} = 0.042695,$$

$$\hat{p} = 0.55123,$$

and the log-likelihood is -64039. This can be considered an improvement over only the model of Section 5.3, because that is the only model for which it is a direct generalization. It has a substantial increase in log-likelihood (639) with only one additional parameter. The models in Sections 5.4.1 and 5.4.2 are also generalizations of the one in Section 5.3, but in a different direction. The continuous parameter model appears to provide a fit somewhere between that of the two-θ and three-θ models.

The parameters α and β of the continuous parameter model are not directly of interest, but they do specify the distribution of θ in the population of offenders. For example, the mean of θ is α/β, which has MLE of 0.01478. This represents the mean imprisonment rate of a randomly selected offender.

5.5. Models With Multiple Offense Types

Since we saw in Section 4.1.1 that street time varied according to the offense that was next committed, we tried to fit models in which different types of offenses led to different waiting times for the next imprisonment. As a start, we distinguished violent from non-violent offenses.

5.5.1. The Simplest Model

The simplest model we fit assumed homogeneity among the offender population, but that the time until the next imprisonment was the minimum of two independent exponential random variables with means $1/\theta_v$ and $1/\theta_{nv}$. Conceptually, the offender is waiting until the next time he/she is imprisoned for a violent crime with rate θ_v and at the same time, waiting until the next time he/she is imprisoned for a non-violent crime with rate θ_{nv}. Whichever comes first determines the type of offense for which he/she is imprisoned and the length of street time between imprisonments. Specifically, the time has an exponential distribution with mean

$1/(\theta_v+\theta_{nv})$, and the next crime is violent with probability $\theta_v/(\theta_v+\theta_{nv})$. If we let $\theta = \theta_v + \theta_{nv}$, then the probability of being in prison at the time of the survey is still given by (5.2), since the distribution of the times between imprisonments is exponential with mean $1/\theta$. The likelihood function for the part of the data consisting solely of the times between imprisonments and the lengths of imprisonments is still given by (5.4). However, we have more data which we did not take into account in (5.4). They are the indicators of the types of offense the imprisonments were for. Each imprisonment for a violent crime contributes a factor of θ_v/θ, and each imprisonment for a non-violent crime contributes a factor of θ_{nv}/θ. These extra factors will make the likelihood smaller, because the factors are all less than 1.0. However, it is easy to see that the model we are now considering is a generalization of the model in Section 5.3. If $\theta_v = \theta_{nv} = 0.5\theta$, then the model of Section 5.3 is a special case of the two-offense type model. The factors θ_v/θ and θ_{nv}/θ, which we did not include in the likelihood are all equal to 0.5 in that case. Hence, in order to compare the log-likelihoods of the two models, we must subtract $\log_e(2)$ from the log-likelihood of the model in Section 5.3 for every imprisonment. The total amount for the 10873 imprisonments is 7537.

Fitting the model described, yields the following MLE's:

$$\hat{\theta}_v = 0.005376,$$

$$\hat{\theta}_{nv} = 0.006554,$$

$$\hat{\delta} = 0.043001,$$

$$\hat{p} = 0.571517.$$

When making comparisons with the earlier models, we must remember that the overall imprisonment rate is $\theta_v + \theta_{nv}$, which has MLE = .011930. The log-likelihood for this model is -72161. This should be compared to -64678 - 7537 = -72215. The improvement over the model of Section 5.3 is 54 with one extra parameter.

5.5.2. A Hierarchical Version of the Model

Since the model of Section 5.4.1 made such a noticable improvement over the simple model of Section 5.3, we might expect the same generalization of the two-offense type model to provide a similar improvement. We generalize in the same way, by assuming that the jth offender is able to "choose" his/her $(\theta_{v,j}, \theta_{nv,j})$ from a distribution of pairs. A simple distribution analogous to the type used in Section 5.3 is to allow two different values for each of θ_v and θ_{nv}, making a total for four pairs of values. Since it seems unlikely a priori that the choice of violent imprisonment rate is independent of the choice of non-violent imprisonment rate, we fit a general distribution over the four pairs of imprisonment rates.

