Methods for Estimating Crime Rates of Individuals

John E. Rolph, Jan M. Chaiken, and Robert L. Houchens

Rand

Prepared under Grant No. 78-NI-AX-0129 from the National Institute of Justice, U.S. Department of Justice. Points of view or opinions stated in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Library of Congress Cataloging in Publication Data

Rolph, John E.
Methods for estimating crime rates for individuals.

"R-2730-NIJ."
Bibliography: p.
1. Criminal behavior, Prediction of. 2. Crime forecasting. I. Chaiken, Jan M. II. Houchens, Robert, 1952- . III. National Institute of Justice (U.S.) IV. Rand Corporation. V. Title.
HV6030.R64 364.2 81-5858
ISBN 0-8330-0317-8 AACR2

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

R-2730-NIJ

Methods for Estimating Crime Rates of Individuals

John E. Rolph, Jan M. Chaiken, and Robert L. Houchens

March 1981

Prepared under a grant from the National Institute of Justice, U.S. Department of Justice

Rand
SANTA MONICA, CA. 90406

PREFACE AND SUMMARY

This work was performed under grant number 78-NI-AX-0129 from the National Institute of Justice, one of several grants awarded for research in criminal justice evaluation methodologies. Methods are developed for analyzing individuals' crime commission data and for drawing conclusions about (1) those individuals' crime commission rates and (2) the distributions of crime commission rates for groups of offenders with specified characteristics. The methods are illustrated using results from the Rand Second Inmate Survey, one of several other related Rand projects concerned with policy-oriented research on criminal careers.

Typical survey responses about numbers of crimes committed are uncertain within ranges and can be treated as censored observations. We use a method by Turnbull to obtain a nonparametric maximum-likelihood estimate for the empirical distribution of observed crime commission rates, including both censored and uncensored observations.

We then examine these empirical distributions of crime rates to determine whether they approximately match standard distributional forms--lognormal, Pareto, or gamma. No single form is satisfactory for all crime types considered in this study. Some types are fit by excluding a portion of the respondents reporting zero crime commissions. Counts of crimes committed--as distinguished from rates--are tested to see whether they follow a negative binomial distribution. Multivariate generalizations are also developed and explored.

The commonly made assumption that each offender's crime commission propensities λ_1, λ_2, ..., λ_K (for K crime types) correspond to K Poisson processes is not consistent with the data. Certain types of crimes appear to be committed according to Poisson processes; others appear to occur according to spurting behavior.

A procedure by Hudson and Tsui is adapted to estimate each individual's crime commission propensity from information about the number of crimes he committed during a measurement period. Offenders are first divided into groups according to their commission or noncommission of

a particular crime prior to the measurement period. A shrinkage estimator of an offender's propensity for committing that crime is then obtained by shrinking his observed crime rate toward a value that is determined by regression for his group. The independent variables in the regression are crime commission covariates such as age, use of drugs, and employment.

Populations of offenders who can be surveyed about their crime commissions (e.g., prisoners) tend to be unrepresentative of target populations of primary interest. We therefore develop models for stochastic processes relating target populations to survey populations. These models yield estimates of sampling probabilities for members of the survey population, allowing estimation of crime commission rate distributions in target populations.

A complete summary of this study is presented in a companion report, *Methods for Estimating Crime Rates of Individuals: Executive Summary*, R-2730/1-NIJ, March 1981.

ACKNOWLEDGMENTS

This work was encouraged and supported by George Silberman and Richard Linster, of the National Institute of Justice. We are indebted to Rand staff members Suzanne Polich, who calculated the crime rate data used for illustrative examples in this study, and Marcia Chaiken, who provided data for examples of potential covariates of crime rates. Naihua Duan and Michael Maltz reviewed an earlier draft of this report with considerable dedication and made many helpful suggestions.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION

CONTEXT AND OBJECTIVES OF THE STUDY

In the criminal justice system, the processing of an individual suspected of committing a crime or convicted of a particular crime involves many decisions based on beliefs or experiences that relate to the degree and seriousness of the offender's criminal behavior. These decisions affect the disposition of the individual's case and the length of time--if any--he or she will be incarcerated after conviction. Differential handling of offenders can be found in many areas of the system, as the following examples illustrate:

- Law enforcement. Some large police departments and sheriffs' departments operate "major offenders units" to keep track of certain known habitual offenders and to arrest them when they commit a crime.* Because of the cost of tracking, only individuals the police anticipate will commit numerous serious crimes are covered by such units.
- Prosecution. Some district attorneys operate "career criminal prosecution units," which bring special resources to bear to increase the probability of conviction of certain individuals when they are arrested, and imprisonment when they are convicted (Bronx County, 1976; Dahmann and Lacy, 1977; INSLAW, 1977).
- Sentencing. In deciding whether a convicted person should be given probation or be incarcerated, judges often consider whether the conviction crime appears to be an isolated incident or part of a pattern of substantial criminal activity. And in establishing the length of sentence to be imposed for

---

*Pate et al. (1976) describe the Perpetrator-Oriented Patrol in Kansas City, Missouri, and two predecessor projects in Miami, Florida, and Wilmington, Delaware. Additional police programs focusing on major offenders were instituted in Rochester, New York; Amarillo, Texas; Pueblo, Colorado; and Norfolk, Virginia.

particular crimes and combinations of crimes, especially when prior criminal record is taken into account, legislators may be influenced by their own beliefs about the kinds of records that indicate a person is a high-rate offender.

- Parole. When parole boards decide whether a prisoner should be released, they consider the chances of "parole success," which essentially means the probability that the individual will commit a crime or violate a condition of parole during a specified future time period (Gottfredson et al., 1978; Hoffman and DeGostin, 1974).

Although decisions based at least in part on individuals' crime commission rates are made daily within the criminal justice system, and the people who make these distinctions may feel quite confident of the correctness of their decisions because of extensive personal experience, research shows that it is exceptionally difficult to predict accurately who *will be* a high-rate criminal offender, or even to determine from personal descriptors and criminal records who *has been* a high-rate offender during a specified period of time in the past. Nonetheless, research also shows that most people who commit crimes commit only a relatively small number, while a few people commit crimes at substantially higher rates (Wolfgang et al., 1972; Peterson et al., 1980; Greene, 1977). In short, very-high-rate offenders do exist, but it is not easy to identify them.

This report describes methods that can be used to analyze crime commission rates and thereby shed light on the problems of distinguishing between low-rate and high-rate offenders. Analytically, the problem divides into two general categories of questions:

1. Given the best possible information about an offender and his criminal behavior during a previous period, how can we estimate his crime commission rates during that period?
2. For a group of criminal offenders with specified characteristics, how can we determine the average rate of committing various crimes, the distribution of those crime commission rates, and the extent to which the rates of the specified groups differ from those of another group?

The second category of questions avoids the problem of identifying particular individuals as high-rate offenders by focusing instead on aggregate behavior. Methods for answering questions of this type can be useful for determining whether rules currently being used by police, prosecutors, parole boards, etc., actually distinguish the intended target group of high-rate offenders from others. The methods described in this study can also help in devising better decision rules for identifying offenders who should receive special attention or punishment, although other considerations certainly play a role (e.g., the equity of treatment of similar persons in similar circumstances, the feasibility of obtaining the necessary information for the decision rule in a timely fashion, and "just deserts" as applied to the conviction crime (von Hirsch, 1976; Morris, 1974)).

In addition, our methods can be used to analyze groups of offenders defined by characteristics that are presumably unrelated to their crime propensities, such as the city or state in which they reside. By enabling investigators to determine whether a city with relatively low per capita crime rates has (a) relatively fewer criminals than other cities or (b) lower crime commission rates among those who are offenders (or both), these methods can help in evaluating the effectiveness of various governmental anticrime activities and the deterrent effects of city- and state-level sanction policies.

Other Rand studies are examining prediction of high-rate offenders, decision rules for selective handling of offenders, and deterrence (Greenwood, 1980). To carry out that research, self-report data were collected from incarcerated offenders about the crimes they had committed during specified periods of time, and both self-report and official data were collected about offenders' characteristics that presumably relate to criminal behavior (Peterson, Chaiken, and Ebener, forthcoming). The analyses of these data yield estimates of the crime commission rates of the surveyed offenders, relate these rates to their personal characteristics, and extrapolate the results to more general populations of offenders.

The methods described in this report can assist in policy research and analyses that use any source of data concerning the criminal activity of individuals. Our illustrative examples, drawn from Rand inmate survey data, do not answer the major analytical questions related to that survey, but rather show how our methods can be applied in practice.

METHODOLOGICAL ISSUES ADDRESSED

Our approach is based on the following general model of criminal behavior:* There are K types of crimes of interest, and each criminal offender commits each of the crimes at a specified rate (possibly zero) when free to do so. Thus, offender i has associated with him a vector $(\lambda_1(i), \lambda_2(i), \ldots, \lambda_K(i))$ giving the expected number of crimes of type k he commits per year of "street time," $\lambda_k(i)$, $k = 1, 2, \ldots, K$. We refer to these parameters as "[underlying] crime commission *propensities*" to distinguish them from the various estimated and reported commission rates that we will use in the analysis.

When incarcerated, the offender is assumed not to commit crimes at all, which essentially means that crimes committed in jail or prison are not considered in this model. Naturally, the parameters $\lambda_1(i), \ldots, \lambda_K(i)$ may change with the passage of time, but we assume that they are constant over reasonably short periods--1 or 2 years--that are not interrupted by major events (e.g., imprisonment). Eventually the individual stops committing crimes or dies.

We assume further that for a selected group of offenders, it is possible to determine the number of each of the K crimes that each offender committed during a specified period of time, called the "measurement period." This assumption reflects the conjunction between our work and the other Rand studies noted above. While data obtained from those studies are used in our illustrative examples, the assumptions and methods are not tailored to these particular data and should apply equally to other methods of collecting the required information.

*Details and variations of the model are given later.

Given the data as described, we address the following issues in this report:

- Characterization of the distributions of observed crime commission rates for the selected offenders who provided data, fitting the distributions by parametric forms, and summarizing them by meaningful statistics. We are concerned with both the univariate distribution of commissions for each crime type and the multivariate distribution for all K crime types together.
- Estimation of the crime commission propensities for any *one* of these individuals, taking into account the group's overall distribution of commission rates as well as *the individual's* reported crime commissions and other characteristics.
- Estimation of the distribution of crime commission propensities for more general populations of criminal offenders who differ in known ways from the selected offenders who provided data.

A COMPOUND GAMMA-POISSON MODEL

We can illustrate the interpretation of these methodological issues by a simple analytic example. This example is not intended to be correct in all respects--indeed, the discussions that follow challenge the example in nearly every particular. It was chosen, rather, for its simplicity and because it serves as a base of comparison from which the analysis that follows can show what is *not* a reasonable assumption.

In this example, each individual commits only one type of crime (or, equivalently, crimes are not separated according to type). Offender i is assumed to commit crimes according to a Poisson process with rate $\lambda(i)$, and the number of crimes Y_i that he commits in a period of time of length T is measured. The assumption of a Poisson process means that Y_i has a Poisson distribution with parameter $\lambda(i)T$:

$$P(Y_i = y \mid \lambda(i), T) = \frac{(\lambda(i)T)^y}{y!} \exp(-\lambda(i)T), \qquad y = 0, 1, 2, \ldots \quad (1.1)$$

It should be noted that if $\lambda(i)$ is nonzero, there is still a chance that offender i does not commit any crimes during the measurement period:

$$P(Y_i = 0 | \lambda(i), T) = \exp(-\lambda(i)T).$$

To incorporate into the model the fact that different offenders have different crime commission propensities, we assume that the propensity parameters $\lambda(i)$ are sampled from a gamma distribution with shape parameter α and scale parameter β. Then the probability density function for each $\lambda(i)$ is

$$f_{\alpha,\beta}(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda) \quad \text{for } \lambda > 0, \tag{1.2}$$

where the gamma function Γ is defined by

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u}\, du.$$

The mean of the gamma distribution of crime commission propensities is α/β and its variance is α/β^2.

Even though the underlying crime commission propensities are drawn from a gamma distribution, the distribution of *observed* crime rates Y_1/T, Y_2/T, ..., is not. The gamma distribution is continuous, while the observed crime rates can obviously take on only the discrete values 0, 1/T, 2/T, 3/T, ... In fact, the compounding of the gamma distribution for $\lambda(i)$ and the Poisson distribution for $Y_i|\lambda(i),T$ can be easily shown (Johnson and Kotz, 1972, Chap. 5) to yield a negative binomial distribution for Y_i:

$$P(Y_i = y|T) = \binom{\alpha + y - 1}{y} \left(\frac{T}{T+\beta}\right)^y \left(\frac{\beta}{T+\beta}\right)^\alpha, \qquad y = 0, 1, 2, \ldots .$$

The parameters of this negative binomial are α, $(\beta/(T + \beta))$; its mean is $\alpha T/\beta$; and its variance is $(\alpha T(T + \beta))/\beta^2$.

If the model is correct, data giving the number of crimes Y_i committed by a random sample of offenders during time period T would be fit well by a negative binomial distribution. Then the parameters for the underlying gamma distribution are the α and β that belong to the estimated negative binomial distribution.

If a subsample of offenders with known characteristics, e.g., those that committed at least one crime in the measurement period ($Y_i > 0$), the parameters that best characterize the resulting truncated negative binomial distribution could be estimated, and the parameters of the underlying gamma distribution would be deduced from that binomial distribution.

Consider now a randomly selected offender who has committed Y_i crimes during a measurement period of duration T. Given that the crime propensities λ are assumed *a priori* to have a gamma distribution with parameters α, β, the relevant information for estimating *this* offender's $\lambda(i)$ is the *posterior* distribution of $\lambda(i)$ given Y_i and T. A standard calculation (De Groot, 1970, Chap. 9) shows this distribution to be gamma with parameters $(\alpha + Y_i)$, $(\beta + T)$. Hence the expected value of $\lambda(i)$, given Y_i and T, is the mean of this distribution, or

$$E(\lambda(i)|Y_i, T) = \frac{\alpha + Y_i}{\beta + T}.$$

This is the Bayes estimator for the crime commission propensity of offender i. It differs from both the ordinary estimator Y_i/T and the *a priori* estimator α/β, and in fact it is a weighted average of the two:

$$\frac{\alpha + Y_i}{\beta + T} = (1 - w)\frac{Y_i}{T} + w\frac{\alpha}{\beta},$$

where $w = \beta/(\beta + T)$. It is the "best" estimate of λ_i, given the data for offender i and the assumptions of the model.

This example has illustrated the distinction between the observed crime commission rate and the underlying crime commission propensity, the possibility of estimating the distribution of crime commission propensities even if the sample of offenders is not representative of the entire offender population, and the possibility of using an estimate of $\lambda(i)$ other than Y_i/T. These notions appear repeatedly in the following sections.

II. UNIVARIATE DISTRIBUTIONS OF OBSERVED CRIME COMMISSIONS

This section presents a method for characterizing the distribution of observed crime commission rates and estimating parameters of the distribution. The task is not straightforward because of problems inherent in obtaining data of the type required. We begin, then, with an overview of the nature of those problems as they would occur in almost any framework for collecting data. Next we describe the idiosyncratic difficulties associated with the particular survey instrument used in the Rand Second Inmate Survey (from which we obtained the data used to illustrate the methods in this report). We then give a nonparametric method for representing the data, along with parametric distributional forms. Finally, we examine the correctness of the Poisson assumption that is used in some of the models.

THE FORM OF CRIME COMMISSION DATA

In order to focus on issues related to the *form* of the data themselves, rather than on problems of inaccurate data, we envision a nearly perfect method of data collection. Suppose that 100 invisible and undetectable observers were assigned to follow 100 presumed criminal offenders for a year, one observer for each offender. Suppose further that each observer recorded all criminal offenses committed by the offender to whom he was assigned, listing each of the offenses in one of K categories. For specificity, we assume K = 8. After the data were assembled at the end of the year, the following issues would arise in the analysis.

Too Many Zeros

The analysts might be interested in one of the eight crime types, say, motor vehicle theft (or auto theft for short). Upon examining the data, they would immediately find that a large number of the offenders had not committed any auto thefts during the year. These people would belong to one of three groups:

1. People who appeared to have experience in stealing cars, contemplated doing so from time to time, but never got around to it during the measurement year. Perhaps if the measurement period had started on a different date or had lasted longer than a year, these individuals would have been observed stealing a car. In terms of our simplified model, they are people whose λ for auto theft is nonzero but their Y (number of auto thefts committed) happened to be zero.
2. People who never stole a car (or perhaps stole cars at one time in the past, but not recently), didn't contemplate stealing cars during the measurement year, and/or didn't appear to know how to steal cars. In terms of the model, these are people whose λ for auto theft is zero. Perhaps they were chosen for observation as part of the sample 100 because they commit assault or some other crime, but their behavior in regard to auto theft is essentially irrelevant to an understanding of auto thieves.
3. People who were in prison, jail, or a hospital for all of the measurement year (or nearly all) and therefore did not have a genuine opportunity to steal a car. They may have committed various other crimes while in custody, but not auto theft. Basically, the fact that their Y for auto theft was zero is not informative. Some of them may be very active auto thieves when free; others, not.

In short, the fraction of the sample who committed no auto thefts in the measurement year may not logically be expected to bear any particular relationship to the fractions that committed 1, 2, 3, ... auto thefts. In fact, if the sample of 100 is stratified in any way (e.g., by the type of crime for which the offender was previously arrested), the stratification design can obviously affect the number of zero counts for auto theft. On the other hand, it is not conceptually correct simply to ignore the zeros. In particular, we cannot define a

member of the sample as an "auto thief" if he stole one or more cars during the measurement year, and then attempt to determine the distribution of propensities for auto theft by treating the data as if there were no other auto thieves in the sample.

Having identified the issue of "too many zeros," we can analyze the data in various ways that take into account the fact that the offenders with Y = 0 for a particular crime are an unknown mixture of offenders with $\lambda = 0$ and offenders with $\lambda > 0$. We shall try several of these approaches and show empirically which of them seem most useful. It will be seen that ignoring the problem--i.e., assuming that all offenders have a nonzero propensity (perhaps very small) for every crime type--is not a practical approach.

Multiple Crimes in a Single Incident

Consider the clever auto thief who has devised a method for diverting a railroad car containing 20 automobiles, removing the automobiles, and returning the empty railroad car to the train. Suppose he does this ten times during the measurement year. Depending on the observer's instructions, the thief will be recorded with a count of either ten auto thefts (Y = 10) or 200 auto thefts (Y = 200). The models used to analyze the data then have to match the definition of "crime" followed by the observers. If the definition results in a count of Y = 200 in the above example, then evidently the Poisson assumption in Sec. I is inappropriate.

The situation is more analytically troublesome in the case of a single incident involving several different types of crimes. For example, an offender steals a car, robs a liquor store, and shoots the cashier. Standard police practice would be to record the crime once, under the most serious category--in this case, commercial robbery. However, such a recording method may easily obscure the fact that this offender did commit an auto theft during the measurement year. Both the distribution of observed crime commission rates for a single crime type and the observed structure of covariance among the commission rates for different crime types will be affected by the recording method.

There is no general solution to this problem. We must simply be alert to the ambiguity and interpret the data correctly in light of the definitions used.

Nobody is Average

Typically, data of this type indicate that among those who did a given type of crime during the year, the majority did very few (say, 1, 2, or 3 crimes). At the high end of the distribution are a small number of people who committed the crime at very high rates: In a sample of 100 offenders, one or two may commit more than 200 auto thefts in a year.

Figure 2.1 illustrates this phenomenon for a subsample of prisoners reflected in the Rand Second Inmate Survey. More than half of this sample of prisoners who committed any of the listed crimes did fewer than 20 crimes per year.*

It can be seen that the commission rates do not tend to cluster around the mean; in fact, nearly all rates are either near zero or over 300. (This observation is succinctly, if somewhat imprecisely, summarized by the heading, "Nobody is Average.") Hence the mean is not a satisfactory descriptor of the distribution. Moreover, the estimate of the mean is highly sensitive to the observed crime rates for the few offenders in the high tail of the distribution. Excluding these high values as outliers in the analysis is not satisfactory, because the people with high crime commission propensities are the very offenders in whom the greatest policy interest resides.

The analysis, then, has to use robust methods that are not unduly sensitive to the exact estimated crime rates for high-rate offenders but, on the other hand, do not ignore them.

SURVEY INSTRUMENTS

The respondents to the Rand Second Inmate Survey (which is described in Peterson, Chaiken, and Ebener (forthcoming)) were male jail

---

*The wording of the survey questions for the crimes included in Fig. 2.1 is explained below.