The MLE's for this model are

$$\hat{\theta}_{v,1} = 0.003228,$$

$$\hat{\theta}_{v,2} = 0.014732,$$

$$\hat{\theta}_{nv,1} = 0.001494,$$

$$\hat{\theta}_{nv,2} = 0.015926,$$

$$\hat{\delta} = 0.042974,$$

$$\hat{p}_{1,1} = 0.419749,$$

$$\hat{p}_{1,2} = 0.335583,$$

$$\hat{p}_{2,1} = 0.077643,$$

$$\hat{p}_{2,2} = 0.167025,$$

$$\hat{p} = 0.577104.$$

Here $p_{1,2}$ is the probability that an offender will choose the first (lower) value for violent imprisonment rate and the second (higher) value of non-violent rate. The other $p_{i,j}$ are defined similarly. One can easily calculate that the estimated probability that an offender chooses both rates low or both rates high is larger than one would get by assuming independence with the same marginal probabilities. The log-likelihood for this model is -71242, which is 919 larger than the model in Section 5.5.1 with 5 extra parameters. It can also be compared to the model of Section 5.4.1, which had a log-likelihood of -64070 - 7537 = -71607. The improvement is 365, with 4 extra parameters.

5.5.3. Differing Imprisonment Lengths

It was evident in Table 8 that the times spent in prison for different offenses varied somewhat by offense. In particular, violent crimes (especially murder and robbery) led to longer imprisonment lengths than the other types. We considered extending the model of Section 5.5.2 to allow the distribution of imprisonment lengths for violent crimes be exponential with mean $1/\delta_v$ and the imprisonment lengths for non-violent crimes to be exponential with mean $1/\delta_{nv}$.

The major difficulty in fitting the model just described is that the probability of an offender being in prison at the time of the survey is very difficult to obtain in closed form. As a first approximation, we used formula (5.3), with $\delta = 2/\{(1/\delta_v)+(1/\delta_{nv})\}$, which appears to provide a good approximation to the true probability of being in prison for moderately disparate values of δ_v and δ_{nv}. After finding the approximate MLE's, we then calculated the exact probability of being in prison and recalculated the log-likelihood at the approximate MLE's, so that we could compare it to the other models.

The estimates under this model are

$$\hat{\theta}_{v,1} = 0.003149,$$
$$\hat{\theta}_{v,2} = 0.013907,$$
$$\hat{\theta}_{nv,1} = 0.001555,$$
$$\hat{\theta}_{nv,2} = 0.016249,$$
$$\hat{\delta}_{v} = 0.033053,$$
$$\hat{\delta}_{nv} = 0.060450,$$
$$\hat{p}_{1,1} = 0.409300,$$
$$\hat{p}_{1,2} = 0.323485,$$
$$\hat{p}_{2,1} = 0.103216,$$
$$\hat{p}_{2,2} = 0.164000,$$
$$\hat{p} = 0.579683.$$

The log-likelihood is -70334, which is 908 higher than the same model with only one value of δ. Notice that the estimated average time spent in prison for violent crimes is $1/\hat{\delta}_v$ = 30.3 months, while the average time spent for non-violent crimes is estimated to be $1/\hat{\delta}_{nv}$ = 16.5 months.

5.6. Models With Late Starters

In Table 6, we saw that the time until the first imprisonment was generally longer than the times between later imprisonments. This suggests that a model for incarceration careers should include a provision for the first street time to have a different distribution than the later street times. The simplest way to do this is to assume that the imprisonment career does not start until after some initial waiting time. In this section, we consider two models with this property.

5.6.1. Everyone Starts Late

First, we assume that all offenders wait an amount of time before embarking on their imprisonment careers. We assume that the waiting time has an exponential distribution with mean $1/\lambda$. This makes the time until the first imprisonment have a distribution which is the convolution of two exponential distributions, one with mean $1/\lambda$ and the other with mean $1/(\theta_v+\theta_{nv})$. The contribution to the likelihood function for that first imprisonment for subject i, $x_{i,1}$ is

$$\lambda\theta_t\{\exp(-\theta x_{i,1})-\exp(-\lambda x_{i,1})\}/(\lambda-\theta),$$

where $\theta = \theta_v+\theta_{nv}$ and t = v or nv depending on whether the first imprisonment is for a violent or non-violent offense.

As in Section 5.5.3, the probability of being in prison at the time of the survey is difficult to calculate, so we will find approximate MLE's and then only calculate the probability at the values of the approximate MLE's. The MLE's for this model are given by

$$\hat{\theta}_{v,1} = 0.003093,$$

$$\hat{\theta}_{v,2} = 0.016435,$$

$$\hat{\theta}_{nv,1} = 0.001641,$$

$$\hat{\theta}_{nv,2} = 0.017914,$$

$$\hat{\delta}_v = 0.032867,$$

$$\hat{\delta}_{nv} = 0.059915,$$

$$\hat{p}_{1,1} = 0.42405,$$

$$\hat{p}_{1,2} = 0.29142,$$

$$\hat{p}_{2,1} = 0.08385,$$

$$\hat{p}_{2,2} = 0.20068,$$

$$\hat{p} = 0.55274,$$

$$\hat{\lambda} = 0.17862.$$

The log-likelihood is -70285, which is 49 higher than the model of Section 5.5.3 with one extra parameter. This is a somewhat smaller improvement than one would expect in light of Table 6. We believe that this may be due to the fact that not all offenders have longer first street time than later ones. Hence, we should not require that all offenders in the survey start their careers late. This assumption is modified in the next section.

5.6.2. Not Everyone Starts Late

Table 6 suggests that not all offenders wait an extended time for their first imprisonments. Those with juvenile records seem to wait only as long for their first imprisonment as those without a juvenile imprisonment wait for their later ones. For this reason, we introduce one further parameter p_L to stand for the probability that an offender is a late starter. With probability p_L the likelihood function is as described in Section 5.6.1. With probability $(1-p_L)$, the likelihood is as in Section 5.5.3. For this model, the MLE's are

$$\hat{\theta}_{v,1} = 0.003482,$$

$$\hat{\theta}_{v,2} = 0.021449,$$

$$\hat{\theta}_{nv,1} = 0.001812,$$

$$\hat{\theta}_{nv,2} = 0.021696,$$

$$\hat{\delta}_{v} = 0.033187,$$

$$\hat{\delta}_{nv} = 0.060744,$$

$$\hat{p}_{1,1} = 0.362441,$$

$$\hat{p}_{1,2} = 0.308661,$$

$$\hat{p}_{2,1} = 0.097110,$$

$$\hat{p}_{2,2} = 0.231788,$$

$$\hat{p} = 0.555675,$$

$$\hat{\lambda} = 0.022333,$$

$$\hat{p}_{L} = 0.579451.$$

The log-likelihood is -69942, which is 343 higher than the model which had all offenders starting late (Section 5.6.1.) Notice that the estimated average time until start of career, for those who start late, is $1/\hat{\lambda} = 44.78$ months compared to 5.60 months in the previous model.

5.6.3. A Model Distinguishing First Imprisonment Lengths

For our final stage of model fitting, we fit a model which allows the length of the first imprisonment to have a different distribution from the later imprisonment lengths. Such a model was suggested both by Table 9 and by Table 10. We introduced a distinction between first and later imprisonment lengths by saying that the first imprisonment has exponential distribution with mean $1/\delta_{0v}$, if it is for a violent offense, or mean $1/\delta_{0nv}$, if it is for a non-violent offense. This is far more convenient mathematically than distinguishing the first imprisonment for each offense type separately.

The contribution to the likelihood function for the first imprisonment length $s_{i,1}$ of offender i is a factor of

- $\delta_{0v}\exp\{-\delta_{0v}s_{i,1}\}$, if violent and not the only imprisonment,
- $\delta_{0nv}\exp\{-\delta_{0nv}s_{i,1}\}$, if non-violent and not the only imprisonment,
- $\exp\{-\delta_{0v}s_{i,1}\}$, if violent and the only imprisonment,
- $\exp\{-\delta_{0nv}s_{i,1}\}$, if non-violent and the only imprisonment.