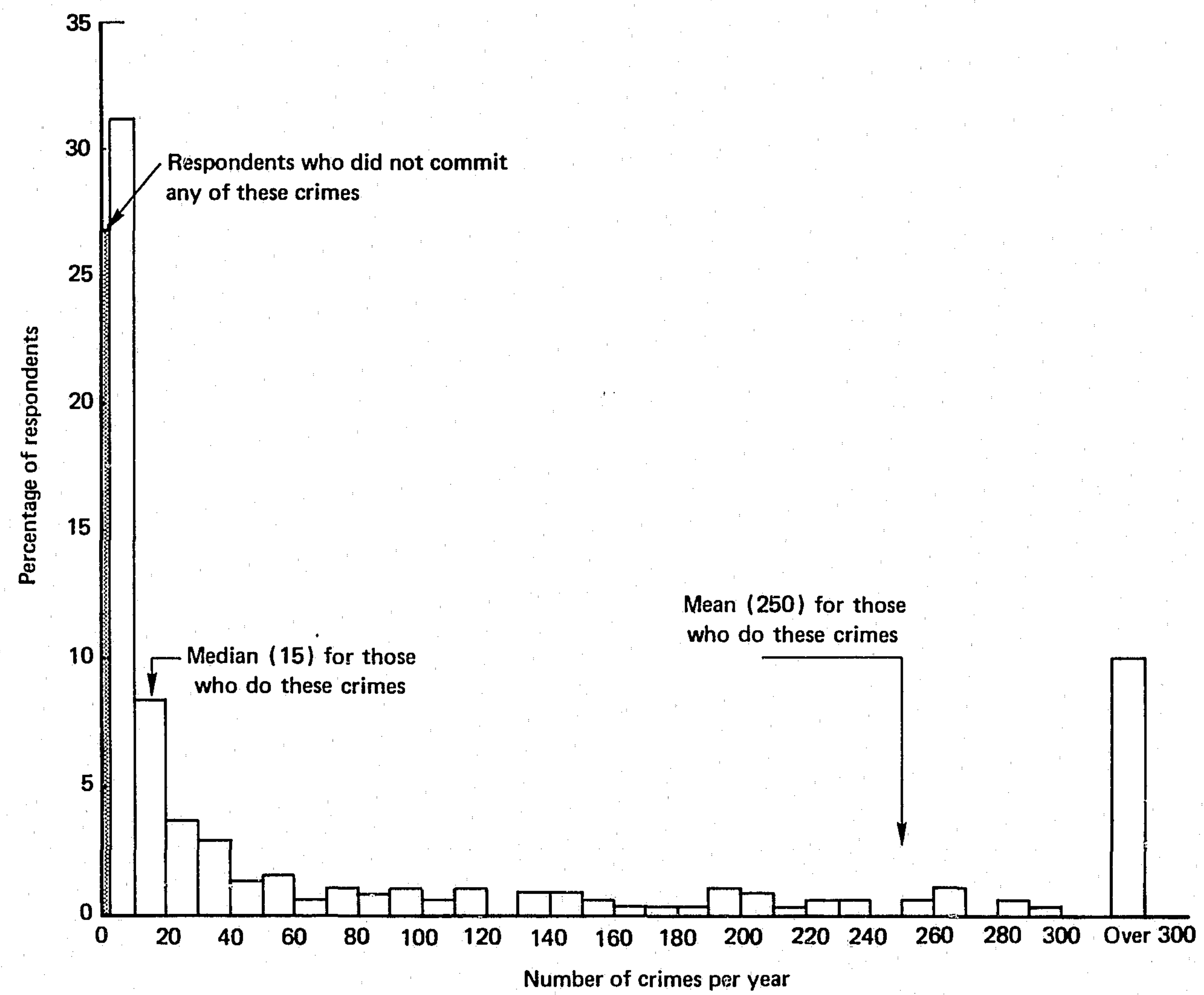

Fig. 2.1 — Observed crime commission rate for burglary, robbery, assault, theft, auto theft, forgery, and fraud

and prison inmates from selected counties in California, Michigan, and Texas. However, only the Michigan prisoner respondents were included in the analysis described in the present report. Since we are concerned with methods of data analysis and not with the implications of these particular data, we shall not describe the sample design, field operations, or response rates. We shall examine only those aspects of the survey instrument format that are relevant for understanding the data analysis.

The questionnaire asked a series of questions about crimes committed by the respondent during a measurement period of from 13 to 24 calendar months. The period ended with the (earliest) month in which the respondent was arrested for the crime he was then serving time for, and it began with January of the preceding year. The respondent used a calendar printed on a separate card and followed a series of instructions in the survey booklet to calculate the number of "street months" in his measurement period, i.e., the number of months he was not incarcerated. The number of street months varied from 1 to 24. Some respondents made obvious calculational errors and/or changed their minds about the number of street months after answering a few questions, so the data we received and processed included a minimum and a maximum estimate of each respondent's number of street months. If the number of street months was missing in the survey booklet, it was assigned a minimum value of 1 and a maximum of 24.

Ten types of crimes were included in the survey, but two of them (related to assault and murder) followed a different format from the other eight and are not included in the present analysis. The first question for each of the eight crime types asked whether or not the respondent did that crime during the street months. The eight crime types and the wording of the opening questions were as follows:

> *Crime 1: Burglary.* During the STREET MONTHS ON THE CALENDAR did you do any burglaries? (Count any time that you broke into a house or a car or a business in order to take something.)
>
> *Crime 2: Business robbery.* During the STREET MONTHS ON THE CALENDAR did you rob any businesses? That is, did you hold up a store, gas station, bank, taxi or other business?

> *Crime 3: Person robbery.* During the STREET MONTHS ON THE CALENDAR did you rob any persons, do any muggings, street robberies, purse snatches, or hold-ups in someone's house or car? (Do not include any business robberies or hold-ups during a burglary that you already mentioned.)
>
> *Crime 4: Theft.* During the STREET MONTHS ON THE CALENDAR did you do any theft or boosting? That is, did you steal from a till or cash register, shoplift, or pick pockets, or take something from someone without their knowledge? (Do not include car theft.)
>
> *Crime 5: Auto theft.* During the STREET MONTHS ON THE CALENDAR did you steal any cars, trucks, or motorcycles?
>
> *Crime 6: Forgery and cards.* During the STREET MONTHS ON THE CALENDAR did you ever forge something, use a stolen or bad credit card, or pass a bad check?
>
> *Crime 7: Fraud.* During the STREET MONTHS ON THE CALENDAR did you do any frauds or swindles (illegal cons) of a person, business, or the government?
>
> *Crime 8: Drug dealing.* During the STREET MONTHS ON THE CALENDAR did you ever deal in drugs? That is, did you make, sell, smuggle, or move drugs?

Respondents who answered "yes" to one of these questions were next asked whether the number of crimes of that type they did was between 1 and 10, or 11 or more. (See Fig. 2.2 for an example of the format.) If the number was between 1 and 10, they were asked to specify it. If 11 or more, they were asked to tell the number of months in which they did the crime and supply a rate (number of crimes per month, per week, or per day).

Because some respondents provided a range (e.g., "3 to 5 times per week") and others answered the question in several different places (e.g., "2 times per week" and "6 times per month"), the data we received and processed included both a minimum and a maximum estimate of the number of crimes done. In many instances, respondents checked the box "1 to 10" but provided no further information, so the minimum was considered to be 1, and the maximum, 10. Others checked only the box "11 or more" so their minimum was 11, and their maximum (effectively), infinity.

Data for respondents who answered "no" to a screening question were given a special code indicating that their number of crimes of

The next questions are also only about the STREET MONTHS ON THE CALENDAR. Look at the calendar to help you remember what you were doing during these months. These are months that do not have X's or lines in them.

I. 1. During the STREET MONTHS ON THE CALENDAR did you do any burglaries? (Count any time that you broke into a house or a car or a business in order to take something.)

YES ☐$_1$ NO ☐$_2$ ➧ go on to page 18 30/

2. In all, how many burglaries did you do?

☐ 11 OR MORE ☐ 1 TO 10 31/
How many?
☐ Burglaries 32 ''/

3. Look at the total street months on the calendar. During how many of those <u>months</u> did you do one or more burglaries?

go on to next page →

_______ Months 34 ''/

4. In the months when you did burglaries, how <u>often</u> did you usually do them?

(CHECK ONE BOX)

EVERYDAY OR ALMOST EVERYDAY ☐ → How many per *day?* ☐ → How many days a week usually? ☐ 36/ 37 ''/ 39/

SEVERAL TIMES A WEEK ☐ → How many per *week?* ☐ 40/ 41 ''/

EVERY WEEK OR ALMOST EVERY WEEK ☐ → How many per *month?* ☐ 43/ 44 ''/

LESS THAN EVERY WEEK ☐ → How many per *month?* ☐ 46/ 47 ''/

Fig. 2.2—Format of survey questions on business robbery

that type was reportedly zero. Data for respondents who failed to answer any of the questions about a particular crime were coded "missing."

Thus, the actual data are less satisfactory than the data that would be produced by our ideal data collection method, in several ways. First, the sample of offenders studied is not a random subset of offenders in the community. Second, because the data were not collected contemporaneously with the crime commissions, but later, the number of crimes counted is subject to response error (lack of candor, poor recall, uncertainty about the relationship of the crime commission to the measurement period, etc.). Third, if the number of street months and/or the number of crimes is given as a range, the reported rate for the respondent is unclear. In addition, due to the format of the questions, many respondents have exactly the same range in their response.

Elsewhere in the survey booklet, the respondents were asked whether or not they had done each of the same eight crimes (a) in the two calendar years before the measurement period and (b) in the two calendar years before that. These data are also used in our analysis.

<u>NONPARAMETRIC REPRESENTATION OF DISTRIBUTION OF OBSERVED CRIME RATES</u>

In light of the ambiguities and omissions in self-report data on the number of crimes an individual commits and his street time, it is not at all obvious how such data should be used to generate descriptive statistics on observed crime commission rates. With data having missing values and ranges, the empirical cumulative-distribution function is not well defined. Therefore, we shall describe how methods appropriate for censored data can be used to estimate the distribution of crime commission rates. This exploratory data analysis procedure is an important first step toward understanding the data before turning to statistical modeling, using specific functional forms of the distributions. The estimated empirical distribution function described below can serve as a standard of comparison when fitting functional forms to the data.

The data to be used for this illustrative example came from Second Rand Inmate Survey responses from 440 Michigan prison inmates.

Three types of data elements were used for each of the respondents. The first element, "street months," may be either a number or a range. The second element, "number of crimes," is provided for the eight crime types listed above, each of which may have a specified number or a range. The third element, annualized "observed crime rate," is determined for each of the eight crimes on the basis of number of street months and number of crimes. For example, a respondent who reports 6 street months and 2 crimes has an annualized observed crime rate of 4. For interval responses, a minimum crime rate and a maximum crime rate are calculated. For example, if the minimum number of crimes is 1 and the maximum is 10, and the minimum number of street months is 3 and the maximum is 4, then the minimum observed crime rate is $12 \cdot 1/4 = 3$ and the maximum observed crime rate is $12 \cdot 10/3 = 40$.*

Among the various methods available for handling this type of data, we chose to use a technique proposed by Turnbull (1976) for estimating, by maximum likelihood, the empirical distribution function with arbitrarily censored data. Suppose that the crime commission rate for the i^{th} individual is known to fall in the interval $[L_i, R_i]$, where R_i may be infinite, and L_i is finite with $L_i \le R_i$. Then for any nondecreasing function F, we can define the likelihood

$$L(F) = \prod_{i=1}^{n} (F(R_i+) - F(L_i-)), \qquad (2.1)$$

where $F(R_i+)$ is the greatest lower bound of $F(x)$ for $x > R_i$, and $F(L_i-)$ is defined analogously. Turnbull (1976) shows that L is maximized by a function F* which is flat on a certain number of intervals (depending on the censoring configuration), just as a histogram is, except that the intervals are of varying length dictated by the data. Between the "flat spots," the value of F* may be arbitrary, as long as

*This calculation yields a conservative estimate of the range for observed crime commission rate. Alternatively, the number of crimes and the number of street months can be considered as separately censored random variables, although this would require adapting methods for bivariate censoring (Muñoz, 1980).

it is increasing. The algorithm for computing Turnbull's maximum-likelihood estimate is an iterative one using the self-consistency algorithm. It requires an initial estimate of the size of the jump between the flat spots of the distribution function. If there are m flat spots, then m initial values summing to unity must be given (for instance, all m values may be initialized at 1/m).

Table 2.1 and Fig. 2.3 illustrate how the Turnbull empirical distribution function is calculated from observed commission rates for person robbery. The second column in Table 2.1 gives the nonparametric estimate of the distribution function on the indicated interval. The locations of flat spots and gaps in Fig. 2.3 clearly highlight some aspects of the data-collection instrument. For example, the estimated cumulative distribution function is flat between 52 and 60 robberies per year. This occurs because if a respondent committed over 10 robberies, he can give a daily, weekly, or monthly frequency. Since once a week yields an annualized rate of 52.0, and five times a month yields 60.0, if a respondent is filling in the questionnaire properly, there is no way he can give a rate between 52.0 and 60.0. (Of course, a respondent who reported, say, nine crimes in 2 street months would have an annualized rate between 52.0 and 60.0, but combinations like this are rare.)

Similarly, the instrument design implies many other such "flat intervals." The censoring in responses is also apparent in the "gaps" in the distribution function where it is arbitrarily defined. For example, there is a jump from .949 to .955 somewhere in the interval between 25.8 and 34.4 robberies per year. All the respondents whose commission rates *may* be in the interval (25.8, 34.4) gave uncertain responses at least as broad as the interval.

An advantage of plots like Fig. 2.3 is that they show exactly how artifacts and ambiguities of the data-collection instrument translate into uncertainties and irregularities in observed crime rates. Note that *no assumptions about the functional form of the crime rate distribution* are needed in this analysis. For this reason, plots of cumulative frequencies are useful as first steps in the exploratory analysis of crime rate data and can permit comparisons with parametric distributional forms, discussed in the following section.

Table 2.1

TURNBULL EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTION
FOR PERSON ROBBERIES

Interval	Value of Distribution Function	Interval	Value of Distribution Function
0.000-0.546	0.739	5.455-5.647	0.902
0.546-0.571	0.741	5.647-5.714	0.905
0.571-0.600	0.747	5.714-6.000	0.909
0.600-0.632	0.751	6.000-6.400	0.909
0.632-0.667	0.756	6.400-6.750	0.913
0.667-0.800	0.758	6.750-6.947	0.917
0.800-0.923	0.761	7.200-7.333	0.918
0.923-1.000	0.775	7.500-7.636	0.919
1.000-1.091	0.776	8.000-9.600	0.920
1.091-1.200	0.785	10.000-10.500	0.924
1.200-1.412	0.788	10.500-12.000	0.929
1.412-1.500	0.792	12.000-14.400	0.934
1.500-1.714	0.806	14.4 -18.9	0.938
1.714-1.846	0.819	18.9 -20.0	0.942
1.846-1.895	0.824	20.0 -22.5	0.943
1.895-2.000	0.829	22.5 -24.0	0.948
2.000-2.400	0.834	24.0 -25.8	0.949
2.400-2.571	0.838	34.4 -50.2	0.955
2.667-3.000	0.840	50.2 -60.0	0.961
3.000-3.429	0.849	60.0 -68.6	0.962
3.429-3.600	0.853	92.9 -110.6	0.963
3.600-3.750	0.858	110.6 -183.5	0.968
3.750-4.000	0.861	190.5 -196.1	0.971
4.000-4.500	0.872	196.1 -206.4	0.981
4.667-4.800	0.879	206.4 -258.0	0.986
4.800-5.143	0.891	258.0 -265.4	0.990
5.143-5.333	0.894	265.4-2528.4	0.995
5.333-5.455	0.898	2528.4-4024.8	0.997
		4024.8- ∞	1.000

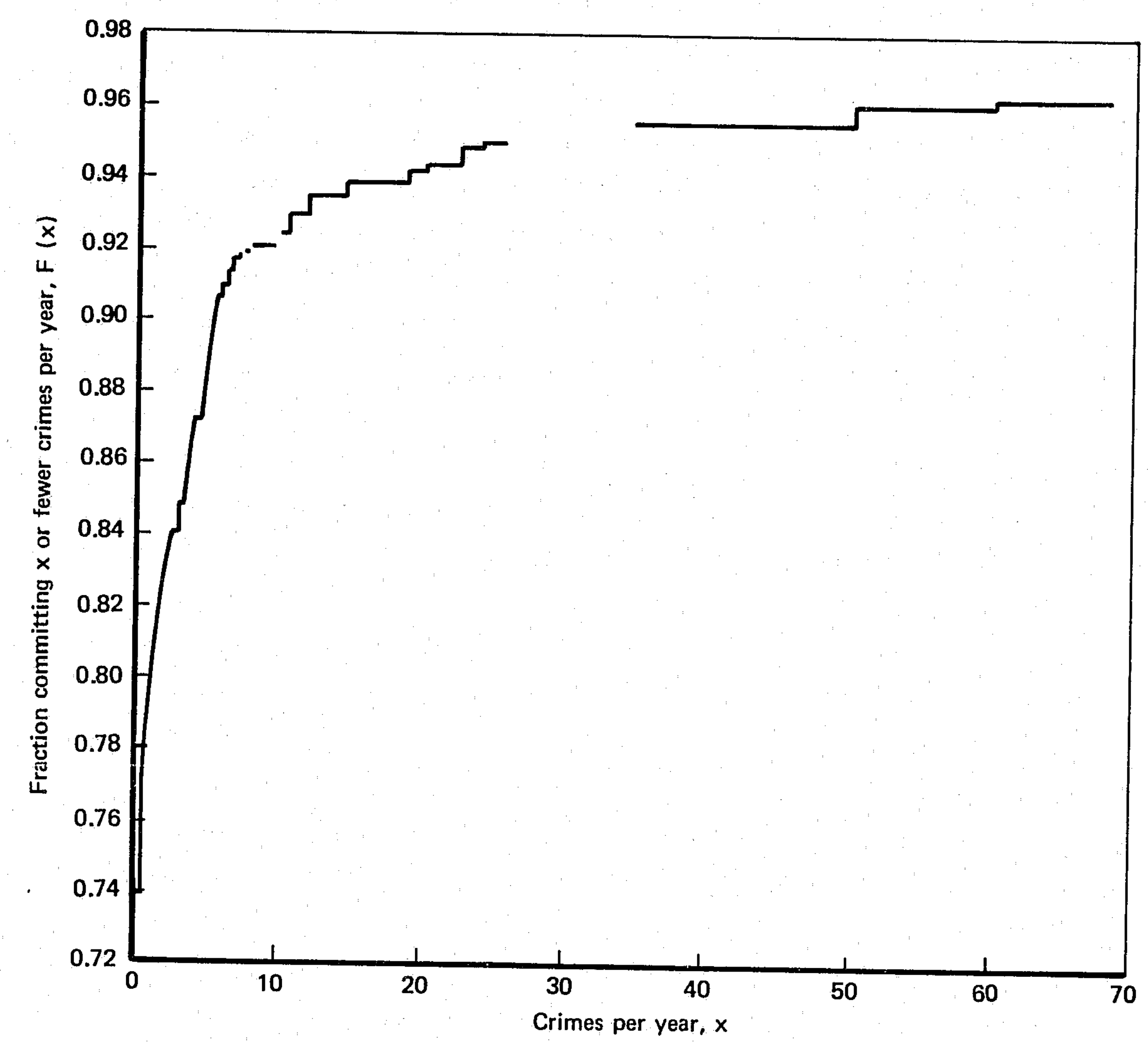

Fig. 2.3—Turnbull empirical cumulative distribution function for person robbery

FITTING PARAMETRIC DISTRIBUTIONS TO OBSERVED CRIME RATES

Parametric distributional forms are helpful for describing data in a parsimonious manner and for developing conceptual models of criminal behavior. We are concerned here with the annualized observed crime rates, Y_i/T_i, where Y_i is the number of crimes committed by individual i during a measurement period of length T_i.

We are not aware of any earlier research that determines the form of the distribution of crime commission rates for some sample of individuals. However, studies of measures of criminal activity that are closely allied with crime commission rates indicate the types of distributions that might be expected. These measures include an individual's time between arrests (Wolfgang, Figlio, and Sellin, 1972), his arrest rate (Greene, 1977), and his failure time (time after first arrest after release from a correctional program) (Stollmack and Harris, 1974). These collectively provide some support for the use of Pareto, gamma, and exponential distributions for the crime rate and negative binomial distributions for the count of crimes, but there is no compelling evidence for any one form.

A distribution of the "total" crime rate from the Rand First Inmate Survey has been graphed by Peterson, Braiker, and Polich (1980, p. 34). The smooth superimposed curve in their graph shows the gamma distribution with parameters $\alpha = .3$ and $\beta = .01875$, which appears visually to match their data well. However, a grouped chi-square test of fit rejects the hypothesis that the data arise from that gamma distribution.

With limited guidance of this type, we chose to examine the Pareto, gamma, and lognormal forms. The two-parameter Pareto distribution with zero origin has distribution function

$$F(x) = 1 - \left(1 + \frac{x}{\sigma}\right)^{-a} \qquad x \geq 0, \qquad (2.2)$$

with shape parameter a and scale parameter σ. It has such a heavy tail that we felt it might potentially fit the data, including counts of zero ($Y_i = 0$). However, since the Pareto distribution has a density

function, any random variable X with this distribution has zero probability of achieving exactly the value X = 0, so it is inappropriate to include a large number of individuals with $Y_i/T_i = 0$ when estimating the parameters of the distribution.

As an *ad hoc* procedure for avoiding this zero problem while still seeing whether the Pareto distribution is suitable as an approximate description of the data, we converted responses with $Y_i/T_i = 0$ into interval responses with a lower terminal of 0 and an upper terminal of $1/(12\ T_i + 1)$. (Here, $12\ T_i$ is the number of street months in the measurement period for individual i.) The parameters of the Pareto distribution were then estimated for each of the eight crime types, using a method of maximum likelihood that takes into account all the observations, uncensored as well as censored (see Appendix A).

The Pareto distributions with the maximum-likelihood parameters were then compared with the estimated empirical distribution functions by grouped chi-square test. The results showed that the Pareto form could be accepted for four of the eight crime types (see Table 2.2). However, the mean of the Pareto distribution (Eq. (2.2)) is finite only when the shape parameter a > 1, which is not the case for the shape-parameter values shown in Table 2.2. Hence, although the Pareto

Table 2.2

FIT OF PARETO DISTRIBUTION TO ILLUSTRATIVE CRIME RATE DATA

Crime Type	Number of Observations: Nonzero	Number of Observations: Zero	Parameters of Pareto Distribution Including all Zero Values: α	Parameters of Pareto Distribution Including all Zero Values: σ
Burglary	174	225	--	--
Business robbery	108	325	0.582	0.0867
Person robbery	109	325	0.485	0.0525
Theft (other than auto)	166	261	--	--
Auto theft	87	321	--	--
Forgery/credit cards	56	365	0.389	0.0087
Fraud/swindles	69	368	0.457	0.0156
Drug dealing	163	254	--	--

form is statistically acceptable as a description of four crime types, including their zero counts, it is not a practically satisfying description of the data, because it has an infinite mean. (Since no other functional form fits the data with zero values of Y_i/T_i treated as if they were very small positive values, the infinite mean of the fit Pareto distribution suggests that the problem of "too many zeros" must be handled in some other way.)

The possibility of a lognormal distributional form for the eight crime types was examined by transforming the data and examining standard probability plots and statistical tests for the normal distribution. The lognormal form was rejected for all eight crime types.

To examine the fit of the gamma distributional form (Eq. (1.2)) to observed crime rates, we took advantage of the fact that quantile ratios of the gamma distribution are independent of the scale parameter β. (For example, in all gamma distributions with shape parameter $\alpha = 0.40$, the ratio of the 75^{th} percentile to the 50^{th} percentile is 3.48, and the ratio of the 50^{th} percentile to the 25^{th} percentile is 6.16; see Fig. 2.4). By selecting three quantiles (we chose the 25^{th}, the 50^{th}, and the 75^{th}) and calculating two quantile ratios from the empirical distribution, it is possible to determine whether there is any *possible* value of α for which a gamma distribution with shape parameter α might match the data. (For example, if the ratio of the empirical 75^{th} percentile to the empirical median is 3.5, and the ratio of the empirical median to the empirical 25^{th} percentile is 6.2, the data might *possibly* correspond to a gamma distribution with α = 0.40. But if the first ratio is 3.5 and the second ratio is 10.0, there is no gamma distribution that matches the data.)

We chose this "quartile ratio" method primarily for its simplicity and robustness. As described earlier, very large and very small observations can have considerable influence in the fit, and we wish to restrict this influence. Using the middle half of the distribution seems a good way to do this.

If a candidate value of the shape parameter α exists by this simple consistency test, the corresponding value of the scale parameter β can be estimated by the method of Särndal (1964), which uses the best

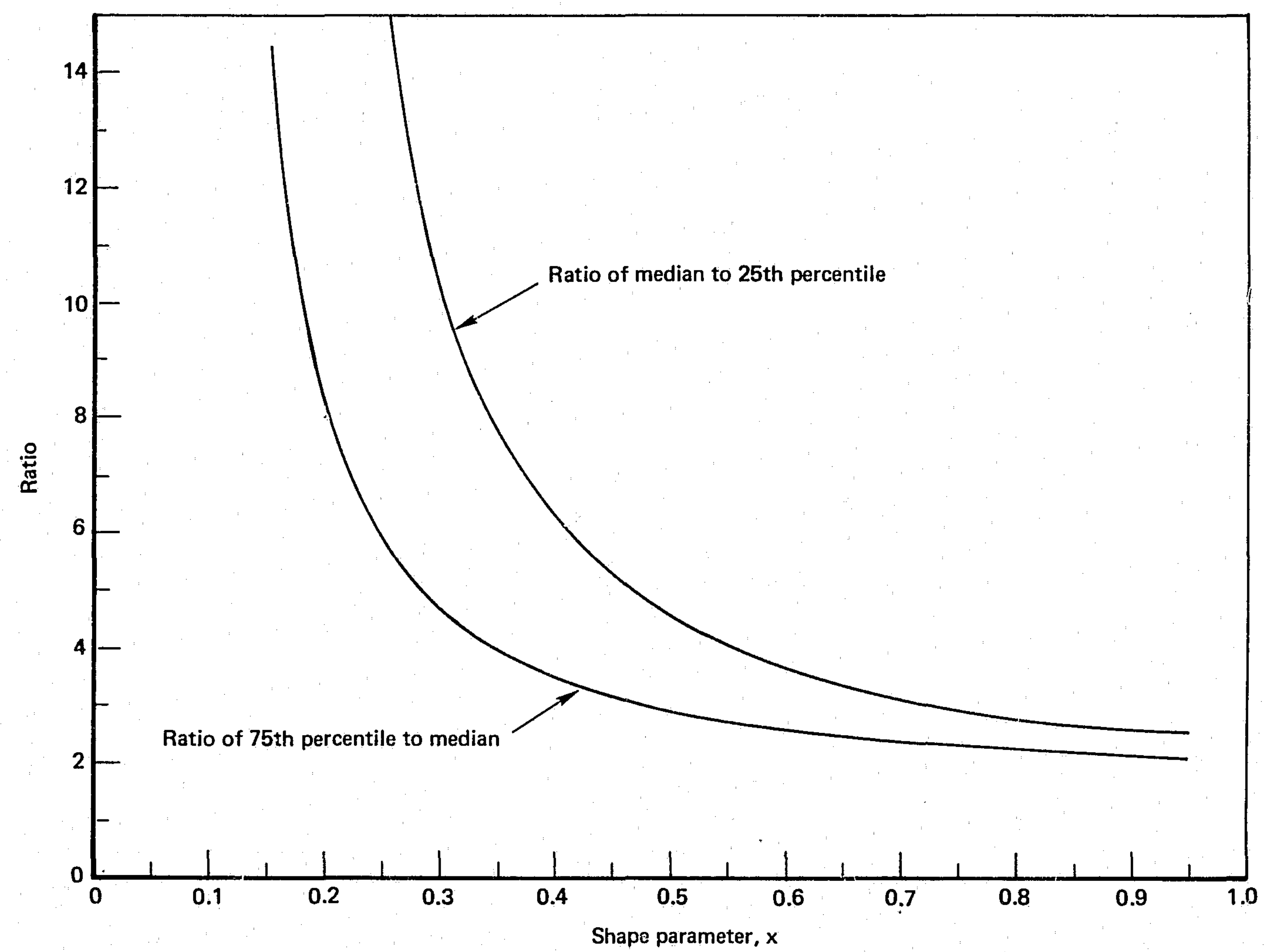

Fig. 2.4 — Quartile ratios of the gamma distribution

linear function of the order statistics used in the calculation. Again, we used the 25^{th}, 50^{th}, and 75^{th} percentiles (see Appendix A). Once the values of α and β have been estimated, the question of whether the gamma distribution with parameters α,β fits the empirical distribution can be tested by probability plots or grouped chi-square test, as in the case of the Pareto distributions.

When we carried out this process using data from Rand's Second Inmate Survey, we found that none of the eight crime types studied have a gamma distribution when all zeros are included, and only two (business robbery and drug dealing) have a gamma distribution when all zeros are excluded. However, when we varied the number of zeros included from 1 to the maximum possible, we found that three additional crime types fit the gamma form when some zeros are included, and business robbery fit a second distribution when some zeros are included. (Inclusion of zeros shifts the quartiles and hence changes the candidate values of α.) Table 2.3 summarizes the results and also shows the values of the means of the estimated distributions. In contrast to the Pareto distributions, the gamma distributions have finite means and are useful for smoothing the data and estimating means.

Table 2.3

FIT OF GAMMA DISTRIBUTION TO ILLUSTRATIVE CRIME RATE DATA

Crime Type	Observations		Zeros Included in Fit	Gamma Distribution		
	Nonzero	Zero		α	β	Mean of Nonzero Values
Burglary	174	225		--	--	
Business robbery	108	325	0	.80	.107	7.4
			20	.58	.086	8.0
Person robbery	109	325	13	.60	.078	8.6
Theft (other than auto)	166	261		--	--	
Auto theft	87	321	24	.40	.040	12.8
Forgery/credit cards	56	365		--	--	
Fraud/swindles	69	368	14	.45	.051	10.7
Drug dealing	163	254	0	.22	.00034	645.5

Figure 2.5 illustrates three different parametric distributions that fit the empirical distribution of annualized crime commission rates for business robbery.* For clarity, only the truncated distributions are shown; that is, individuals with $Y_i = 0$ are excluded from the figure even though they may have been included when fitting the distributions. The ordinate shows the estimated percentage of respondents in each interval of width 2, i.e., $Y_i/T_i \in (0, 2), [2, 4), [4, 6)$, etc., and the percentage with $Y_i/T_i \geq 20$.

The Pareto distribution was fit to data for all respondents, including those who reported zero business robberies during the measurement period. Figure 2.5 shows that this distribution overestimates the frequency of small, but nonzero crime commission rates, but it gives an excellent fit in the tail of the distribution (where crime commission rates $\geq$ 20). By contrast, the two gamma distributions are closer to the empirical distribution at the low end, but they underestimate the tail. One of the gamma distributions was fit to all 108 nonzero values of Y/T, while the other was fit to the original 108 nonzero values plus 20 zero values. Although the parameters of the two gamma distributions are substantially different, they are nearly identical for practical purposes.

Comparison of Tables 2.2 and 2.3 shows that two crime types (burglary and theft other than auto) were not fit by any of the parametric distributional forms we tried. They are presumably probabilistic mixtures of distributions, so the data for subgroups of offenders (defined by some distinguishing characteristics) might well match one of these functional forms.

FITTING PARAMETRIC DISTRIBUTIONS TO CRIME COUNTS

For those crime types whose observed *annualized commission rates* appear to fit a gamma distribution approximately, a reasonable hypothesis is that the distribution of *counts* of crimes committed, conditioned on the length of the measurement period, is negative binomial. This

*Another example is shown in the Executive Summary of this study, R-2730/1-NIJ.

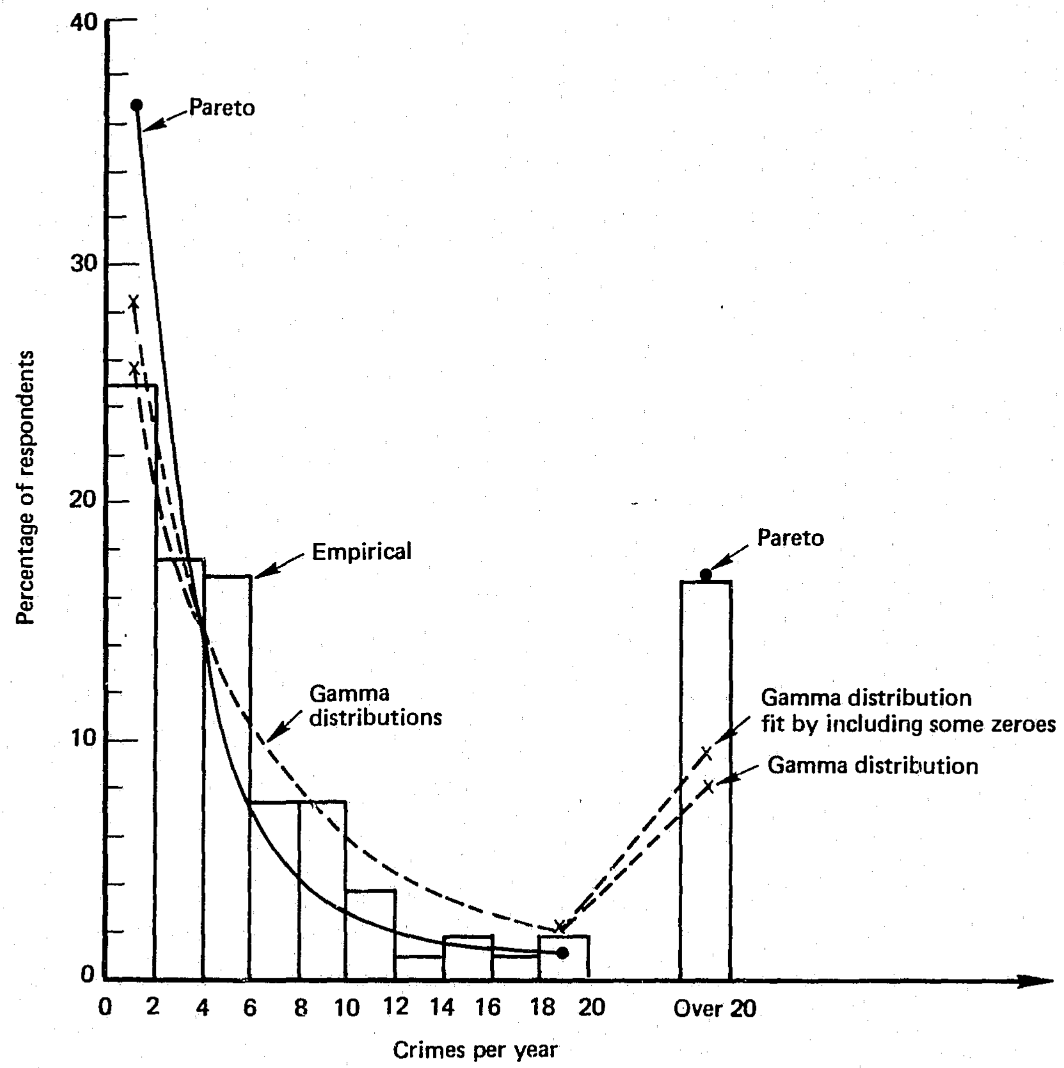

Fig. 2.5 — Distribution of crime rate for business robberies (respondents who commit business robberies)

distribution could arise from a compound gamma-Poisson model or in other ways. If we use the compound gamma-Poisson model, then estimating the parameters of the negative binomial distribution yields estimates of the parameters of the underlying gamma distribution. The negative binomial model has the potential of being formally correct, whereas the observed crime commission rates cannot, in principle, be exactly modeled as having been drawn from a gamma distribution.*

In contrast with the development of the negative binomial distribution, we here permit each individual to have a different length measurement period T_i, and we allow some individuals to have $\lambda(i) = 0$. This implies that the Y_i's are not a random sample from a single negative binomial distribution. Rather, for individuals with $\lambda_i > 0$, the Y_i's are independent draws from negative binomial distributions, each with parameters α, P_i, where $P_i = \beta/(T_i + \beta)$.

Now, let π be the proportion of individuals with $\lambda(i) = 0$, and $1 - \pi$ the proportion of offender individuals whose $\lambda(i)$'s are assumed to be gamma-distributed. The resulting marginal distribution is negative binomial with some extra mass at zero. That is,

$$P(Y_i = y) = \begin{cases} \pi + (1 - \pi)P_i^{\alpha} & \text{if } y = 0 \\ (1 - \pi)\binom{\alpha + y - 1}{\alpha - 1}(1 - P_i)^y P_i^{\alpha} & \text{if } y = 1, 2, 3, \ldots . \end{cases} \tag{2.3}$$

This equation shows explicitly how the zero counts arise from both individuals with $\lambda(i) = 0$ and those with $\lambda(i) > 0$. An estimate of π then specifies how many "real zeros" there are. However, it is less complicated to estimate the parameters α and β (or α and P_i) from the values of Y_i that are positive, and then estimate π separately. That is, we do not ignore the offenders whose $\lambda(i)$ is positive and whose Y_i is zero, but we exclude them from the estimation procedure. The observed frequency function for nonzero values of Y_i is then a *truncated*

---

*The observed crime commission rates are quantized. For example, in the data from the Rand Second Inmate Survey, they cannot have positive values smaller than 0.5 (i.e., 1/(2 years)).

negative binomial and is given by

$$P(Y_i = y | y > 0) = \binom{\alpha + y - 1}{\alpha - 1} \frac{(1 - P_i)^y P_i^{\alpha}}{1 - P_i^{\alpha}}, \qquad y = 1, 2, 3, \ldots \quad (2.4)$$

The use of a truncated distribution is a convenient way to solve the "zero problem." The method is perfectly general, although we apply it here only in the negative binomial case. The idea is to estimate the parameters of the full distribution from the truncated (at zero) sample. The parameter estimates are then used to infer the number of zeros that "belong" to the full distribution. The number of zero λ's is then estimated as the total number of observations minus the above estimate.

Unfortunately, even in the apparently simple case of the truncated distribution shown in Eq. (2.4), the exact computation of the maximum-likelihood estimators of β (which is hidden in the P_i's) and α is intractable. For computational simplicity, then, we analyze a subset of the data, consisting of individuals having the same (or nearly the same) value of T_i. The parameters of the truncated negative binomial can then be fit by maximum likelihood as described in Hartley (1958), and the size of the truncated sample can be estimated as shown in Sanathanan (1977).

We illustrate the method by analyzing reported drug deals from the Rand Second Inmate Survey. The data used are observed frequencies (midpoints of ranges, actually) that occurred with street months = 13, 14, 15, or 16. We treat the data as if all cases involved 14.5 street months. Table 2.4 lists the observed y's and the corresponding frequencies n_y. Two individuals reporting over 4500 drug deals per year were deleted from the analysis as outliers, which left a total of m = 40 nonzero observations. A total of 78 individuals with street months 13, 14, 15, or 16 claimed not to have done drug deals.

We can then estimate the size of the truncated sample (Sanathanan, 1977) as follows: Let N be the total sample size, n the negative binomial sample size, m the number of nonzero observations, and

Table 2.4

TRUNCATED NEGATIVE BINOMIAL FIT TO DRUG DEALING[a]

($\chi_5^2(.10) = 9.236$)

Number of Crimes, y	Observed Number of Individuals, n_y	Chi-Square Test			
		Class - Intervals	Expected	Observed	$(O-E)^2/E$
1	1	1-2	5.266	4.0	.304
2	3	3-7	4.931	6.0	.232
3	1	8-18	4.817	8.0	2.103
4	1	19-42	5.038	3.0	.824
5	3	43-90	4.991	3.0	.794
6	1	91-188	4.975	7.0	.824
8	3	189-422	4.995	3.0	.797
9	1	423-∞	4.987	6.0	.206
10	1				6.084
11	1				
12	1				
16	1				
21	1				
24	1				
32	1				
52	2				
77	1				
103	1				
120	1				
129	1				
146	1				
150	1				
172	1				
181	1				
323	1				
360	1				
391	1				
482	1				
542	1				
587	1				
697	1				
1032	1				
1264	1				
	m = 40				

[a]Truncated negative binomial was fit by maximum likelihood according to the method described in Hartley (1958). For $T_i \equiv 14.5$, the maximum-likelihood estimators are $\hat{P} = .00143$, $\hat{\alpha} = .1698$.

$n_0 = (n - m)$ the number of zeros from the negative binomial. The conditional maximum-likelihood estimate $\hat{n}$ of the true sample size n is then the integer part of $m/(1 - \hat{P}^{\hat{\alpha}})$. Using m = 40, $\hat{P}$ = .00143, and $\hat{\alpha}$ = .1698 yields the estimate, $\hat{n}$ = 59. Hence, we estimate the number of zeros in the negative binomial as $n_0 = \hat{n} - m = 59 - 40 = 19$. We therefore estimate that of the 78 people who claimed zero drug deals and had 13, 14, 15, or 16 street months, 19 have nonzero $\lambda(i)$ for drug deals and the remaining 59 have $\lambda(i) = 0$.

Table 2.4 shows a chi-square goodness-of-fit test for the truncated negative binomial. The fit is satisfactory at the level of aggregation used, so the approach is satisfactory for the subset of data concerning drug dealing that we analyzed. To the extent that drug deal offenders with street times of 13, 14, 15, or 16 months can be considered representative, these estimates of α and β can be used for all drug deal offenders. Otherwise, we recommend the much less tractable procedure of computing maximum-likelihood estimates from the complete data set, using Eq. (2.4).

Remarkably enough, although it would seem that 59 "true zeros" is a large proportion of the 78 observed zeros, the data do not permit rejecting the hypothesis that all the observed zeros come from a negative binomial distribution. We tested the hypothesis H_0: $\pi = 0$ against the alternative hypothesis H_1: $\pi \neq 0$. Under H_0, the maximum-likelihood estimates of the parameters are $\hat{\alpha}_0$ = .06135 and $\hat{P}_0$ = .00125. Defining the function u by

$$u(\underset{\sim}{y}, \alpha, P, \pi) = \prod_{i=1}^{N} \text{Prob}(Y_i = y_i),$$

where each term on the right is given by Eq. (2.3), the test statistic is

$$-2 \log \frac{u(\underset{\sim}{y}, \hat{\alpha}_0, \hat{P}_0, 0)}{u(\underset{\sim}{y}, \hat{\alpha}, \hat{P}, \hat{\pi})},$$

which is asymptotically distributed as chi-square with 1 degree of freedom. The value of the statistic for our data is 1.32, which is not statistically significant.

We also fit the negative binomial distribution to the data for person robbery. In this case, $\hat{n}$ = 84, m = 37, and n_0 = 47, and the maximum-likelihood algorithm was numerically unstable. An alternative method, that of moments estimation (Sampford, 1955), was used to get the values $\hat{\alpha}$ = .108 and $\hat{P}$ = .00485. Table 2.5 gives the data, the fitted values, and associated chi-square values.

Table 2.5

TRUNCATED NEGATIVE BINOMIAL FIT TO PERSON ROBBERY

Number of Crimes, y	Observed Number of Individuals, n_y	Chi-Square Test			
		Class Intervals	Observed	Expected[a]	$\frac{(O-E)^2}{E}$
1	7	1	7	5.11	0.70
2	5	2–3	7	4.79	1.02
3	2	4–9	12	6.29	5.18
4	2	≥ 10	11	20.82	4.63
5	3				
6	4				11.53
7	0				
8	2				
9	1				
≥ 10	11				
	m = 37				

NOTE: $\chi^2_1(.01) = 6.63$. Reject truncated negative binomial.

[a]Parameters estimated by method of moments; $\hat{\alpha}$ = 0.108 and $\hat{P}$ = 0.00485.

The difficulties with person-robbery data illustrate the shortcomings of fitting the negative binomial by the truncation method. However, the truncation method can be used for any functional form—it does not have to be negative binomial. Thus if the truncated negative binomial fit is not satisfactory for given data sets, other functional forms can be tried with the truncation approach. A satisfactory fit solves the problem of "too many zeros."

APPROPRIATENESS OF THE POISSON ASSUMPTION

In examining the applicability of the compound gamma-Poisson model (at least to certain crime types), it is important to verify not only that the consequences of the model are correct (e.g., that the distribution of counts of crimes, conditioned on the length of the measurement period, is negative binomial), but also that the assumptions of the model are correct. One of these assumptions is that each offender commits crimes according to a Poisson process when he is "on the street," i.e., free to commit crimes. In examining the data, however, we find that the Poisson assumption may be correct for some types of crimes but not for others.