The MLE's for this model are:

$$\hat{\theta}_{v,1} = 0.003575,$$

$$\hat{\theta}_{v,2} = 0.019169,$$

$$\hat{\theta}_{nv,1} = 0.001533,$$

$$\hat{\theta}_{nv,2} = 0.021291,$$

$$\hat{\delta}_{v} = 0.039549,$$

$$\hat{\delta}_{nv} = 0.060423,$$

$$\hat{\delta}_{0v} = 0.029355,$$

$$\hat{\delta}_{0nv} = 0.060246,$$

$$\hat{p}_{1,1} = 0.310992,$$

$$\hat{p}_{1,2} = 0.301169,$$

$$\hat{p}_{2,1} = 0.102791,$$

$$\hat{p}_{2,2} = 0.285048$$

$$\hat{p} = 0.568655,$$

$$\hat{\lambda} = 0.020295,$$

$$\hat{p}_{L} = 0.610366.$$

The log-likelihood is -69249, which is 693 higher than for the model of Section 5.6.2 with two extra parameters.

There is something odd, however, about the estimates for this model. The estimated length of the first imprisonment is $1/\hat{\delta}_{0v}$ = 34.07 months for violent offenses and $1/\hat{\delta}_{0nv}$ = 16.60 months for non-violent offenses. On the other hand, the estimated length of later imprisonments is $1/\hat{\delta}_{v}$ = 25.29 months for violent offenses and $\hat{\delta}_{nv}$ = 16.55 months for non-violent offenses. This is counter to what is suggested in Tables 9 and 10. One reason may be that, although Tables 9 and 10 suggest that first imprisonments are shorter than later ones, those tables do not include the current imprisonment. There are many offenders who were in prison for the first time at the time of the survey with rather long imprisonments. Tables 9 and 10 may be misleading because they only compare first and later imprisonments for those offenders who managed to finish their first imprisonment already. For example, there are 447 inmates in Table 10 whose first adult imprisonment was for a violent offense, and who completed that first imprisonment. The average length of those imprisonments was 25.57 months. Not included in Table 10 are 2667 inmates who are currently serving their first and only imprisonment for a violent offense. The average of the elapsed times in those imprisonments is 30.63 months. The distinction between current first imprisonments and completed first imprisonments for non-violent offenses is not nearly as dramatic.

Just to make sure that the addition of eleven new parameters to the model of Section 5.3 did not destroy the nice fit of the empirical age distributions seen in Figure 13, we used the model of this Section to predict the empirical age distribution, and the result is in Figure LASTAGECOMPARE. As one can see, this model also predicts an age distribution very much like what was actually observed.

5.7. Conclusions From the Models Fit

The conclusions from fitting the above models can be summarized as follows:

- It seems apparent that there is heterogeneity in the offender population with regard to length of time spent on the street between imprisonments. The mean and standard deviation of the fitted distribution of imprisonment rates are 0.022738 and 0.01146 respectively.

- There appears to be a difference between the rates at which offenders are imprisoned for different types of offenses.

- Many offenders (estimated to be about 61%) start their imprisonment careers later than the rest. The estimated amount of time is 49.3 months later.

- There is evidence that offenders drop out of their imprisonment careers at with moderately high probability, accounting for the over representation of young offenders in prisons.

- First imprisonments for violent offenses are estimated to be somewhat longer than later imprisonments for violent offenses (by 8.8 months), while there is essentially no difference between the estimated lengths of first and later imprisonments for non-violent offenses.

One must be careful not to over interpret the estimates in the fitted models. For example, the fact that the hierarchical model with two different imprisonment

Figure 14: Comparison of Observed and Predicted Age Distributions for Final Model

rates (Section 5.4.1) fits much better than a model with only one imprisonment rate does not mean that there actually are two groups of offenders, some who are imprisoned at a high rate and others who are imprisoned at a low rate. Rather, it suggests that there is indeed heterogeneity in the population of offenders, which is better modeled by imagining two groups of offenders, whose street times are more similar to each other than to those in the other group, than by assuming the variation in street time is large but all offenders vary the same. It will take a new survey which is better constructed to be able to sort out the heterogeneity in a more exact manner.