To carry out a formal test of the validity of the Poisson assumption (i.e., the "Poissonicity") of crime occurrences, it would be desirable to have interval data--that is, data showing the exact dates on which each individual's crimes were committed and the exact periods when he was on the street. By stringing together the street periods of numerous offenders who have approximately the same crime commission rates, we would obtain a realization of a stochastic process with enough data points to test whether the process is Poisson. (Such a test is carried out for fire alarms by Carter and Rolph, 1979.)

However, the data from the Rand Second Inmate Survey do not include the dates on which crimes were committed, and in the absence of our ideal observation method, it is difficult to imagine how such data could be successfully obtained. The survey responses simply provide an estimate of the total count of crimes committed. However, the responses also tell us how many street months the offender had during the measurement period, and for those offenders who committed 11 or more crimes of a given type during the measurement period, the number of months during which those crimes were committed. We shall discuss here what can be learned from such data on counts and duration of activity. See Cox and Lewis (1966) for a more general discussion of tests of Poissonicity.

Recall that the respondent's measurement period ended with the month he was arrested for his current conviction crime. This could be any month between January and December and, apart from possible seasonal variations in arrests, is essentially randomly chosen for each offender.

Hence one would expect that offenders with long street time would have higher *counts* of crimes Y_i than offenders with short street time, although their rates Y_i/T_i would be (collectively) approximately the same. (The argument is not entirely correct, because it ignores the fact that offenders with high arrest rates would be more likely to be arrested early in the year or to lose street time by incarceration. But the basic pattern should be as described.)

Surprisingly, the data confirm this expectation only for certain crime types. For example, Table 2.6 shows crime counts against street months for person robbery, and Table 2.7 gives analogous information for business robbery. The variation for person robbery is generally in the direction anticipated, although it is not statistically significant, and the same is true for auto theft, forgery, fraud, and drug deals. (The pattern for drug deals is statistically significant because of the large number of "doers" in the sample.) By contrast, Table 2.7 shows a complete lack of variation in business robbery counts with the number of street months, and we found a similar, but less striking lack of pattern for burglaries and thefts. We conclude, then, from this superficial examination of the data, that at least three crime types are probably not committed according to a Poisson process.

Table 2.6

RELATIONSHIP OF CRIME COUNTS TO STREET TIME: PERSON ROBBERY[a]

	Number of Robberies									
	1-3		4-6		7-9		10 or more		Total	
Street Months	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
1-3	4	67	10	40	16	38	12	34	42	39
4-6	2	33	10	40	9	21	7	20	28	26
7-9	0	0	2	8	5	12	3	9	10	9
10 or more	0	0	3	12	12	29	13	37	28	26

[a]The pattern is similar for auto theft, forgery, fraud, and drug deals.

Table 2.7

RELATIONSHIP OF CRIME COUNTS TO STREET TIME: BUSINESS ROBBERY[a]

	Number of Robberies					
	1-10		11 or more		Total	
Street Months	N	Percent	N	Percent	N	Percent
1-6	8	11	3	12	11	11
7-12	15	20	5	20	20	20
13-18	33	43	10	40	43	43
19-24	20	26	7	28	27	27

[a]The pattern is similar for burglary and theft.

A more precise test of Poissonicity can be obtained by examining whether offenders who do numerous crimes tend to concentrate their criminal activity in short time periods. The advantage of this test is that no assumption need be made about the similarity of crime rates between offenders with long and short street times.* The test is based on the observation that if events occur according to a Poisson process, and N of them occur during a period of length T, then each of the N events is independently uniformly distributed over the interval of length T. Thus, if an offender committed 18 crimes according to a Poisson process during 12 months, it is unlikely that he committed them all during one, two, or three of those months. In fact, let $P(t|N, T)$ denote the probability that all N crimes occur in exactly t of T months. Then

$$P(t|N, T) = \binom{T}{t}\frac{1}{T^N}\sum_{k=0}^{t}(-1)^k\binom{t}{k}(t-k)^N, \qquad t = 1, 2, \ldots, \min(N, T).$$

$$P(t|N, T) = 0, \qquad t > \min(N, T).$$

*For the data from the Rand Second Inmate Survey, this test has the disadvantage that it applies only to those who reported 11 or more commissions of the crime type in question.

If N_i and T_i are the number of crimes committed and the duration of the measurement period for individual i, then $\sum_i P(t|N_i, T_i)$ is the expected number of individuals whose crimes were committed in t months. This expected distribution can be compared with the actual data by a chi-square test.

To obtain an adequate sample size when applying this test to the Rand Second Inmate Survey data, we used data from all respondents (inmates of prisons and jails in three states, not just the Michigan prisoners whose data were used elsewhere in this study). The test showed that the Poisson hypothesis could be rejected at the .01 level of significance for all crime types other than person robbery and auto theft. The data revealed that offenders ordinarily committed the other crimes (burglary, business robbery, theft, forgery, fraud, and drug dealing) in a much smaller fraction of their street time than would be the case if commissions followed a Poisson process. Table 2.8 shows an example for theft.

Consequently, we conclude that the Poisson assumption, while not rigorously tested by these data, appears to be tenable only for certain crime types. Other crime types may require treatment by a "switching model," i.e., from time to time the crime commission process turns on and off, and within the "on" periods a high-rate Poisson process applies. Such a model has been proposed and used by Maltz and Pollock (1980). They envision two behavioral states for an offender, which they call "quiescent" and "active." In either state, the offender commits crimes according to a Poisson process, with a state-dependent rate, and transitions between states occur according to a continuous-time Markov process. The offender's long-term average crime commission rate (e.g., the rate observed in the Rand survey) is then a function of his transition rates between states as well as his commission rates in each of the states.

Table 2.8

TEST OF POISSON ASSUMPTIONS FOR COMMISSIONS OF THEFTS

Number of Months, t	Number of Individuals With All Crimes Committed in t Months: Expected		Observed
1	10.1	<	16
2	4.6	<	17
3	5.1	<	20
4	7.6	<	19
5	8.0	<	12
6	15.1	<	27
7	11.0	<	16
8	13.1	<	14
9	12.2	>	3
10	12.5	<	21
11	9.1	>	6
12	10.0	<	15
13	18.1	>	9
14	10.3	>	9
15	21.7	>	13
16	10.9	>	7
17	13.5	>	4
18	9.3	>	5
19	7.4	>	2
20	16.4	>	8
21	12.1	>	9
22	9.2	>	2
23	4.0	>	1
24	5.6	>	2
Total			257

χ^2 = 167 with 23 degrees of freedom.

NOTE: Only individuals who committed 11 or more thefts and reported the number of months in which they did those thefts are included.

III. ESTIMATES OF INDIVIDUAL CRIME PROPENSITIES: SHRINKAGE ESTIMATORS

In Sec. II, we emphasized fitting a distribution to a sample of reported crime rates or to reported numbers of crimes. But frequently, more micro-level questions are of interest. These range from needing good estimates of individual crime propensities, say, for evaluating incapacitation strategies (Chaiken and Rolph, 1980), to needing estimates of crime propensities for small groups of offenders with similar characteristics. In this section, we combine Bayesian and shrinkage estimators with regression methods to derive estimators of individual crime propensities that are appropriate for both purposes.

The problem of estimating individual crime commission propensities is nontrivial. Let us consider the kind of situation we postulated in the gamma-Poisson model, namely that the $\lambda(i)$'s have a probability distribution, rather than being all the same. The gamma distribution of the $\lambda(i)$'s implies a particular distribution of the usual estimator of λ, $\hat{\lambda}(i) = Y_i/T_i$.

Because each $\hat{\lambda}(i)$ is equal to $\lambda(i)$ plus a random error, the spread of the $\hat{\lambda}(i)$'s is greater than the spread of the distribution of the $\lambda(i)$'s. Figure 3.1 illustrates this phenomenon in an example of the compound gamma-Poisson model. Note the greater spread of the observed counts, as compared to the gamma distribution. From this example, it is intuitively plausible that a better estimate of $\lambda(i)$ than Y_i/T_i can be obtained by some sort of shrinking of Y_i/T_i toward the center. In the following, we present such estimators of $\lambda(i)$ that are superior to Y_i/T_i.

BAYES ESTIMATORS

Description

With the assumption that the $\lambda(i)$'s have an *a priori* distribution, the relevant information for inference about $\lambda(i)$, given the data, is in the *a posteriori* (or posterior) distribution of $\lambda(i)$, given Y_i and T_i.

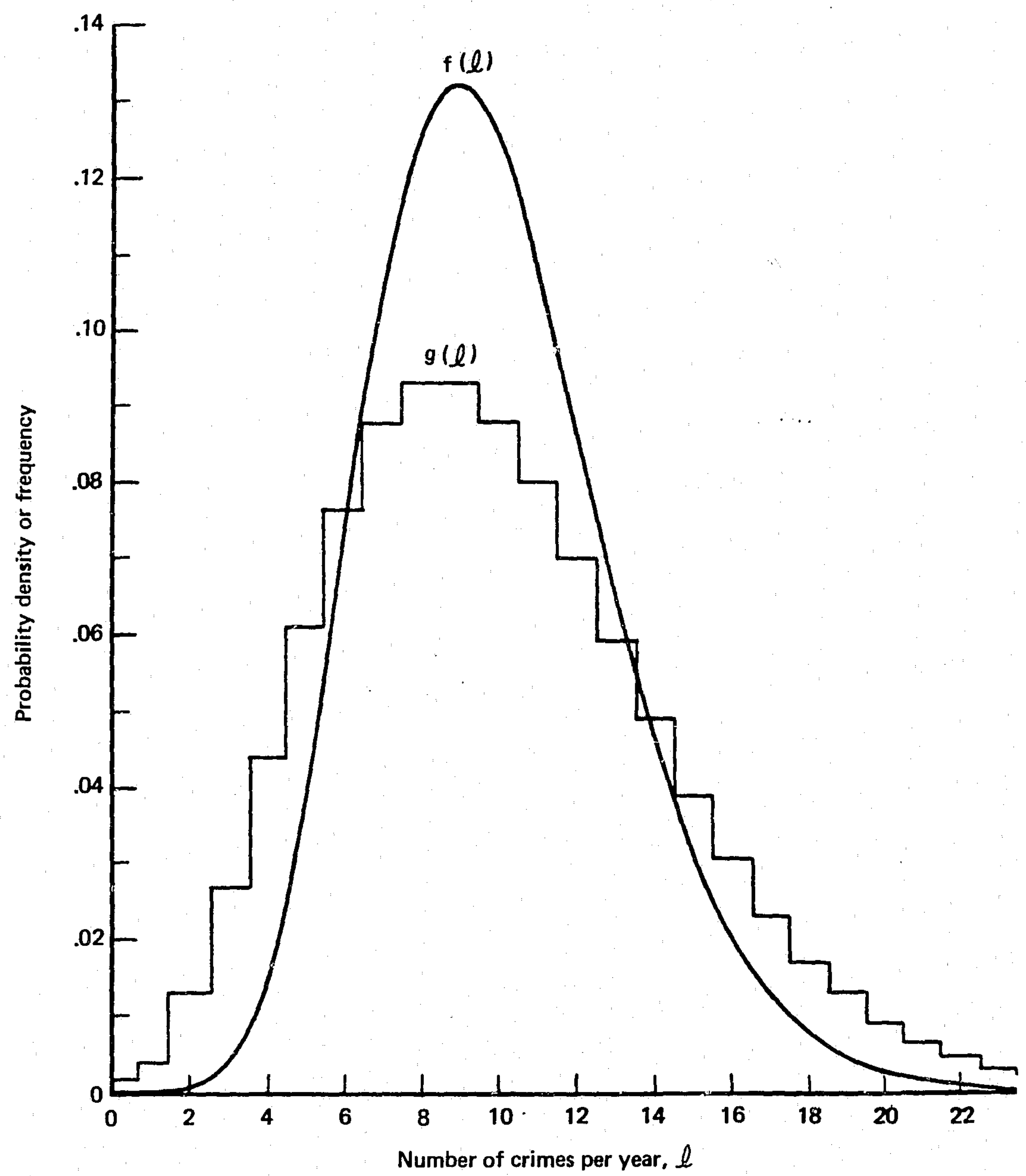

Fig. 3.1 — Example of underlying and "spread" distribution. (The underlying distribution has a density f, which is gamma with shape parameter 10 and scale parameter 1, so the mean crime commission rate is 10/year. The estimate is obtained by counting crimes over a 1-year period and has a frequency function g)

We have shown that under the negative binomial model, this distribution is gamma with parameters $(\alpha + Y_i, \beta + T_i)$. Therefore, the mean and variance of the posterior distribution of $\lambda(i)$, given Y_i and T_i, are $(\alpha + Y_i)/(\beta + T_i)$ and $(\alpha + Y_i)/(\beta + T_i)^2$, respectively. The Bayes estimator of $\lambda(i)$ is defined to be the mean of this posterior distribution, or

$$\hat{\lambda}(i) = (1 - w_i)\,\frac{Y_i}{T_i} + w_i \cdot \frac{\alpha}{\beta}, \qquad (3.1)$$

where

$$w_i = \frac{\beta}{\beta + T_i}.$$

In this case, the usual estimator Y_i/T_i is shrunk toward the *a priori* mean of $\lambda(i)$, namely α/β. Note that the longer individual i is observed (the larger T_i), the closer $\hat{\lambda}(i)$ is to the usual estimator $\hat{\lambda}(i) = Y_i/T_i$. Thus, the estimated $\lambda(i)$ for a criminal whose behavior is observed for a very short time, T_i, will be close to α/β, since $w_i = \beta/(\beta + T_i)$ is close to 1. (While Eq. (3.1) depends on the gamma-Poisson assumptions for its optimality properties as a Bayes estimator, more generally it is a linear Bayes estimator (Hartigan, 1969).) If the parameters α and β are known *a priori*, e.g., from earlier studies or from experience, Eq. (3.1) can be used as an estimator of $\lambda(i)$. Otherwise, α, β, and w_i in Eq. (3.1) must themselves be estimated from the data, so that

$$\hat{w}_i = \frac{\hat{\beta}}{\hat{\beta} + T_i}. \qquad (3.2)$$

We can then estimate $\lambda(i)$ by

$$\hat{\lambda}(i)' = (1 - \hat{w}_i)\,\frac{Y_i}{T_i} + \hat{w}_i\,\frac{\hat{\alpha}}{\hat{\beta}}. \qquad (3.3)$$

Such an estimator is called an *empirical Bayes estimator* of $\lambda(i)$.

Where the data themselves are uncertain, as in the Rand Second Inmate Survey, the above formulation can easily allow for the uncertainty--the Y_i's being interval data. Since in our formulation we condition on the values of T_i, allowing for uncertainty in the street time is more difficult and will not be discussed here. Suppose observation i comes in the form $[L_i, R_i]$ as the minimum and maximum counts during street time T_i. Then it is straightforward to show that the density of the posterior distribution of $\lambda(i)$ is

$$f(\lambda | L_i \le Y_i \le R_i, T_i) = C \sum_{y=L_i}^{R_i} (\lambda T_i)^{\alpha+y-1} \exp[-\lambda(\beta + T_i)],$$

where

$$C = \left[\sum_{y=L_i}^{R_i} T_i^{\alpha+y-1} \Gamma(\alpha + y)/(\beta + T_i)^{\alpha+y} \right]^{-1}.$$

The mean of this distribution, analogous to Eq. (3.1), is

$$E(\lambda | L_i \le Y_i \le R_i, T_i) = \sum_{y=L_i}^{R_i} w(y) \frac{\alpha + y}{\beta + T_i} \Bigg/ \sum_{y=L_i}^{R_i} w(y),$$

where

$$w(y) = T_i^{\alpha+y-1} \Gamma(\alpha + y)/(\beta + T_i)^{\alpha+y}.$$

Thus the posterior expected crime propensity, given that y is in the interval $[L_i, R_i]$, is a weighted average of the posterior expected values for each value of y in that interval.

An Example

In Sec. II, we fit a truncated negative binomial distribution to reported drug deals and got $\hat{\alpha} = 0.1698$ and $\hat{P}(14.5) = 0.00143$. Since $P_i = \beta/(T_i + \beta)$, we take T to be the average T_i, or 14.5 months. Then $\hat{\beta} = T\hat{P}/(1 - \hat{P}) = 0.02076$.

Since $\hat{w} = \hat{P}$, we have for $T_i = 14.5$, $\hat{\lambda}(i) = 0.99857\ Y_i/T_i + 0.00143 \times 0.1698/0.02076$. In this case, the values of Y_i/T_i are shrunk less than 1 percent of the way to the estimated average rate. The weights $w_i = \beta/(\beta + T_i)$ differ accordingly for other values of T_i.

The gamma-Poisson example is a particularly simple shrinkage estimator. The form of the estimator of $\lambda(i)$ itself, Eq. (3.3), depends on how α and β are estimated. The optimality properties of either the Bayes estimator (Eq. (3.1)) or the empirical Bayes estimator (Eq. (3.3)) depend critically on the assumption that the $\lambda(i)$'s have a gamma distribution. In the past few years, the problem of simultaneously estimating Poisson means has received increasing attention in the statistical literature. The emphasis has been on finding estimators (or families of estimators) that perform well, in the sense of improving the expected squared-error of estimation as compared to the estimator Y_i/T_i, no matter what the values of $\lambda(i)$ are. Clevenson and Zidek (1975) gives Bayes and empirical Bayes estimators of $\lambda(i)$. The work of Peng (1975), Tsui (1978), Tsui and Press (1977), Hudson (1978, 1980), and finally Hudson and Tsui (1981) has produced successive improvements in simultaneously estimating a set of Poisson means.

MORE GENERAL SHRINKAGE ESTIMATORS

We turn now to adapting some results of Hudson and Tsui (1981) and Hudson (1980) to the estimation of underlying crime propensities.

Historical Background

The concepts underlying our methods were first developed by James and Stein (1961) in their work on simultaneously estimating the means of n normal distributions $(\mu_1, \ldots, \mu_n)$. The James-Stein estimator dominates (in the sense of minimizing mean-square error) the usual (maximum-likelihood) estimator (MLE) for all values of $\mu_1, \ldots, \mu_n$. In contrast to the usual estimator, the benefits of the James-Stein estimator are greatest when $\mu_1, \ldots, \mu_n$ are all close to zero. This result can be extended so that it has important practical consequences whenever hypotheses exist about how the μ_i's are related to one another, e.g., the means are close to some specified values or, less specifically, the means follow a trend or are similar in value. Equation (3.3) is an example in the Poisson case, where the n means are the values of $\lambda(i)$ for n individuals and the estimators are all shrunk toward the same number, namely, the estimated mean of the gamma distribution.

In the same spirit, the James-Stein theory has been applied where the means are adjusted for covariates (see Efron and Morris (1975) and Fay and Herriot (1979)). For normal distributions, this amounts to choosing the preassigned values of the μ_i's by fitting a linear regression to the data and shrinking toward the predicted values given by the regression in the same way the James-Stein estimator shrinks toward zero.

No matter what predicted values are used in either the normal or the Poisson case, the shrinkage estimators have a smaller expected squared error of estimation than the MLE *when all n means are being estimated simultaneously*. That is, when the estimation procedure is being used collectively--say, by an agency such as a major offenders unit--shrinkage does much better *on average* than the alternative methods. However, there is no guarantee that the shrinkage estimators will outperform the MLE *for every individual*. They improve on average, but not necessarily for every single component μ_i. Since a judge who is confronted with specific cases would want to operate on a "case-by-case" criterion, shrinkage estimators may or may not be helpful in such a context.

When estimating individual crime propensities for a collection of n offenders, we must simultaneously estimate the means of n Poisson distributions. Developing shrinkage estimators for n Poisson means that are uniform improvements on the maximum-likelihood estimators is considerably more complicated than the analogous problem for n normal means, as described above. Only in the past few years have statisticians made any real headway in the Poisson case. Clevenson and Zidek (1975) developed the first improvement, while Peng (1975) derived the estimators that are basis for the Hudson-Tsui estimator used in this study (Hudson and Tsui, 1981, and Hudson, 1980). As will be shown below, the form of the estimator is both complex and not intuitively transparent.

In the Poisson case, as in the normal case, the improved estimators can shrink toward zero, toward some grand mean, or toward some predicted value obtained from a regression. The problem of estimating underlying crime propensities $\lambda(i)$ in a Poisson model is usually facilitated by the availability of a considerable amount of background information on offenders, including age, race, socioeconomic characteristics, and previous criminal behavior.

<u>Some General Theory</u>

We begin by describing some estimators derived by Hudson (1980) and Hudson and Tsui (1981). Suppose there are n individuals in the group. Let Y_i and T_i be the reported number of crimes and street time, and let $\hat{H}_1, \ldots, \hat{H}_n$ be the values we wish to shrink toward (in a transformed space to be described). The values of $\hat{H}$ will vary with the application and are discussed below. Then, assuming that $Y_i \sim \text{Poisson}(\lambda(i)T_i)$, $i = 1, \ldots, n$, the Hudson-Tsui estimator of $\lambda(i)T_i$ is given by

$$\widehat{\lambda(i)T_i} = Y_i - \frac{R}{S}(H_i - \hat{H}_i) \quad \text{for } i = 1, \ldots, n, \qquad (3.4)$$

where

$$H_i = \begin{cases} h(Y_i) = \sum_{j=1}^{Y_i} (1/j) & \text{for } Y_i \text{ a positive integer,} \\ 0 & \text{for } Y_i = 0, \end{cases}$$

$$R = \max(0, n - N_0 - 2 - q),$$

N_0 = the number of observed zeros,

$$S = \sum_{i=1}^{n} (H_i - \hat{H}_i)^2,$$

and q depends on the application and is to be specified. For estimating

$\lambda(i)$, we divide Eq. (3.4) by T_i to get

$$\hat{\lambda}(i) = Z_i - \frac{R}{ST_i}(H_i - \hat{H}_i) \quad \text{for } i = 1, \ldots, n. \tag{3.5}$$

where $Z_i = Y_i/T_i$.

For estimating reported crime rates, we use three methods to specify values of $\hat{H}_i$: $\hat{H}_i \equiv 0$; $\hat{H}_i \equiv \bar{H}$, the mean of $H_1, \ldots, H_n$; and $\hat{H}_i$ is the appropriate predicted value of H_i from a regression. These methods are described below.

Choosing Groups and Shrinkage Centers

In applying the theory outlined above, the analyst is faced with two sets of choices. First, should the offenders be broken into separate groups with independent shrinkage estimators for each group? If so, on what basis should these groups be defined? Second, in a given group, what should the counts Y_i be shrunk toward? That is, how is the shrinkage center $\hat{H}$ defined? We will discuss the rationale for these choices in some detail in the discussion of applications that follows.

If the offenders are divided into subgroups according to characteristics other than their data Y_i, the centers of shrinkage $\hat{H}_i$ can be estimated differently for each subgroup. It is desirable to make such a division into subgroups if the relationship between individual characteristics and crime rates is believed to differ among subgroups. (See Carter and Rolph (1979) and Efron and Morris (1973a), for a discussion of division into subgroups in the case of normally distributed variables.) In our illustrative applications, we divide the data into three subgroups and compare three methods for obtaining values of $\hat{H}_i$:

1. $\hat{H}_i$ is set to zero for all individuals in the subgroup.
2. $\hat{H}_i$ is the mean of the transformed counts; that is, $\hat{H}_i \equiv \frac{1}{n}\Sigma H_i$, where n is the size of the subgroup.
3. $\hat{H}_i$ is the value of H_i predicted from a regression equation as follows: The dependent vector in the regression is $H' = (H_1, H_2, \ldots, H_n)$ where n is the number of offenders in the subgroup. If p independent variables describing characteristics of individuals are considered relevant for predicting crime commission propensities, the regression has an n by (p + 1) design matrix X. (It incorporates a row of 1's for the constant term, as well as the values of the p independent variables for the n individuals.) The vector of estimates $\hat{H}' = (\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_n)$ is determined by the least-squares fit as $\hat{H}' = X(X'X)^{-1} XH'$.

The values of q in the definition of R (Eq. (3.4)) are as follows: In Case 1 ($\hat{H}_i \equiv 0$), q is also zero. In Case 2, q is equal to one. In Case 3, q is the rank of the design matrix in the regression, namely p + 1--that is, one more than the number of regressor variables. (Note that p = 0 obtains in Case 2, so that q = 1.)

SOME APPLICATIONS

We return to the Rand Second Inmate Survey to illustrate the use of these methods. Several practical issues must be dealt with before the shrinkage estimators can be applied. First, the subgroups must be selected. In this example we will simply define three subgroups, although elaborations involving more subgroups are possible and would be desirable in analyzing the survey data. Second, where regression methods are to be used to estimate the $\hat{H}_i$'s, we must specify what set of regressors will be entered into the regression.

Grouping of Offenders

Ideally, the division of offenders into subgroups would separate offenders with $\lambda_k(i) = 0$ from those with $\lambda_k(i) > 0$. (Obviously, personal characteristics are not needed to predict the value of λ_k for those with $\lambda_k(i) = 0$.) However, we have repeatedly observed that offenders cannot be separated into "doers" and "nondoers" of crime k simply on the basis of whether their crime k count during the measurement period is zero or nonzero. The theory underlying the Hudson-Tsui estimator assumes that Y_i given $(\lambda(i), T_i)$ has a Poisson distribution, not a truncated Poisson distribution, so some zero counts should normally occur among the "doers." Moreover, it is not appropriate to select $\hat{H}_i = 0$ purely because $Y_i = 0$.

In light of these considerations, we decided to separate individuals into a "previous doer" group and a "previous nondoer" group according to whether or not they had committed the crime in question during the four-year period just preceding the measurement period for the survey. We expected the "previous nondoer" group to have primarily (but not entirely) zero Y_i's, and the "previous doer" group mostly positive Y_i's.

The "previous doers" were further subdivided by separating out offenders whose self-reported crime counts we had some reason to suspect. An external validity check (to be reported in later publications by the Rand survey group) compared responses to questions concerning age, arrests, race, education, prior prison terms, and similar items with external data for the same items. Individuals whose external validity was especially poor were assumed to have potentially invalid self-reported crime counts also.

Regression to Obtain $\hat{H}_i$

The procedure for obtaining $\hat{H}_i$ will be illustrated for our earlier crime-type example, business robbery. A regression model for predicting H_i was generated for the 53 individuals in the "previous doer" group who had not been excluded for poor validity. Intuitively, we would expect the Hudson-Tsui estimator to perform best if the Y_i's for the "previous nondoer" group are shrunk according to Eq. (3.5), with $\hat{H}_i = 0$, while the Y_i's for the "previous doer" group are shrunk using the value of $\hat{H}_i$ estimated from the regression on covariates. By carrying out the analysis, we shall show that this expectation is correct.

The independent variables we used for the regression are shown in Table 3.1.* Construction of the variables was guided by sociological theory and was carried out primarily by Guttman scaling techniques applied to survey responses. (Responses related to the crime rates, Z_i,

*These variables were constructed and provided to us by M. Chaiken.

Table 3.1

FITTED REGRESSION MODEL FOR PERSON ROBBERY: DOERS WITH GOOD VALIDITY

(Number of observations = 50; dependent variable: H_i; $R^2 = .32$)

Variable	Estimate	t-statistic	Significance	Standard Error of Estimate
Intercept	-.6931	-0.51	0.61	1.3606
Badkid	-.5988	-1.77	0.08	0.3375
Kidcrime	.5990	2.03	0.05	0.2957
Pcntjob	-2.7305	-2.13	0.04	1.2793
Longwk1	1.5938	1.88	0.07	0.8487
Longwk2	1.3128	1.23	.23	1.0670
Age16-17	1.7932	1.89	.07	0.9506
Logst	.6624	1.42	0.16	0.4671

were not considered in constructing the independent variables.) "Badkid" is a scaled variable taking on integer values from 0 to 4 indicating the degree to which the individual had contact with the juvenile justice system. "Kidcrime" is similarly scaled and indicates the degree to which individuals committed criminal offenses as juveniles. "Pcntjob" ranges in value from 0 to 1 and reflects the percentage of street time the offender was legally employed. "Longwk1" and "Longwk2" are 0-1 dummy variables, the first indicating partial employment during a period beginning four years prior to the survey measurement period, the second indicating full employment. "Age16-17" is a 0-1 variable telling whether an individual's age during the measurement period was 16 or 17, used to indicate that the juvenile variables have been declared missing. (If the offender was 16 or 17 years old during the measurement period, then "Kidcrime" and "Badkid" cannot be viewed as *predictors* of Z_i; rather, they are alternative descriptions of Z_i.)

"Logst" is $\log(T_i)$. We transformed the number of street months T_i by the logarithm because the other independent variables are conceptually related to the crime rate $Z_i = Y_i/T_i$, not the crime count Y_i, whereas the dependent variable is a transformation $h(Y_i)$. Since the transformation h in Eq. (3.4) is very similar to a logarithmic transformation up to an additive constant (except near zero), we have

$$h(Y_i) \approx \log(Y_i) = \log \frac{Y_i}{T_i} + \log(T_i).$$

Table 3.1 shows that much residual variation remains unexplained after the regressions are carried out. However, our purpose here is simply to demonstrate the use of regression techniques with the Hudson-Tsui shrinkage estimator. Using regression to choose the center of shrinkage for empirical Bayes estimators frees the analyst from the rigidity usually present in regression modeling. That is, to the extent that the regression equation is misspecified or has omitted important explanatory variables, the empirical Bayes estimator compensates by shrinking the observations less. (See Efron and Morris (1975) for a discussion of this point.)

As stated earlier, we shrink the Z_i for those in the "previous nondoer" group toward zero by choosing $\hat{H} = 0$, while for those in the "previous doer" group we choose $\hat{H}_i = \hat{\beta}X_i$ in Eq. (3.5). The regression equation in Table 3.1 was fit to the "previous doer with good validity" group. In the estimates presented below, both the good validity and poor validity "previous doer" groups are shrunk independently toward this regression surface. The offenders with poor validity were excluded because their values for dependent variables are considered untrustworthy. However, application of the regression equation to their values of independent variables could potentially yield reasonable estimates of their crime commission rates.

Note that the shrinkage is actually taking place with respect to Y_i and is then translated to shrinkage on Z_i upon division by T_i. Also, in shrinking toward the regression surface, the "shrinkage" can take place in either direction. That is, if $\hat{H}_i$ is greater than H_i, $\hat{\lambda}(i)$ will be greater than Z_i; otherwise it will be the same as Z_i or less. In particular, if Z_i is zero for offender i in the "previous doer" group, $\hat{\lambda}(i)$ is very likely to be greater than zero, since $\hat{H}_i$ is usually positive. On the other hand, shrinkage will always be toward zero for the previous nondoer group, and, in particular, $\hat{\lambda}(i)$ is the same as Z_i if $Z_i = 0$.

As we stated earlier, considerable gains can be made by applying the estimator separately to relatively homogeneous groups with respect to the variable of interest. We have identified three such groups: previous nondoers, previous doers with good validity, and previous doers with poor validity. Although the zero group estimates are shrunk toward zero and the nonzero group estimates are shrunk toward the regression surface, we apply the estimator to each of the three groups independently.

Estimates are tabulated separately for each group and each crime type. Tables 3.2, 3.3, and 3.4 present estimates for the three groups for person robbery. The number of observations in the group and the values of R and S as defined in Eq. (3.4) are given at the head of each table. The column headed T is street time (midpoint); Y is the number of crime commissions (midpoint); Z = Y/T; and Hudson-Tsui is the value, $\hat{\lambda}$, of the Hudson-Tsui estimator. A total of 350 instances of person robbery were reported in the previous nondoer group, so only a representative subsample of this group is shown in Table 3.2.

Table 3.2 shows that for the previous nondoer group, values of Z_i equal to zero are unmoved, while positive values are pulled closer to zero, as expected. The amount of shrinking depends critically on the values of R and S and, more specifically, on R/S. In turn, the values of R and S depend not only on the sample values, but on the values $\hat{H}_i$ used in the estimator as well. As explained earlier, it is most beneficial to use shrinkage estimators when the H_i are close to $\hat{H}_i$. With this in mind, it is interesting to compare values generated using different values of $\hat{H}_i$ within a group.

While Tables 3.2 through 3.4 show only one version of the Hudson-Tsui estimator for each of the three groups, it is possible to calculate other alternatives and compare them. Table 3.5 presents information on three different versions of the Hudson-Tsui estimates. It shows an estimated lower bound on the improvement in mean square error that the Hudson-Tsui shrinkage estimators give over the maximum-likelihood estimators, Y_1, ... Y_n (Hudson, 1980), and it shows an estimated bound on the percentage gain achieved, by comparing the lower bound of the mean-square-error gain of the Hudson-Tsui estimator to the mean square error

Table 3.2

PERSON ROBBERY DATA FOR 48 PREVIOUS NONDOERS
(Total observations = 350)

R = 51.0
S = 275.020508

T	Y	Z	Hudson-Tsui[a]
14.00000	0.0	0.0	0.0
15.00000	0.0	0.0	0.0
8.00000	1.00000	0.12500	0.10667**
5.00000	0.0	0.0	0.0
10.00000	0.0	0.0	0.0
21.00000	0.0	0.0	0.0
24.00000	5.50000	0.22917	0.21471*
22.00000	0.0	0.0	0.0
14.00000	6.00000	0.42857	0.40291*
22.00000	1.00000	0.04545	0.03879**
14.00000	2.00000	0.14286	0.12715**
13.00000	3.00000	0.23077	0.21009*
19.50000	16.00000	0.82051	0.79509
8.00000	1.00000	0.12500	0.10667**
24.00000	0.0	0.0	0.0
16.00000	5.50000	0.34375	0.32206*
24.00000	0.0	0.0	0.0
10.00000	0.0	0.0	0.0
15.00000	0.0	0.0	0.0
15.00000	0.0	0.0	0.0
14.00000	0.0	0.0	0.0
4.00000	0.0	0.0	0.0
3.00000	0.0	0.0	0.0
12.00000	0.0	0.0	0.0
21.00000	0.0	0.0	0.0
14.00000	0.0	0.0	0.0
15.00000	0.0	0.0	0.0
19.00000	1.00000	0.05263	0.04491**
10.00000	0.0	0.0	0.0
8.00000	0.0	0.0	0.0
24.00000	0.0	0.0	0.0
14.00000	0.0	0.0	0.0
22.00000	0.0	0.0	0.0
13.50000	0.0	0.0	0.0
16.00000	0.0	0.0	0.0
17.00000	0.0	0.0	0.0
19.00000	0.0	0.0	0.0
19.00000	0.0	0.0	0.0
24.00000	0.0	0.0	0.0
18.50000	7.00000	0.37838	0.35783*
17.00000	0.0	0.0	0.0
22.50000	0.0	0.0	0.0
12.00000	6.00000	0.50000	0.47006*
8.00000	0.0	0.0	0.0
23.50000	9.00000	0.38298	0.36533
21.00000	1.00000	0.04762	0.04064**
11.00000	0.0	0.0	0.0
9.00000	0.0	0.0	0.0

[a]Estimates followed by double asterisks indicate at least 10 percent shrinkage; a single asterisk indicates at least 5 percent. If y = 0, a single asterisk indicates that $\hat{\lambda} > .01$.

Table 3.3

PERSON ROBBERY DATA FOR 53 PREVIOUS DOERS WITH GOOD VALIDITY
(Total observations = 53)

R = 23.0
S = 158.486069

T	Y	Z	Hudson-Tsui[a]
13.50000	1.00000	0.07407	0.06463**
6.00000	1.00000	0.16667	0.15444*
10.50000	2.00000	0.19048	0.16974**
12.00000	0.0	0.0	0.03529**
17.50000	43.00000	2.45714	2.43930
15.00000	0.0	0.0	0.01279**
13.00000	1.00000	0.07692	0.08403*
6.00000	2.00000	0.33333	0.35237*
20.00000	84.00000	4.20000	4.18830
19.50000	322.50000	16.53845	16.51846
5.00000	0.0	0.0	0.05048**
6.00000	0.0	0.0	0.0
14.00000	0.0	0.0	0.02952**
18.00000	1.00000	0.05556	0.06536**
16.00000	0.0	0.0	0.00061
3.00000	3.00000	1.00000	0.99973
10.00000	4.00000	0.40000	0.38044
13.50000	0.0	0.0	0.00338
23.00000	0.0	0.0	0.00873*
5.00000	6.00000	1.20000	1.14920
21.00000	464.39990	22.11427	22.09216
19.50000	8.00000	0.41026	0.40621
9.00000	424.00000	47.11110	47.04543
7.00000	1.00000	0.14286	0.14121
11.00000	5.00000	0.45455	0.46782
1.00000	0.0	0.0	0.16021**
17.00000	0.0	0.0	0.01331**
12.00000	5.50000	0.45833	0.46675
20.00000	5.50000	0.27500	0.27589
10.00000	4.00000	0.40000	0.39054
19.00000	3.00000	0.15789	0.15694
8.00000	7.00000	0.87500	0.83324
13.00000	0.0	0.0	0.01730**
13.00000	154.79993	11.90769	11.89161
15.00000	0.0	0.0	0.01279**
23.50000	0.0	0.0	0.01091**
5.00000	0.0	0.0	0.06287**
16.00000	0.0	0.0	0.02664**
9.50000	3.00000	0.31579	0.32737
11.00000	8.00000	0.72727	0.69142
21.00000	10.00000	0.47619	0.47562
17.00000	0.0	0.0	0.02541**
24.00000	6.00000	0.25000	0.25595
13.50000	275.19995	20.38518	20.36606
10.00000	0.0	0.0	0.01464**
16.00000	4.00000	0.25000	0.24147
10.00000	0.0	0.0	0.02398**
15.00000	8.00000	0.53333	0.51468
14.00000	0.0	0.0	0.00724*
13.50000	232.19995	17.19998	17.16554
16.00000	318.00000	19.87500	19.85393
4.00000	2.00000	0.50000	0.46812*
22.00000	0.0	0.0	0.01155**

[a]Estimates followed by double asterisks indicate at least 10 percent shrinkage; a single asterisk indicates at least 5 percent. If y = 0, a single asterisk indicates that $\hat{\lambda} > .01$.

Table 3.4

PERSON ROBBERY DATA FOR 29 PERVIOUS DOERS WITH POOR VALIDITY
(Total observations = 29)

R = 11.0
S = 188.600494

T	Y	Z	Hudson-Tsui[a]
20.00000	40.00000	2.00000	1.99312
16.00000	344.00000	21.50000	21.48146
14.00000	129.00000	9.21428	9.18725
14.50000	42.59999	2.93793	2.92227
20.00000	1.00000	0.05000	0.05885**
20.00000	25.50000	1.27500	1.26671
13.00000	0.0	0.0	0.01961**
21.00000	0.0	0.0	0.00538*
11.00000	0.0	0.0	0.03320**
20.00000	6.00000	0.30000	0.30717
16.00000	2.00000	0.12500	0.13266*
13.00000	45.50000	3.50000	3.48816
6.00000	0.0	0.0	0.04670**
11.00000	0.0	0.0	0.02394**
18.00000	8.00000	0.44444	0.44585
10.50000	14.00000	1.33333	1.31194
12.50000	1.00000	0.08000	0.09210**
12.50000	180.00000	14.40000	14.38551
12.00000	5.50000	0.45833	0.45417
12.50000	0.0	0.0	0.00110
12.00000	4.00000	0.33333	0.33482
20.00000	326.80005	16.34000	16.32600
10.50000	5.50000	0.52381	0.51675
22.00000	958.89990	43.58635	43.56403
16.00000	9.00000	0.56250	0.55233
3.00000	5.50000	1.83333	1.76606
2.00000	0.0	0.0	0.02532**
16.00000	0.0	0.0	0.01248**
13.00000	27.95000	2.15000	2.13304

[a]Estimates followed by double asterisks indicate at least 10 percent shrinkage; a single asterisk indicates at least 5 percent. If y = 0, a single asterisk indicates that $\hat{\lambda} > .01$.

of the maximum-likelihood estimator. This mean square error is

$$\frac{1}{n} \sum_{i=1}^{n} \lambda_i T_i ,$$

which must be estimated because its value depends on the unknown parameters $\lambda_1, \ldots, \lambda_n$. One easy way to estimate the mean square error of the maximum-likelihood estimator is

$$\widehat{MSE} = \frac{1}{p} \sum_{i=1}^{p} Y_i . \qquad (3.6)$$

This is the denominator used in computing the bound on the estimated percentage gain shown in Table 3.5.

Table 3.5

A COMPARISON OF MEAN-SQUARE-ERROR-GAIN BOUNDS OF DIFFERENT VERSIONS OF THE HUDSON-TSUI ESTIMATOR FOR THREE GROUPS OF PERSON ROBBERS

	Group					
			Previous Doers			
	Previous Nondoers		Good Validity		Poor Validity	
Shrinkage Center	MSE Gain Bound	% Gain	MSE Gain Bound	% Gain	MSE Gain Bound	% Gain
Zero	.021	.16	.043	.09	.031	.01
Mean	.002	.02	.032	.07	.025	.01
Regression estimate	.006	.04	.063	.14	.057	.02

NOTE: The first number in the entries in the table is the estimated bound on the mean-square-error gain, given by $(1/n)R^2/S$. The second number is the estimated percent-gain bound given by 100 x $R^2/(S\Sigma Y_i)$. All regression estimates are derived from the group of previous doers with good validity.

Table 3.5 shows that the lower bound on the mean-square-error gain may not be sharp--that is, it may not be very close to the actual gain. (Unfortunately, good estimates of the gains are not available.) But the relative sizes of the bounds show that the regression center of shrinkage is best for the two previous doer groups, and the zero center of shrinkage is best for the previous nondoers.

We shall now compare the amount of shrinking between groups of data. Regression estimates are employed to set values of $\hat{H}_i$ for both of the previous doer groups (good validity and poor validity), so we can easily compare them. One would expect the good validity group to be rather more homogeneous than the poor validity group. One would also expect the values H_i of the good validity group to be closer on the average to $\hat{H}_i$, since the H_i's are the very values used in estimating the regression. We therefore expect to realize relatively more shrinkage and higher mean-square-error gains when estimating in the good validity group. This is in fact the case for person robbery, where $\frac{1}{n} R/S^2 = .0630$ for the good validity group, and $\frac{1}{n} R^2/S = .0568$ for the poor validity group.

The Hudson-Tsui estimator involves shrinking the transformed counts H_i toward a regression surface and then reversing the transformation. Hudson (1980) proves that this estimator is superior to Y_i itself for estimating $\lambda(i)T_i$ in the sense that it has a smaller mean square error of estimation *for all* values of $\lambda(i)$, if $n \geq 3$. Hudson's (1980) empirical comparison of this estimator to one that is analogous to the empirical Bayes estimator (Eq. (3.3)) shows that the Hudson-Tsui estimator (Eq. (3.4)) dominates, although not by much. Since the empirical Bayes estimator does not necessarily dominate Y_i for *all* $\lambda(i)$, we conclude that the Hudson-Tsui estimator is the best that can be used in these circumstances.

While the Hudson-Tsui estimator is not derived directly from explicitly considering a probability distribution on the λ's, Hudson (1980) shows that his estimator is quite close to one derived from a stochastic model in which $\lambda(i)T_i$ varies around a predicted value from a regression in a lognormal-like way. He compared the empirical Bayes estimator from this lognormal model to his "H" estimator and showed

that they give similar values. Since the lognormal distribution and gamma distribution are similar, it is not surprising that as a practical matter, the gamma-Poisson empirical Bayes estimators and the Hudson-Tsui estimators are similar.