Similarly, the estimated 61% of the offenders who may have started their careers late by an estimated 49.3 months did not all start 49.3 months later. Rather, we fit an exponential distribution to the late start time which had a mean of 49.3 months, but was spread over all possible late start times from 0 months to indefinitely long.

One must also be careful when interpreting the estimated imprisonment rates. These will be estimated rates at the beginning of an offender's career (either age 18 or start of career for late starters). Given that an offender has survived to an older age in his/her career, the estimated imprisonment rate for that individual will be lower. This is due in part to the feature of the model which lets high rate offenders drop out of their careers early.

6. Further Analyses to be Done

6.1. Problems With The Data

There are some problems which we have not yet dealt with in this data set, but which deserve further attention. First there is the problem of inmate recall. This has manifested itself already in that some inmates report a certain number of imprisonments, but then describe a different number. Even though the set of such inmates is itself large (4,364) these are only the ones we *know* have had recall problems. There are the thousands of others who reported apparently complete histories, but who also did not recall completely. It will be a serious modelling task to introduce the uncertainty due to recall problems. A second problem is the voluntary nature of participation in the survey. It would also be difficult to model the relationship between criminal career parameters and willingness to participate in the survey.

A third, minor problem with the data, is the fact that occasionally an interviewer did not ask the survey questions in the correct order. This led to certain inmate records being difficult to extract correct information from. Finally, there is the problem which arises from possible falsification of information by the offender. Dispite the best assurances of confidentiality, not every inmate is likely to believe an interviewer who says that no further prosecutions will result from any answers they give.

6.2. Sampling Plan

In none of our analyses, did we make explicit use of the sampling plan which was used to collect the data set. In doing a likelihood based analysis, as we have done, one need not consider the sampling plan so long as one explicitly models the joint distribution of all relevant variables in the population of interest and models the probability of selecting each particular subject. We have modelled the population of offenders as a prior exchangeable in the sense that we have no information to distinguish one from the next until we begin the interview process. We have explicitly modelled the probability of selecting each inmate via the bias correction mechanism described in Section 5.2

6.3. Other Covariates

Although our exploratory analysis indicated that race did not make as large a difference in either street times or imprisonment times as did the other factors we considered, race does seem to make a large difference in a person's initiation into the offender population. That is, there are far more blacks in the prisons than one would predict based simply on their proportion in the population at large. Although this phenomenon is similar to that of the age distributions, which also differed dramatically, there is no mechanism such as dropout which will correct this imbalance in the model. The only way to account for the difference would be to introduce differential initiation rates for blacks and nonblacks. This should be done in future models for this data.

Another intuitively plausible covariate which we failed to examine is drug use. In future analyses, we would hope to explore the relations between drug use and both street and imprisonment times as well as any possible relation to choice of crime type. The same analyses should be done for sex, although the inmate populations are predominately male. Once again, differential initiation rates for males and females would be needed to account for the difference between prison and population distributions.

Acknowlegments

The authors would like to thank Alice Kraycar for diligent typing and C. E. Kim for dataset management. They would also like to thank Mr. Kim and C. W. Ahn for some of the theoretical calculations. Special thanks go to Professor A. Blumstein for selflessly offering insight and suggestions at several times during the course of this research.

References

Deely, J. J. and Lindley, D. V. "Bayes empirical Bayes," Journal of American Statistical Association, 76, 833-841, 1981.

Dempster, A. P., Rubin, D. B. and Tsutakawa, R. W., "Estimation in covariance components models," Journal American Statistical Association, 76, 341-353, 1981.

Karlin, S. and Taylor, H. M., A First Course in Stochastic Processes, 2nd ed., New York, Academic Press, 1975.

Morris, C. N., "Parametric empirical Bayes inference: Theory and applications," Journal American Statistical Association, 78, 47-65, 1983.

Rolph, J. E. Chaiken, J. M. and Houchens, R. E., "Methods for estimating crime rates of individuals," Rand Corporation Report R-2730-NIJ, 1981.

Survey of Inmates of State Correctional Facilites, Bureau of Justice Statistics, U.S. Department of Justice, 1979.