IV. MULTIVARIATE MODELING: THE RELATIONSHIPS AMONG SEVERAL CRIME TYPES

The methods presented thus far consider each crime type individually. In analyzing and modeling the behavior of groups of offenders, it is important to understand how crime rates for different crime types are related. That is, some, but not all, burglars do robberies. Some robbers deal in drugs, etc. Depending on the crime types being analyzed, the interrelationships will vary. It is apparent, from the data we obtained from the Rand Second Inmate Survey, that there are varying amounts of overlap in doing different types of crimes. Table 4.1 shows the number of individuals in our subsample who said they did zero through 8 of the different crime types. Even this crude summary shows that a fairly rich family of distributions is needed to represent multivariate frequencies of crime commissions.

Table 4.1

NUMBER OF TYPES OF CRIMES DONE BY INDIVIDUAL OFFENDERS (Total sample = 440)

Number of Crime Types	Number of Offenders
0	79
1	85
2	84
3	66
4	57
5	36
6	18
7	11
8	4

Modeling the joint distribution of commissions of several different crime types is substantially more difficult than modeling the distribution of single crime types because of the need for a flexible, analytically tractable representation of the dependence structure across crime

types. Further, the "zero problem" is not amenable to the sample-truncation method, as it was in the univariate case. For K crime types, there are $2^K - 1$ potential combinations of "zeros," ranging from doing all but one type of crime to doing none of them. Separate truncation estimation for each crime type does not produce a consistent multivariate model.

In this section we attempt to solve the above problems. The zeros are handled in the same way as in Sec. III--we include only those individuals who have a prior history of committing one of the types of crime in question. That is, the data for the measurement period are not truncated. Assuming that the individual crime types have univariate negative binomial distributions (the gamma-Poisson process), we derive a multivariate generalization of that distribution. Finally, we fit this distribution to three crime types from the Rand Second Inmate Survey.

THE GENERALIZED MULTIVARIATE NEGATIVE BINOMIAL DISTRIBUTION

Suppose there are data for n individuals and K types of crime. Individual i is observed for period T_i, during which he commits Y_{ik} crimes of type k. We shall now develop a generalization of the univariate negative binomial distribution and present a method of estimating the parameters of this distribution. We call this distribution the generalized multivariate negative binomial (GMNB) distribution.

Assumptions. Let $\underset{\sim}{Y}_i' = (Y_{i1}, \ldots, Y_{iK})$ and $\underset{\sim}{\lambda}(i)' = (\lambda_1(i), \ldots, \lambda_K(i))$. Then $Y_{ik} | \lambda_k(i) \overset{ind}{\sim} \text{Poisson}(\lambda_k(i)T_i)$; i = 1, ..., n; k = 1, ..., K; and

$$\lambda_{ij} = \gamma_{i0} + \gamma_{ij} \qquad \text{where } \gamma_{i\ell} \overset{ind}{\sim} \text{gamma}(\alpha_\ell, \beta) \text{ for } \ell = 0, \ldots, K.$$

The notation "$\overset{ind}{\sim}$" means "is independently drawn from." Note that the distribution of $\underset{\sim}{\lambda}$ is multivariate gamma, thus allowing for correlations of $\lambda_k(i)$ and $\lambda_\ell(i)$, individual i's propensity for committing type k and type ℓ crimes, respectively. For reference, the mean propensity for crime type k is $(\alpha_0 + \alpha_k)/\beta$, the variance is $(\alpha_0 + \alpha_k)/\beta^2$, and the covariance between λ_{ik} and $\lambda_{i\ell}$ is α_0/β^2.

In Appendix B, we derive the probability-generating function and the moments of the GMNB distribution. The moments are given by

$$E(Y_{ik}) = \frac{(\alpha_0 + \alpha_k)T_i}{\beta},$$

$$Var(Y_{ik}) = \frac{(\alpha_0 + \alpha_k)T_i(T_i + \beta)}{\beta^2},$$

and

$$Cov(Y_{ik}, Y_{i\ell}) = \frac{\alpha_0 T_i(T_i + \beta)}{\beta^2} \quad \text{for } k \neq \ell.$$

Appendix B also gives an iterated weighted least-squares method for estimating the parameters $\alpha_0, \ldots, \alpha_k$ and β.

AN APPLICATION OF THE GMNB DISTRIBUTION

The data from the Rand Second Inmate Survey will be used to illustrate the fitting of the GMNB distribution. Let Y_{i1}, Y_{i2}, Y_{i3} be the number of business robberies, person robberies, and frauds, respectively, that individual i reports in period T_i. We have restricted ourselves to the 105 individuals who have committed at least one of these three crime types in the period prior to the measurement period. The GMNB distribution was fit to these data using the method outlined above and described more fully in Appendix B. The estimated parameter values are given in Table 4.2.

It is interesting to note that the pairwise correlations implied by the values of $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$ are close to 0.1 for all pairs. These are shown in the correlation matrix in Table 4.2.

It was shown in Sec. II that the univariate negative binomial may not give a good fit to the reported counts of crimes. The GMNB distribution imposes the additional constraint that the scale parameters of each of the marginal univariate distributions be the same. Thus, *a priori*, we might expect poor agreement between the data and the fitted GMNB

Table 4.2

THE GMNB DISTRIBUTION FIT TO THREE CRIME TYPES

Crime Type, k	$\hat{\alpha}_k$	$\hat{\mu}_k$
1	0.03432	-0.18219
2	0.04390	0.34103
3	0.03475	-0.15883

$\hat{\alpha}_0$ = 0.00387, $\hat{\beta}$ = 0.01831, $\hat{\bar{\mu}}$ = 2.26810

Correlation matrix estimated from the model:

1.000	0.09	0.10
0.09	1.00	0.09
0.10	0.09	1.00

distribution. In addition, poor agreement might be due to fitting the full GMNB distribution rather than some truncated version, as in the univariate case.

Approximate chi-square tests were computed to check the fit of each of these marginal univariate distributions. These were all statistically significant, confirming our expectations. The contributions to chi-square indicate that if the zero problem could be handled better, the fit might well improve. Thus, there is reason to hope for better fits with other data sets.

ALTERNATIVES TO THE GMNB DISTRIBUTION

There are several possible ways of generalizing the GMNB distribution to get a richer, more flexible family of distributions. Such generalization may be desirable when working with data sets that are not fit well by the GMNB distribution. One of the possible methods is described below.

The GMNB distribution constrains the scale parameter β of the univariate NB distributions to be the same for all crime types. It also constrains the underlying propensities to be symmetric, in that there is one unique gamma variable γ_{i0} shared by all crime types. We use some concepts from the literature on individuals' proneness to accidents (e.g., automobile accidents) to present a generalization of the univariate negative binomial distribution (called the negative binomial beta) that allows for a more general distribution of the underlying λ's and to indicate how this distribution can be made into a multivariate distribution.

For the univariate case, we can assume that the gamma scale parameter β follows some distribution. If we choose a "Beta II" density for β, namely,

$$f(\beta) = \frac{1}{B(\delta, \gamma)} \frac{\beta^{\delta-1}}{(1+\beta)^{\delta+\gamma}}, \qquad 0 < \beta < \infty \quad \delta,\gamma > 0 \tag{4.1}$$

where

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx,$$

the resulting final distribution of counts is called a "negative binomial beta" or a "generalized Waring" distribution:

$$\Pr[Y = y] = \frac{\Gamma(\delta+\gamma)\Gamma(\alpha+\gamma)}{\Gamma(\gamma)\Gamma(\alpha)\Gamma(\delta)} \frac{\Gamma(y+\alpha)\Gamma(y+\delta)}{\Gamma(y+1)\Gamma(y+\alpha+\delta+\gamma)} \tag{4.2}$$

where

$$\delta > 0,\ \gamma > 0,\ \alpha > 0$$
$$y = 0, 1, 2, \ldots .$$

Irwin (1968) used these assumptions to model numbers of accidents experienced by individuals. He showed that the variance of the negative binomial beta distribution can be split into three components arising from

1. Differences in an individual's "proneness" to have an accident (internal factors).

2. Differences in an individual's "liability" to have an accident (external factors/environment).
3. Random fluctuation.

(In the case of criminal behavior, "proneness" corresponds to inherent propensity to commit crime, and "liability" corresponds to opportunity.) Irwin regarded the negative binomial as the distribution for individuals with proneness β arising from a Poisson distribution with parameter $\lambda|\beta$, where $\lambda|\beta$ represents an individual's liability λ at a given fixed level of proneness β, and where $\lambda|\beta \sim \text{gamma}(\alpha, \beta)$. Irwin then let proneness β have the Beta II distribution (Eq. (4.1)). The variance of the final generalized Waring distribution is then the sum of three variance components due to proneness, liability, and randomness. Other properties of this distribution are also described by Irwin (1975).

In applying this model to counts of crimes committed by individuals, the implied distribution of crime propensities is not a gamma distribution but is a three-parameter distribution called the gamma product-ratio distribution GPR (α, δ, γ) with density

$$f(w) = \frac{\Gamma(\alpha+\gamma)\Gamma(\delta+\gamma)}{\Gamma(\alpha)\Gamma(\delta)\Gamma(\gamma)} w^{\alpha-1}\, U(\alpha+\gamma, \alpha-\delta+1, w),$$
$$0 < w < \infty \quad \alpha,\delta,\gamma > 0 \tag{4.3}$$

where U is a solution for Kummer's equation and is defined by

$$U(\alpha+\gamma, \alpha-\delta+1, w) = \frac{1}{\Gamma(\alpha+\gamma)} \int_0^\infty e^{-wt} \frac{t^{\alpha+\gamma-1}}{(1+t)^{\delta+\gamma}}\, dt. \tag{4.4}$$

This distribution is naturally more flexible than the gamma for fitting crime commission propensities.

A multivariate generalization of the Waring distribution (also called the multivariate inverse Polya-Eggenberger distribution) has been described and analyzed by Sibuya (1980). Its frequency function is

$$P(\underset{\sim}{Y} = \underset{\sim}{y} | \underset{\sim}{\alpha}, \beta, \gamma) = \frac{\Gamma(\alpha_+ + \gamma)\Gamma(\beta + \gamma)}{\Gamma(\alpha_+ + \beta + \gamma)\Gamma(\gamma)} \cdot \frac{(\beta, y_+)}{(\alpha_+ + \beta + \gamma, y_+)} \cdot \prod_{i=1}^{K} \frac{(\alpha_i, y_i)}{y_i!}, \tag{4.5}$$

where

$$y_i = 0, 1, 2, \ldots$$

$$y_+ = \sum_{i=1}^{K} y_i, \qquad y_+ > 0, \qquad \alpha_+ = \sum_{i=1}^{K} \alpha_i, \qquad \alpha_i > 0$$

and $(\beta, y) = \beta(\beta + 1) \ldots (\beta + y - 1)$. Additional properties of this distribution are presented in Sibuya (1980), including a treatment of truncating zero. The parameters can be estimated by maximum likelihood in a straightforward way. Thus, Eq. (4.5) offers an attractive generalization of the GMNB distribution that appears to be an excellent candidate for future work on modeling crime counts.

V. EXTRAPOLATION TO TARGET POPULATIONS OF OFFENDERS

Thus far, we have treated the data as if the distribution of crime rates for the sampled offenders was itself of general interest. This would be true if the sample had been drawn randomly from a large population--for example, all offenders in a given city or county. However, rarely if ever can data about crime commissions be obtained from a representative population. Data about *arrests* for a representative population are more easily obtained, especially in jurisdictions that compile criminal career histories from official records.

For practical reasons, self-reported data on the number of crimes of various types committed by individuals have been systematically collected only from offenders who have come into official contact with the criminal justice system. Such persons are necessarily nonrepresentative of offenders in general, because the probability that a particular offender will be arrested (or convicted, or incarcerated) in a given time period depends on the types of crimes he commits, how often he commits them, and various other factors. The Rand survey respondents, whose data have been used here, are clearly nonrepresentative, because they were all incarcerated in prison at the time they answered the survey questions.

The general impression of incarcerated offenders--and especially imprisoned offenders--is that they are an extraordinarily atypical group, the "losers" among criminals.* For this reason, most people who are interested in individual crime commission rates want to know the distribution of rates for groups other than prisoners. For example, prosecutors might be interested in the crime commission propensities of *arrestees* or the characteristics of high-crime-propensity arrestees. Researchers concerned with deterrence or the sociodemographic factors related to crime would be more interested in a general population of active offenders than in arrestees. Judges, on the other hand, might

---

*Wolfgang called them a "subservience of captured and confined sinners" (private communication, 1979).

like to know the characteristics of convicted persons with high crime commission propensities.*

In order for data collected from one group of offenders, such as prisoners, to be extrapolated into estimates of distributions for some other target population, it is not necessary for the sample to be collectively, or on the whole, a representative group. Rather, it is only necessary that members of the target population have a nonzero probability of appearing in the sampled population and that the sampling probabilities of respondents can be estimated.

We have therefore developed models for estimating a surveyed offender's sampling probability with respect to a target population that differs from the surveyed population in specified ways. By weighting each respondent's estimated crime commission propensity according to his sampling probability, we obtain an empirical estimate of the distribution of crime commission propensities in the target population.

In extrapolating distributions of crime commission rates from sampled populations to target populations, the sampling probability is not necessarily strongly related to the crime in question. For example, suppose the target population includes auto thieves. In most states an arrest for auto theft is very unlikely to lead to a prison sentence, so it might be argued that auto thieves have practically no chance of being in a sampled population of prisoners. But in fact, some offenders who are primarily auto thieves wind up in prison upon conviction for assault, manslaughter, or other crimes. Such offenders provide information about the distribution of crime commission rates for auto theft, even though they are not incarcerated for auto theft.

The following discussion gives several examples of target populations and sampled populations. The first example has fairly uncomplicated mathematical assumptions, but it does not correspond to the Second Rand Inmate Survey sample.

*"Convicted persons" is not synonymous with "prisoners," because many convicted persons do not go to prison.

MODELS FOR A SAMPLED ARREST COHORT

Parameters of the Model

Consider a sampled population drawn randomly from an arrest cohort, that is, from all individuals who were arrested in a given jurisdiction in a given time period (say, 1 year).* The target population is defined to consist of everyone who had a nonzero chance of being in the cohort, i.e., everyone who had a nonzero commission propensity for crimes for which it is possible to be arrested in that jurisdiction.

Assume that when offender i is not incarcerated, his arrests occur according to a Poisson process with rate $\mu(i)$. This situation could arise in many different ways. For example, offender i commits crime type k according to a Poisson process with rate $\lambda_k(i)$, $k = 1, \ldots, K$, and each of his commissions of crime k results in an arrest with probability $Q_{k1}(i)$, independent of prior events. In this case, his arrest rate is

$$\mu(i) = \sum_k \lambda_k(i) Q_{k1}(i).$$

This model is quite general, since it does not assume that offender i has the same arrest probability for different crimes, nor does it assume that different offenders have the same arrest probability for the same crime k. Nonetheless, the model may not be valid for several reasons, including possible invalidity of the Poisson assumption for crime commissions, as discussed in Sec. II, and possible increases or decreases in arrest probabilities immediately following an arrest. (Once an offender has been arrested for crime k, the police might be more likely to pick him up for subsequent crimes of the same type; or he may gradually learn what he does wrong and take steps to decrease his arrest probability.)

*As a practical matter, it may be difficult to obtain data from arrestees, since they have merely been charged by the police with committing a crime and have not necessarily been prosecuted or convicted. However, the example is intended only to illustrate the method.

Actual data collected from offenders ordinarily show that each individual has only a small number of arrests, so it is practically impossible to test at an individual level whether there is any correlation between an offender's arrests at one time and his prior arrests. Moreover, since arrests correspond to a small subset of crime commissions, a so-called "thinning process," they are closer to a Poisson process than the original crimes are (Maltz, 1980). Consequently, the assumption we have adopted--that the sequence of arrests for offender i constitutes a Poisson process with rate $\mu(i)$ when he is free--will be generally consistent with the data.

Given information about the number of arrests for crimes of type k ($k = 1, \ldots, K$) for each sampled offender during the measurement period, the total arrest rate $\mu(i)$ for offender i can be estimated by shrinkage methods analogous to those described in Sec. III for estimating crime commission propensities $\lambda_k(i)$. The key question of judgment is how the "center of shrinkage" $\hat{\mu}_i$ is to be obtained. Two suitable procedures follow:

1. Estimate the average arrest probability $\overline{Q}_{k1}$ for crime k in the jurisdiction from data external to the survey of offenders, and take as the center of shrinkage

$$\hat{\mu}_i = \sum \hat{\lambda}_k(i)\ \overline{Q}_{k1},$$

where $\hat{\lambda}_k(i)$ is the previously estimated value of offender i's commission propensity for crime type k.

2. Consider the average arrest probabilities for the sampled offenders as coefficients $q_1, q_2, \ldots, q_k$ in a regression whose independent variables are $\hat{\lambda}_1, \ldots, \hat{\lambda}_K$, namely

$$\mu(i) = \sum q_k\ \hat{\lambda}_k(i) + \varepsilon_i,$$

where ε_i is the unexplained error for offender i. Then the center of shrinkage for offender i is his expected value of μ from this regression.

Although it would also be possible to shrink toward a rate estimated by taking into account the characteristics of each offender, this approach does not appear attractive thus far, since we have not yet found any independent covariates that are significantly related to arrest probabilities.

Probability of Being in the Sample

An offender who is a member of the target population is eligible for the arrest cohort sample if he is arrested during the year. If offender i happens to be on the street at the beginning of the year, then his probability of eligibility is simply

$$P(\text{elig}|\text{on the street}) = 1 - \exp(-\mu(i)).$$

However, there is a nonzero probability that he was incarcerated at the beginning of the year. To estimate this probability we introduce the cumulative distribution function $F_k(i)$ for the length of time "off the street" when offender i is arrested for crime type k. The distribution $F_k(i)$ includes mass at zero, reflecting the fact that the offender might not be incarcerated after an arrest, and is assumed to have finite mean $t_k(i)$. Although this distribution is permitted to vary according to the characteristics of offender i, the available data sources provide such distributions only for large groups, such as all arrested individuals in a jurisdiction or state.

Let F(i) be the cumulative distribution function for the incarceration time of offender i, given arrest, namely, the probabilistic mixture

$$F(i) = \sum \lambda_k(i) Q_{k1}(i) F_k(i)/\mu(i),$$

and let t(i) be its mean. (In practice it is ordinarily not possible to identify the dependence of $F_k(i)$ and $Q_{k1}(i)$ on i, but t(i) can differ from t(j) if the crime commission propensities of offender i are different from those of offender j.) Then, assuming a steady state, the probability that offender i is incarcerated at any given moment is

$$z(i) = \frac{\mu(i)}{\mu(i) + \frac{1}{t(i)}} .$$

Hence z(i) is the probability that offender i was incarcerated at the start of the year for which the arrest cohort is defined. Given that he is incarcerated, his probability of release by time x is

$$r(x) = \frac{1}{t(i)} \int_o^x S_F(u)\, du \qquad 0 \le x \le 1,$$

where S_F is the survivor function 1 - F(i). Given release at x, his probability of arrest in the year is 1 - exp(-μ(i)(1-x)). Hence, the overall probability of eligibility for the sample, given incarceration at the start of the year, is

$$P(\text{elig}|\text{incarcerated}) = \int_o^1 \frac{1}{t(i)} S_F(x)(1 - e^{-\mu(i)(1-x)})\, dx.$$

In sum, the probability that offender i is eligible for the sample is

$$P(\text{elig}) = \frac{1/t(i)}{\mu(i) + 1/t(i)} [1 = e^{-\mu(i)} + \mu(i) \int_o^1 S_F(x)(1 - e^{-\mu(i)(1-x)})\, dx].$$

The probability of his being in the sample is simply the overall sampling fraction times this P(elig). An illustrative application of this formula is given below.

SAMPLING FROM AN INCOMING COHORT TO INCARCERATION

Calculating Eligibility for the Sample

Instead of selecting from arrestees, we might select a sample of people who become incarcerated in a given jurisdiction during a given year. In this case, the target population consists of individuals with nonzero commission propensity for crimes for which it is possible to

be incarcerated. The above results for an arrest cohort apply here with the following modifications:

1. Replace $Q_{k1}(i)$ with $Q_k(i)$, the probability that offender i is incarcerated, given commission of crime type i. Consequently, $\mu(i) = \sum \lambda_k(i)Q_k(i)$ is the average rate at which offender i is incarcerated, or the inverse of his average time between incarcerations.
2. Interpret $F_k(i)$ to be the cumulative distribution function of incarceration time for crime k, given incarceration, and $t_k(i)$ as the mean incarceration for crime k, given incarceration. (Consequently, F(i) and t(i) are reinterpreted.)

For an incoming cohort to incarceration, it is somewhat reasonable to assume, as an approximation, that the distribution of incarceration time is negative exponential. Chaiken (1980) showed that the duration of prison terms in California in 1976 corresponded approximately to an offset exponential distribution. The offset can be interpreted as arising from the fact that individuals sentenced to terms of less than a year go to jail, not prison. Hence the overall distribution (including jail and prison) could well be exponential. (In 1976, California prisoners were sentenced to indefinite terms, with release determined by a parole board. Subsequently, California legislation was changed to provide for determinate sentences. The distributional form that was correct for California in 1976 is more likely to be valid for states having indeterminate sentences than for California at the present time.)

Under the exponential assumption, we have

$$S_F(u) = e^{-u/t},$$

where t stands for t(i), dropping the index i (below, we also use μ for μ(i)). Hence,

$$P(\text{elig}|\text{incarcerated}) = \frac{1}{t}\int_0^1 e^{-x/t}(1 - e^{-\mu(1-x)})\,dx$$

$$= 1 - e^{1/t} - \frac{1}{1-\mu t}(e^{-\mu} - e^{-1/t}) .$$

$$P(\text{elig}) = \frac{1}{1+\mu t}(1 - e^{-\mu})$$

$$+ \frac{\mu t}{1+\mu t}[(1 - e^{-1/t}) - \frac{1}{1-\mu t}(e^{-\mu} - e^{-1/t})]$$

$$= 1 - \frac{1}{1-(\mu t)^2}e^{-\mu} + \frac{(\mu t)^2}{1-(\mu t)^2}e^{-1/t}.$$

As a very different alternative to the exponential assumption, we can assume that all incarcerations have the same length $t_o > 1$ year, i.e.,

$$S_F(u) = \begin{Bmatrix} 1 & u \le t_o \\ 0 & u > t_o \end{Bmatrix}.$$

Then $P(\text{elig}) = \mu/(1 + \mu t_o)$, which is simply the steady-state probability of entering incarceration. (Under this assumption, it is impossible to enter incarceration twice or more in the sampling year.)

We have now derived two different equations for P(elig), based on two assumptions about the form of the distribution of incarceration time, and we have indicated how either of these probabilities can be estimated for an individual member of an incoming cohort to incarceration, using

- Estimates of his crime commission propensities for crimes 1, 2, ..., K.
- Estimates of the probabilities of incarceration and average incarceration times in the jurisdiction for crimes 1, 2, ..., K.

This is all the information we need to estimate the sampling probability of each member of the incoming cohort, relative to the target population. The sample can now be weighted to estimate characteristics of the target population.

<u>Example of Estimating Procedure</u>

Suppose, for simplicity, that there are only two crime types. The first (which we shall call robbery) will be assumed to have Q = (0.2)(0.4) = 0.08 and t_1 = 4 years; the second (which we shall call burglary) will have Q = (0.08)(0.3) = 0.024 and t_2 = 2 years. Different combinations of crime commission result in different values of mean incarceration time t, as well as of arrest rate $\mu = .08\lambda(\text{robbery}) + .024\lambda(\text{Burglary})$. The results for P(elig) are shown in Fig. 5.1.

Offenders with very low crime rates (say, less than one burglary per year and one robbery per 5 years) have less than a 2 percent chance of eligibility, while those with high crime rates have from about a 25 to 35 percent chance of eligibility. The two different models of the distribution of time served (exponential and constant) yield nearly identical estimates of P(elig) when both λ(robbery) and λ(burglary) are less than 10 per year, and they do not yield substantially different estimates of P(elig) for any values of the crime commission rates. Hence the details of the assumptions are immaterial in a practical sense.

As shown in Fig. 5.1, offenders who have high robbery rates as well as high burglary rates are less likely to be sampled than those who simply have high burglary rates (because high-rate robbers are more likely to be in prison at the start of the year).

<u>Example Showing Effect of Weighting on the Distribution of Crime Commission Propensities</u>

Assume that there is a group of offenders who commit only one crime type, and their distribution of crime commission propensities is gamma with mean 2 per year. Figure 5.2 shows an example of such a distribution with parameters $\alpha = 0.4$, $\beta = 0.2$. The more widely spread distribution represents the distribution of crime commission rates in an incoming cohort to incarceration. This distribution was obtained by

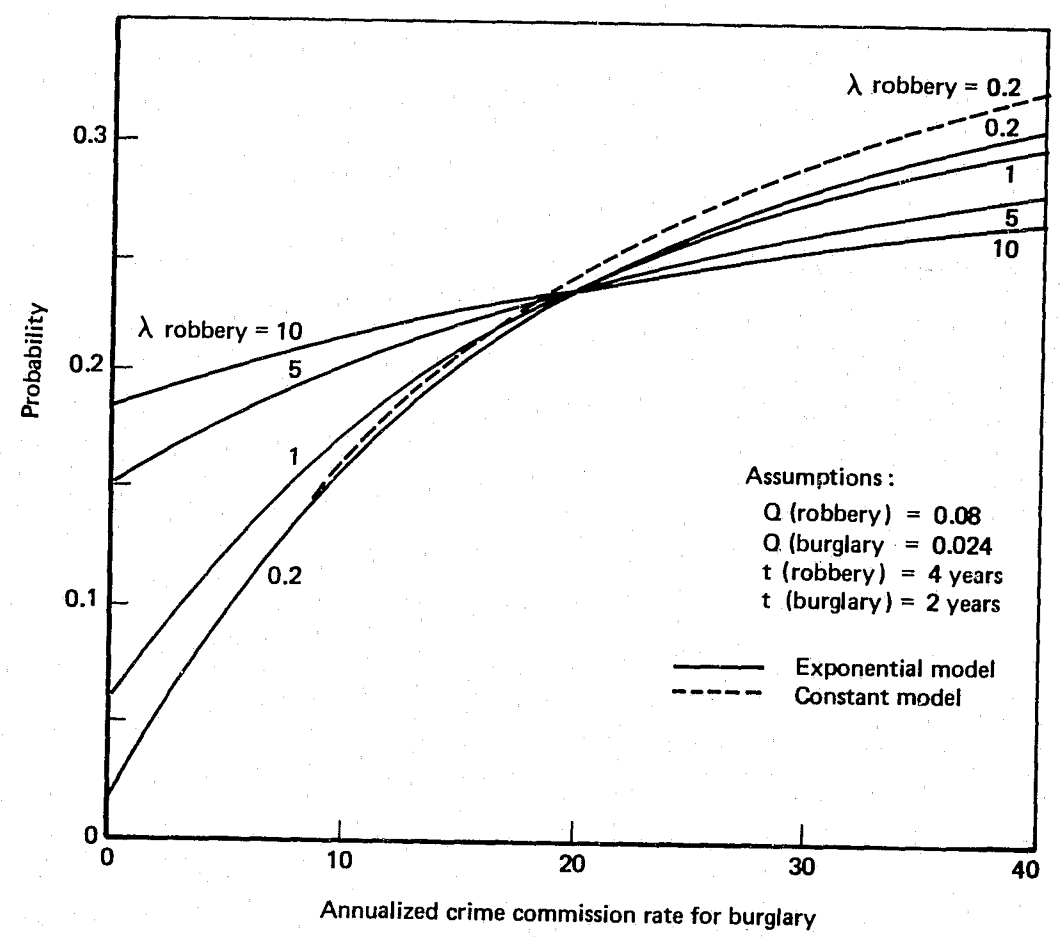

Fig. 5.1 — Illustrative probability of eligibility for sample (incoming cohort to incarceration)

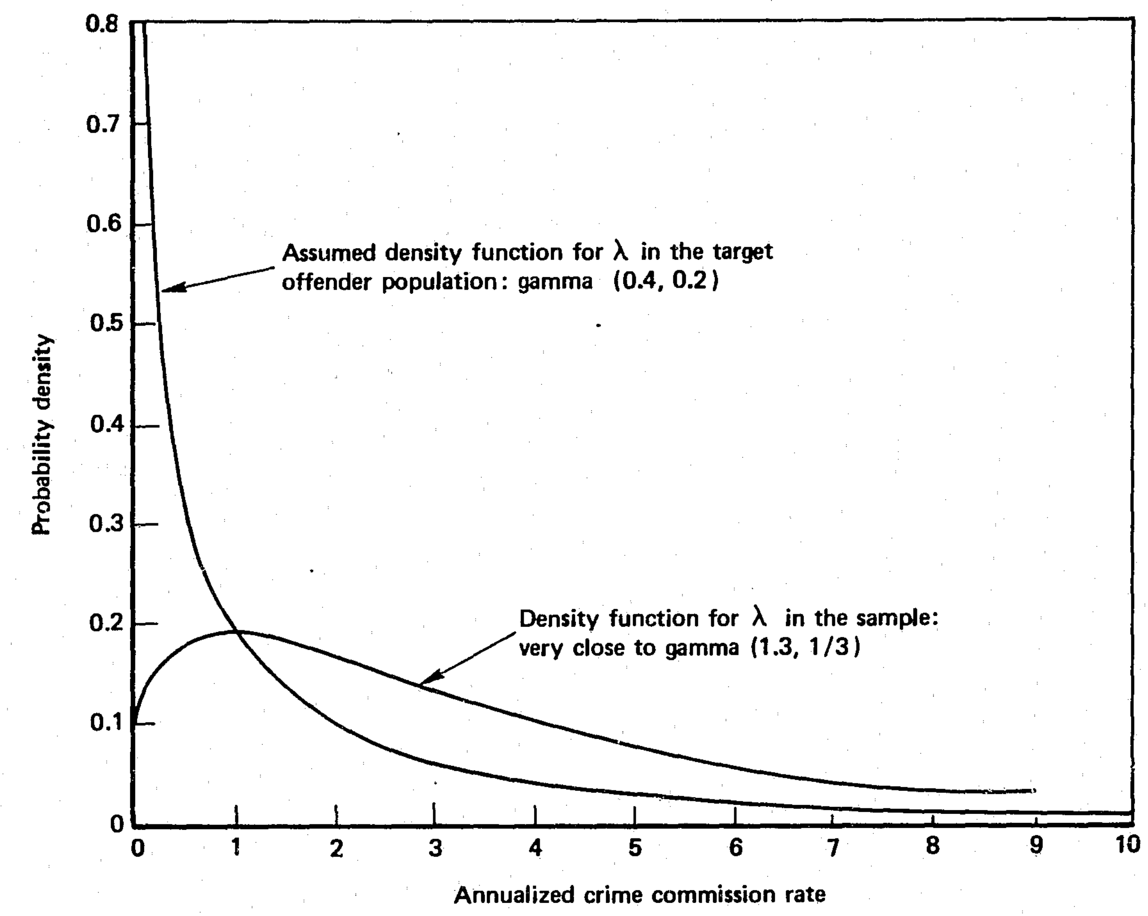

Fig. 5.2 — Application of model for incoming cohort to incarceration

applying the incoming-cohort model described above with the parameters $Q = 0.08$ and $t = 4$ years. Although the distribution of crime commission propensity in the sample cannot be exactly gamma (under our assumptions), it is almost indistinguishable from a gamma distribution with parameters (1.3, 1/3).

The distribution for the sample shows a distinct dropoff below $\lambda = 0.9$ even though 50 percent of the target offender population has $\lambda < 0.7$. An extraordinary concentration at low values of λ in the target population produces only a modest concentration at low values in the distribution of crime rates in the sampled population. Since typical distributions of crime commission rates in sampled populations have fairly heavy concentrations near zero, the implication of the model is that the target population has an even more dramatic concentration near zero and a less substantial tail.

AN IN-PRISON SAMPLE

The case of selecting a sample of offenders in prison was described by Chaiken (1980). A model for estimating an offender's probability of eligibility for such a sample must incorporate three additional features not included in the previous two models:

1. In an adult prison there are no prisoners under the age of 18, so there are strong age dependencies in the probability of 18-, 19-, and 20-year-olds being sampled in prison.
2. An offender may remain in prison past the time when he would have stopped doing crime if he weren't in prison. Hence the phenomenon of criminal career termination, which is somewhat irrelevant for the previous models, becomes important in this case. (To avoid confusion concerning whether an inactive offender is or is not in the target population, we will include all imprisoned individuals in the target population.)
3. The sentence lengths for members of the sample are typically longer than average, due to time-biased sampling. Hence, if we wish to estimate the parameters $t_1, t_2, \ldots, t_K$ (lengths of sentences) from the data for the sampled population, additional models are needed.

The third feature is illustrated by Fig. 5.3, which shows simulated criminal careers of 20 offenders, 10 of whom have a "low" crime rate of 9 per year and 10 of whom have a "high" rate of 18 per year. The light line indicates time not in prison; the heavy line, time in prison. Different offenders start committing crimes at different times in relation to the sampling date. For each crime, the probability of (arrest and) incarceration is assumed to be 0.02. The length of time spent in prison was simulated as an offset exponential distribution, with an average of 38 months and a minimum of 20 months. Figure 5.3 shows that high-rate offenders are oversampled in relation to low-rate offenders and that the sampled prisoners have much longer terms than the average term.

Adapting the notation above, let $\lambda_k(i)$ be the annualized crime commission propensity for offender i, $Q_k(i)$ his probability of (arrest, conviction, and) imprisonment, given commission of a crime of type k, $F_k(i)$ his distribution of staying time in prison, given imprisonment for crime type k, and $t_k(i)$ the mean of $F_k(i)$. As before, the distribution F(i) is defined as

$$F(i) = \Sigma\lambda_k(i)Q_k(i)F_k(i)/\mu(i),$$

where $\mu(i) = \Sigma\lambda_k(i)Q_k(i)$ is the inverse of the average time between *imprisonments*, and t(i) is the mean of F(i). Time spent incarcerated in facilities other than prison is ignored in this version of the model, since prison terms are substantially longer than jail terms, but it can be incorporated without much difficulty. The largest population consists of offenders, imprisoned or not, who have nonzero commission propensities for crimes for which it is impossible to be imprisoned, together with all other imprisoned offenders.

To account for termination of criminal careers, we assume that the duration of criminal activity for offender i has a distribution G(i), and an offender can become inactive either while on the street or while in prison.

To derive an analytic expression for the probability that offender i would be found in prison at a specified age, we must make

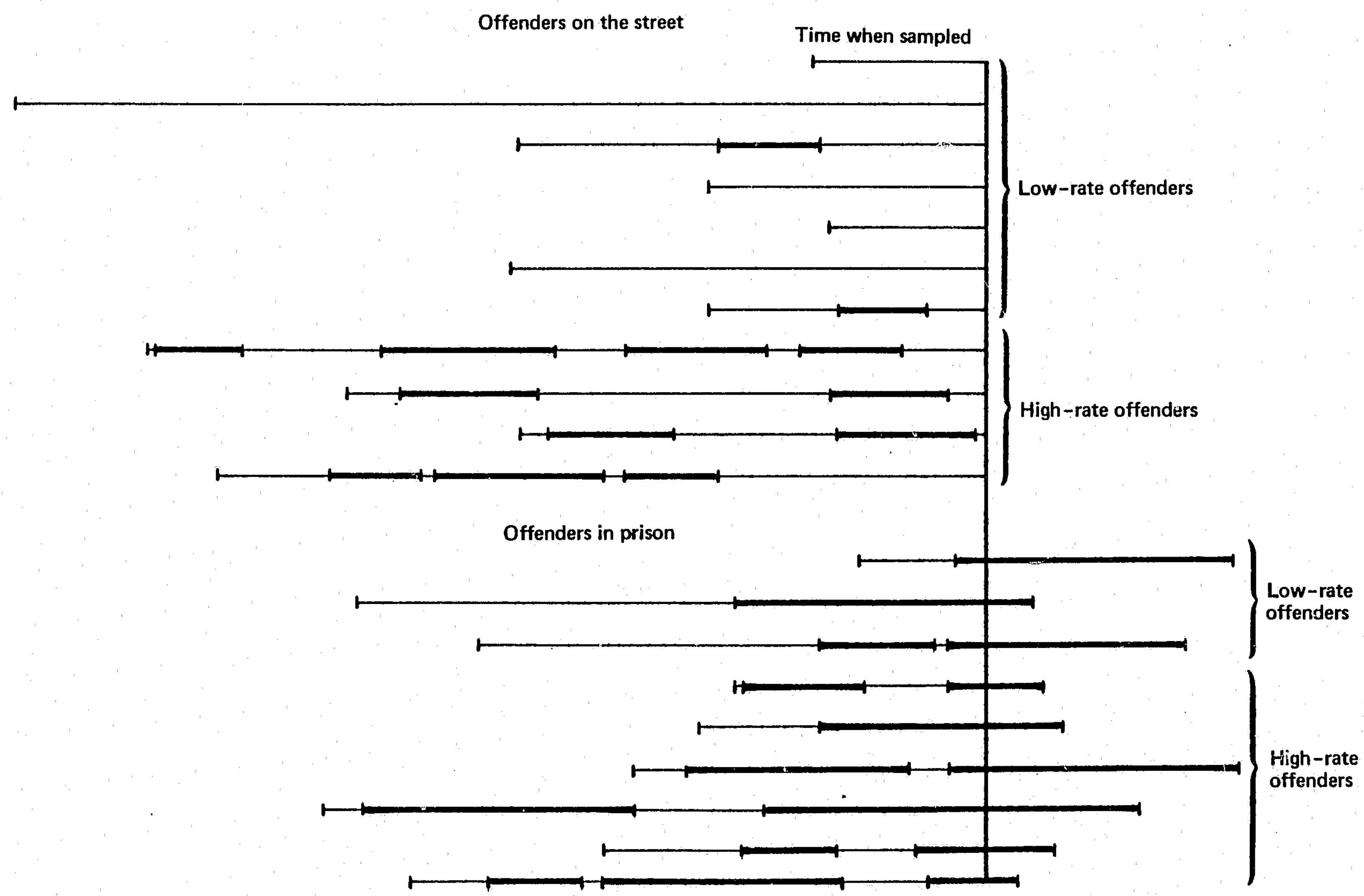

Fig. 5.3 — Simulated offenders selected for the sample if in prison on the sampling date

CONTINUED

1 OF 2

further simplifying assumptions. As an illustration, we adopt the following (not very realistic) assumptions, recognizing that the result will be only approximately correct:

- The distribution $F(i)$ of the length of time that offender i spends in prison is exponential.*
- Offender i ends his criminal activity after an exponentially distributed time whose mean is $1/\omega(i)$, unless his end-of-career time occurs when he is in prison. That is, offenders become inactive at a constant rate ($\omega(i)$ for offender i) when they are on the street.
- When in prison, offender i can be "labeled" as having become inactive at the same rate $\omega(i)$, but he remains a member of the target population until he leaves prison. (For the model to work, we must have $\omega(i) < 1/t(i)$.)
- All offenders start their criminal activity before age 18 and become eligible for (adult) imprisonment beginning with crimes committed after their 18th birthday.

Under these assumptions, the offender's status as a function of time can be modeled as a Markov process with these states: in prison, active on the street, and inactive. Denote by $P_1(18 + x)$ the probability that offender i is in prison at age 18 + x, and $P_0(18 + x)$ the probability that he is active and on the street. Then the assumptions listed above imply the following differential equations† for P_0 and P_1:

$$\frac{dP_0(18 + x)}{dx} = - (\omega(i) + \mu(i))P_0(18 + x) + (- \omega(i) + 1/t(i))P_1(18 + x),$$

$$\frac{dP_1(18 + x)}{dx} = - P_1(18 + x)/t(i) + \mu(i)P_0(18 + x),$$

$$P_0(18) = 1.$$

---

*This assumption is not literally possible, since $F(i)$ is a probabilistic mixture of distributions, and an exponential distribution is not a mixture of any plausible individual distributions.

†These are called the equations of detailed balance.

For those who are still members of the target population at age 18 + x, the fraction in prison is

$$\phi_i(18 + x) = \frac{P_1(18 + x)}{P_0(18 + x) + P_1(18 + x)} .$$

By solving the differential equations for P_0 and P_1, we obtain the result for ϕ_i:

$$\phi_i(18 + x) = \frac{\mu(i)}{\mu(i) - \omega(i) + 1/t(i)} (1 - \exp(-(\mu(i)-\omega(i)+1/t(i))x)).$$

Thus $\phi_i = 0$ at age 18 and gradually increases to the asymptotic value $\mu(i)/(\mu(i) - \omega(i) + 1/t(i))$. Ignoring the age of offenders, the average probability that offender i is in prison can be shown to be $\mu(i)t(i)/(1 + \mu(i)t(i))$.

Comparing the probabilities across offenders, we see that the functional form has the desired properties. All other things being equal:

- Offenders with high $\lambda_k(i)$ are oversampled in a prison population, compared to offenders with average or low $\lambda_k(i)$.
- Offenders with high incarceration probabilities $Q_k(i)$ are oversampled, compared to other offenders (because $\mu(i) = \Sigma\lambda_k(i)Q_k(i)$).
- Offenders who do crimes that yield long prison terms t(i) are oversampled, compared to other offenders.
- Old offenders are oversampled, compared to young ones.
- Offenders with short career lifetimes $1/\omega(i)$ are oversampled, compared to offenders with long lifetimes.

Since the quality of available data for estimating $Q_k(i)$, $t_k(i)$, and (especially) $\omega(i)$ is ordinarily rather poor, the improvements that could be made by introducing more realistic assumptions in the model do not appear to be warranted. Moreover, since the inverses of the

sampling probabilities are used as sampling weights, any improvement in the model that results in simply multiplying most offenders' sampling probabilities by approximately a constant factor has essentially no effect on the analysis. Chaiken (1980), applying these methods to data from the Rand First Inmate Survey, showed that the weighting scheme resulted in estimated mean λ_k's for active offenders that differed from the means for the in-prison sample by 60 to 80 percent; changing the details of the model affected the adjusted means by less than 10 percent. Hence, models that capture the essence of the sampling process under simplifying assumptions should ordinarily be sufficiently accurate for typical applications.

Appendix A
ESTIMATION OF PARAMETERS OF DISTRIBUTIONS

PARETO DISTRIBUTION

The Pareto cumulative distribution function is

$$F(x) = 1 - \left(1 - \frac{x}{\sigma}\right)^{-a},$$

where a is the shape parameter, and σ is the scale parameter. We estimated these two parameters by a method of maximum likelihood, using the censored data for annualized crime commission rates. Before processing the data, responses of "zero crimes" were converted to interval responses with lower terminal 0 and upper terminal 1/(12 T + 1). Four types of observations then entered the likelihood function:

1. Complete or exact observations for which the minimum rate equals the maximum rate.
2. Closed-interval observations with nonnegative left endpoint and finite right endpoint.
3. Left-closed and right-open interval observations (e.g., we know only that at least 11 crimes were committed).
4. Zero observations, which were converted to interval observations with left endpoint zero and right endpoint finite.

Suppose there are n_1 observations of the first type, n_2 of the second, n_3 of the third, and n_4 of the fourth. Then the likelihood function is given by

$$L = \prod_{i=1}^{n_1} f(x_i) \prod_{i=1}^{n_2} [F(r_i) - F(\ell_i)] \prod_{i=1}^{n_3} (1 - F(y_i)) \prod_{i=1}^{n_4} F(w_i), \qquad (A.1)$$

where x_i is the i^{th} observation of the first type, ℓ_i is the left endpoint and r_i the right endpoint of the i^{th} observation of the second

Preceding page blank

type, y_i is the left endpoint of the i^{th} observation of the third type, w_i is the right endpoint of the i^{th} observation of the fourth type, f is the density function, and F is the distribution function. The maximization was performed by a computer program employing an IMSL (International Mathematical and Statistical Library) subroutine using a quasi-Newton method for minimizing a function of any number of parameters.

Table A.1 shows the Turnbull empirical distribution function* and the values of the fitted Pareto distribution for the crime of person robbery. The Pareto distribution function is evaluated at the interval midpoints. The fitted parameters for burglary rates are $\hat{a} = 0.48512$ and $\hat{\sigma} = 0.05254$.

Since the column labeled "Turnbull" is the estimated value of the nonparametric distribution function on the indicated interval, comparing this with "P(midpoint)" gives some idea of the agreement between the nonparametric distribution function and the Pareto $(\hat{a}, \hat{\sigma})$ distribution.

More formally, we tested the fit of the estimated distribution to determine whether the Pareto form is appropriate. The arbitrary censoring poses problems in applying the usual goodness-of-fit tests. One possibility is to use only the exact observations to test the fit, but this method presents two problems: First, depending on the crime type, only between 8 and 27 percent of the total 440 observations were exact (not given as intervals). (Recall that those offenders who responded that they did not do the crime were included as interval responses.) Second, in accepting this approach we assume that criminals yielding exact observations are a fair representation of the criminals responding with interval observations, an assumption which is difficult to test. Another course of action, the one we have taken, is to use the estimated Turnbull empirical distribution function by comparing it to the estimated Pareto distribution function.

Our formal comparison here is a chi-square goodness-of-fit test. We estimate the observed number of observations falling into a cell using the nonparametric distribution function and the expected number

*This Turnbull function differs from the one in Table 2.1 because we have smeared the zero values in order to fit the Pareto distribution.

Table A.1

TURNBULL EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTION AND PARETO FITTED CUMULATIVE DISTRIBUTION FUNCTION FOR PERSON ROBBERIES

Turnbull	P(midpoint)	Interval	Midpoint
0.675	0.678	0.48000 to 0.50000	0.49000
0.680	0.684	0.50000 to 0.52170	0.51085
0.686	0.690	0.52170 to 0.54550	0.53360
0.693	0.696	0.54550 to 0.57140	0.55845
0.704	0.702	0.57140 to 0.60000	0.58570
0.713	0.709	0.60000 to 0.63160	0.61580
0.722	0.716	0.63160 to 0.66670	0.64915
0.732	0.727	0.66670 to 0.75000	0.70835
0.738	0.737	0.75000 to 0.80000	0.77500
0.745	0.745	0.80000 to 0.85710	0.82855
0.751	0.754	0.85710 to 0.92310	0.89010
0.763	0.762	0.92310 to 1.00000	0.96155
0.769	0.771	1.00000 to 1.09090	1.04545
0.779	0.781	1.09090 to 1.20000	1.14545
0.785	0.791	1.20000 to 1.33330	1.26665
0.789	0.798	1.33330 to 1.41180	1.37255
0.793	0.804	1.41180 to 1.50000	1.45590
0.806	0.813	1.50000 to 1.71430	1.60715
0.818	0.821	1.71430 to 1.84620	1.78025
0.823	0.826	1.84620 to 1.89470	1.87045
0.827	0.829	1.89470 to 2.00000	1.94735
0.834	0.838	2.00000 to 2.40000	2.20000
0.840	0.848	2.40000 to 2.57140	2.48570
0.846	0.857	2.66670 to 3.00000	2.83335
0.854	0.865	3.00000 to 3.42860	3.21430
0.859	0.871	3.42860 to 3.60000	3.51430
0.863	0.874	3.60000 to 3.75000	3.67500
0.867	0.877	3.75000 to 4.00000	3.87500
0.876	0.882	4.00000 to 4.50000	4.25000
0.883	0.888	4.66670 to 4.80000	4.73335
0.891	0.891	4.80000 to 5.14290	4.97145
0.895	0.893	5.14290 to 5.33330	5.23810
0.898	0.895	5.33330 to 5.45450	5.39390
0.900	0.896	5.45450 to 5.64710	5.55080
0.903	0.897	5.64710 to 5.71430	5.68070
0.906	0.899	5.71430 to 6.00000	5.85715
0.907	0.902	6.00000 to 6.40000	6.20000
0.911	0.904	6.40000 to 6.75000	6.57500
0.914	0.906	6.75000 to 6.94740	6.84870
0.915	0.909	7.20000 to 7.33330	7.26665
0.916	0.911	7.50000 to 7.63640	7.56820
0.920	0.917	8.00000 to 9.60000	8.80000
0.925	0.923	10.00000 to 10.50000	10.25000
0.929	0.926	10.50000 to 12.00000	11.25000
0.934	0.932	12.00000 to 14.40000	13.20000
0.939	0.939	14.40000 to 18.94740	16.67369
0.943	0.943	18.94740 to 20.00000	19.47369
0.944	0.946	20.00000 to 22.50000	21.25000
0.947	0.948	22.50000 to 24.00000	23.25000
0.948	0.950	24.00000 to 25.80000	24.89999
0.957	0.961	34.39999 to 50.39999	42.39999
0.961	0.966	50.39999 to 60.00000	55.20000
0.963	0.968	60.00000 to 68.57140	64.28569
0.966	0.975	92.87990 to 110.57130	101.72560
0.971	0.979	110.57130 to 183.46660	147.01892
0.975	0.981	190.52299 to 196.07990	193.30139
0.980	0.982	196.07990 to 206.39990	201.23987
0.983	0.983	206.39990 to 257.99976	232.19983
0.986	0.984	257.99976 to 265.37109	261.68530
0.994	0.993	265.37109 to 2528.39746	1396.88428
0.997	0.995	2528.39746 to 4024.79785	3276.59766
1.000		4024.79785 and greater	

using the estimated Pareto distribution function. We then calculate the chi-square statistic as usual and compare it to the chi-square distribution with p - 3 degrees of freedom, where p is the number of cells used in the test (2 degrees of freedom are subtracted for estimating the parameters a and σ). Of course, the assumptions for the usual chi-square test are not satisfied exactly, since we are estimating the number of observations that fall into a cell via the estimated nonparametric distribution function, but for this type of exploratory analysis that detail can be overlooked.

Table A.2 gives the calculation of the chi-square statistic for person robberies. The column labeled "Interval" defines the cells for calculating the statistic. The first cell has left endpoint zero and right endpoint P(z), where P is the Pareto ($\hat{a}$, $\hat{\sigma}$) distribution function and z is the value that gives the closest multiple of 5 percent exceeding the value P(MIDPOINT) on the first interval on which F(x) has value. The remaining intervals are constructed so that the expected relative frequency is 5 percent. The two columns labeled "Observed" and "Expected" under "Interval Relative Frequency" show the observed and

Table A.2

CHI-SQUARE TEST OF FIT OF PARETO DISTRIBUTION TO RATES OF PERSON ROBBERY

Interval		Interval Relative Frequency		Interval Frequency		$\frac{(O-E)^2}{E}$
From	To	Observed	Expected	Observed	Expected	
0	.5760	.70	.70	308	308	0.0
.5760	.8627	.05	.05	22	22	0.0
.8627	1.3973	.04	.05	17.6	22	.88
1.3973	2.5708	.06	.05	26.4	22	.88
2.5708	5.9986	.06	.05	26.4	22	.88
5.9986	25.2033	.04	.05	17.6	22	.88
25.2033	∞	.05	.05	22	22	0.0
Totals		1.00	1.00	440	440	3.52

NOTE: Observed χ^2 = 3.52 with 4 degrees of freedom. Do not reject Pareto at the 5 percent significance level.

expected fraction of observations for the corresponding interval cell. The columns under "Interval Frequency" show observed and expected number of observations in the corresponding interval. The last column, "$(O-E)^2/E$," is the contribution to the chi-square statistic for that cell and is calculated as

$$\frac{(\text{Observed interval frequency} - \text{Expected interval frequency})^2}{\text{Expected interval frequency}} .$$

For the crime of person robbery, the fit of Pareto is fairly good as measured by the chi-square test. However, this same procedure yielded a less satisfactory fit for the other seven crime types. Based on the chi-square tests, the Pareto model was rejected at the 5 percent level for four of the eight crimes (see Table A.3). The estimated Pareto has a slightly too heavy tail for the crime types for which it was rejected.

Table A.3

FIT OF PARETO DISTRIBUTION

Crime Type	$\hat{\alpha}$	$\hat{\sigma}$	χ^2	Result
Burglary	.390	.153	19.75	Reject at 1 percent
Business robbery	.582	.087	0.0	Do not reject
Person robbery	.485	.053	3.52	Do not reject at 5 percent
Theft (other than auto)	.375	.082	44.07	Reject at 1 percent
Auto theft	.373	.024	9.74	Reject at 5 percent
Forgery/credit cards	.389	.009	0.0	Do not reject
Fraud/swindles	.457	.016	1.76	Do not reject at 5 percent
Drug dealing	.217	.016	94.24	Reject at 1 percent

GAMMA DISTRIBUTION

The gamma density function is

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x),$$

with shape parameter α and scale parameter β. (This scale parameter is the inverse of the scale parameter given in Särndal (1964).) The parameter α was estimated by a quartiles ratio comparison described in Sec. II. Given this assumed value of α, Särndal's estimate of β from the sample quartiles is as follows:

x_1, x_2, x_3 = quartiles of the gamma distribution with shape parameter α and scale parameter 1

z_1, z_2, z_3 = sample quartiles

$$u(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x)$$

$$\begin{aligned}\text{Numerator} = {} & z_1u(x_1)(2x_1u(x_1) - x_2u(x_2)) \\ & + z_2u(x_2)(2x_2u(x_2) - x_1u(x_1) - x_3u(x_3)) \\ & + z_3U(x_3)(2x_3u(x_3) = x_2u(x_2))\end{aligned}$$

$$\begin{aligned}\text{Denominator} = {} & (x_1u(x_1))^2 \\ & + (x_2u(x_2) - x_1u(x_1))^2 \\ & + (x_3u(x_3))^2\end{aligned}$$

$$\widehat{\frac{1}{\beta}} = \frac{\text{Numerator}}{\text{Denominator}} .$$

Appendix B

DERIVATION AND ESTIMATION OF THE GMNB DISTRIBUTION

DISTRIBUTION OF $\underset{\sim}{Y}_i$

Marginally, Y_{ik} is negative binomial with parameters $(\alpha_0 + \alpha_k, P_i)$, where $P_i = \beta/(T_i + \beta)$:

$$P(Y_{ik} = y) = \binom{\alpha_0 + \alpha_k + y - 1}{\alpha_0 + \alpha_k - 1}(1 - P_i)^y (P_i)^{\alpha_0+\alpha_k};$$

$$y = 0, 1, \ldots . \qquad \text{(B.1)}$$

Now the distribution of $\underset{\sim}{\lambda}(i)$ is multivariate gamma, type I (Johnson and Kotz, 1970, Chap. 40). To get the distribution of $\underset{\sim}{Y}$, observe that Y_i has the same distribution as $X_{i0} + X_{ik}$, where

$$X_{i0} | \gamma_{i0} \overset{\text{ind}}{\sim} \text{Poisson}(\gamma_{i0}T_i), \ \gamma_{i0} \overset{\text{ind}}{\sim} \Gamma(\alpha_0, \beta), \qquad \text{(B.2)}$$

and

$$X_{ik} | \gamma_{ik} \overset{\text{ind}}{\sim} \text{Poisson}(\gamma_{ik}T_i), \ \gamma_{ik} \overset{\text{ind}}{\sim} \Gamma(\alpha_k, \beta). \qquad \text{(B.3)}$$

Thus, X_{i0} and X_{ik} have independent negative binomial distributions with parameters (α_0, P_i) and (α_k, P_i), respectively. The probability-generating function of $\underset{\sim}{Y}_i$ can be obtained from the probability-generating function of the univariate negative binomial, which is

$$P_Y(t) = P^{\alpha}(1 - Qt)^{-\alpha}, \qquad \text{(B.4)}$$

where $Q = 1 - P$ and $Y \sim NB(\alpha, P)$. Thus, the multivariate probability-generating function is

$$P_{\underset{\sim}{Y}_i}(t_1, \ldots, t_K) = E\left(\prod_{j=1}^{K} t_j^{Y_{ij}}\right)$$

$$= E\left[\left(\prod_{j=1}^{K} t_j\right)^{X_{i0}}\left(\prod_{j=1}^{K} t_j^{X_{ij}}\right)\right].$$

By independence we get

$$P_{\underset{\sim}{Y}_i}(\underset{\sim}{t}) = P_i^{\alpha_0}\left(1 - Q_i \prod_{j=1}^{K} t_j\right)^{-\alpha_0} \prod_{j=1}^{K} P_i^{\alpha_j}(1 - Q_i t_j)^{-\alpha_j}, \qquad (B.5)$$

where

$$Q_i = 1 - P_i \text{ and } \underset{\sim}{t} = (t_1, \ldots, t_k).$$

We call this the generalized multivariate negative binomial (GMNB) distribution, in contrast to the multivariate negative binomial distribution of Johnson and Kotz (1972, Chap. 11). The formula for the frequency function is fairly complicated and will not be presented here.

The moments of $\underset{\sim}{Y}_i$ follow easily from the probability-generating function or from the negative binomial representation of the distribution:

$$E(Y_{ik}) = \frac{(\alpha_0 + \alpha_k)Q_i}{P_i} = \frac{(\alpha_0 + \alpha_k)T_i}{\beta},$$

$$\text{Var}(Y_{ik}) = \frac{(\alpha_0 + \alpha_k)Q_i}{P_i^2} = \frac{(\alpha_0 + \alpha_k)T_i(T_i + \beta)}{\beta^2}, \qquad (B.6)$$

$$\text{Cov}(Y_{ik}, Y_{i\ell}) = \text{Cov}(X_{i0} + X_{ik}, X_{i0} + X_{i\ell}) \quad \text{for } k \neq \ell$$

$$= \text{Var}(X_{i0}) = \frac{\alpha_0 T_i(T_i + \beta)}{\beta^2},$$

so that

$$\rho(Y_{ik}, Y_{i\ell}) = \frac{\alpha_0}{\sqrt{(\alpha_0 + \alpha_k)(\alpha_0 + \alpha_\ell)}}, \quad \text{for } k \neq \ell.$$

ESTIMATION OF PARAMETERS

As noted above, the likelihood function of $\underset{\sim}{Y}_i$ is quite complicated, thus making maximum-likelihood estimation of the parameters an unattractive option. Our alternative is to write the expected value of Z_{ij} as a particular variance component-like model and use least squares to estimate the various expected values. There are some added complications in getting back to the original parameters $(\alpha_0, \alpha_1, \ldots, \alpha_K, \beta)$, but this method of estimation should be quite efficient.

Let $\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_n$ be a random sample from the above GMNB distribution, so that

$$Y_{ik} = X_{i0} + X_{ik},$$

where $X_{i\ell} \sim NB(\alpha_\ell, P_i)$. It will be convenient to use $Z_{ik} = Y_{ik}/T_i$ as the dependent variable. Now $E(Z_{ik}) = (\alpha_0 + \alpha_k)/\beta$, so that

$$Z_{ik} = \bar{\mu} + \mu_k + \varepsilon_{ik}, \qquad (B.7)$$

where

$$\bar{\mu} = (\alpha_0 + \bar{\alpha})/\beta, \; \bar{\alpha} = \frac{1}{K}\sum_{j=1}^{K} \alpha_j,$$

$$\mu_j = (\alpha_j - \bar{\alpha})/\beta, \; \sum_{j=1}^{K} \mu_j = 0,$$

and

$$E(\varepsilon_{ik}) = 0.$$

The covariance matrix of $\underset{\sim}{\varepsilon}_i$ has elements

$$\sigma_{k\ell}^{(i)} = \alpha_0 \tau_i \quad \text{if } k \neq \ell, \tag{B.8}$$

$$= (\alpha_0 + \alpha_k)\tau_i \quad \text{if } k = \ell,$$

where

$$\tau_i = (T_i + \beta)/T_i\beta^2.$$

Thus the correlation matrix of the errors $\underset{\sim}{\varepsilon}_i$ has off-diagonal elements

$$\alpha_0/\sqrt{(\alpha_0 + \alpha_j)(\alpha_0 + \alpha_\ell)}.$$

Note that ε_{ik} could also be written as a random-effects or components-of-variance model as

$$\varepsilon_{ik} = V_i + e_{ik},$$

where V_i and e_{ik} are independent, and

$$V_i \sim (0, \alpha_0\tau_i) \text{ and } \varepsilon_{ik} \sim (0, \alpha_k\tau_i).$$

Using Eq. (B.7), we could estimate $\bar{\mu}$, μ_1, ..., μ_K in a minimum-variance unbiased way if the covariance matrix of $\underset{\sim}{\varepsilon}_i$ were known, using generalized least squares. Two modifications are needed. First, the covariance matrix depends on the unknown parameters being estimated; and second, Eq. (B.7) gives estimates of only K parameters, while we need estimates of the K + 2 parameters α_0, α_1, ..., α_K, β. Our modifications are as follows: Initially, fit Eq. (4.7), using a one-way analysis-of-variance model with ordinary least squares. This will give initial estimates of $\bar{\mu}$, μ_1, ..., μ_K. Form the residuals r_{ik} and estimate

the functions $\alpha_0\bar{\tau}$ and $(\alpha_0 + \bar{\alpha})\bar{\tau}$ by

$$\hat{\sigma}_V^2 = \widehat{(\alpha_0 + \bar{\alpha})\bar{\tau}} = \sum_{i=1}^{n} \sum_{j=1}^{K} r_{ij}^2/K(n-1),$$

$$\hat{\sigma}_{CV}^2 = \widehat{\alpha_0\bar{\tau}} = \sum_{i=1}^{n} \sum_{j<\ell} r_{ij}r_{i\ell}\Big/ \frac{K(K-1)(n-1)}{2}. \tag{B.9}$$

Solve for values of $\hat{\alpha}_0$, $\hat{\alpha}_1$, ..., $\hat{\alpha}_K$, $\hat{\beta}$ as described below. Rerun the one-way analysis-of-variance model using generalized least squares with covariance matrix $\hat{\sigma}_{ij}^{(i)} = \hat{\alpha}_0\hat{\tau}_i$ if $j \neq \ell$ and $\hat{\sigma}_{jj}^{(i)} = (\hat{\alpha}_0 + \hat{\alpha}_j)\hat{\tau}_i$ if $j \equiv \ell$. For exploratory purposes, one iteration should be sufficient; for a final model, we may want to iterate to convergence. We may also want to consider using a weighted average in Eq. (B.9) to estimate the average variances and covariances.

Equation (B.9) and the estimates from Eq. (B.7) can be combined as follows: To keep the formulas less cluttered, we drop the "hats" that indicate the quantities are estimates. From Eqs. (B.7) and (B.9), we have

$$\bar{\mu} = (\alpha_0 + \bar{\alpha})/\beta,$$

$$\mu_j = (\alpha_k - \bar{\alpha})/\beta; \qquad k = 1, \ldots, K,$$

$$\bar{\sigma}_V^2 = (\alpha_0 + \bar{\alpha})\bar{\tau},$$

$$\bar{\sigma}_{CV}^2 = \alpha_0\bar{\tau}.$$

Since $\bar{\tau} = (1/n)\Sigma(T_i + \beta)/T_i\beta^2$, where T_1, ..., T_n are known, we have effectively K + 2 equations with K + 2 unknowns that can be solved straightforwardly. In particular,

$$\bar{\tau} = \frac{1}{n}\sum_{i=1}^{n} \frac{T_i + \beta}{T_i\beta^2} = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{1}{\beta^2} + \sum_{i=1}^{n} \frac{1}{T_i\beta}\right] = \frac{1}{\beta^2} + \frac{1}{n\beta}\sum_{i=1}^{n} \frac{1}{T_i}.$$

Substituting the above expression for $\bar{\tau}$, the equations are

$$\bar{\mu} = (\alpha_0 + \bar{\alpha})/\beta, \tag{B.10}$$

$$\mu_j = (\alpha_j - \bar{\alpha})/\beta, \tag{B.11}$$

$$\bar{\sigma}_V^2 = (\alpha_0 + \bar{\alpha}) \left(\frac{1}{\beta^2} + \frac{1}{n\beta} \sum_{i=1}^{n} \frac{1}{T_i} \right), \tag{B.12}$$

$$\bar{\sigma}_{CV}^2 = \alpha_0 \left(\frac{1}{\beta^2} + \frac{1}{n\beta} \sum_{i=1}^{n} \frac{1}{T_i} \right). \tag{B.13}$$

To get estimates, substitute Eq. (B.10) into Eq. (B.12) to get

$$\bar{\sigma}_V^2 = \bar{\mu} \left(\frac{1}{\beta} + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \right) \longrightarrow \hat{\beta} = \frac{\bar{\mu}}{\bar{\sigma}_V - \bar{\mu} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i}} .$$

Now substitute $\hat{\beta}$ into Eq. (B.13) to get

$$\hat{\alpha}_0 = \frac{\bar{\sigma}_{CV}^2}{\frac{1}{\hat{\beta}^2} + \frac{1}{n\hat{\beta}} \sum_{i=1}^{n} \frac{1}{T_i}} .$$

Then substitute $\hat{\alpha}_0$ and $\hat{\beta}$ into Eq. (B.10) to get

$$\hat{\bar{\alpha}} = \bar{\mu}\, \hat{\beta} - \hat{\alpha}_0 .$$

Finally, substitute $\hat{\beta}$ and $\hat{\bar{\alpha}}$ into Eq. (B.11) to get

$$\hat{\alpha}_j = \hat{\mu}_j\, \hat{\beta} + \hat{\bar{\alpha}} .$$

BIBLIOGRAPHY

Anscombe, F. J., "The Transformation of Poisson, Binomial and Negative Binomial Data," *Biometrika*, Vol. 35, 1948, pp. 246-254.

Bronx County (New York) District Attorney's Office, "The Major Offense Bureau: An Exemplary Project," A report prepared for the Law Enforcement Assistance Administration, U.S. Department of Justice, 1976.

Carter, G., and J. Rolph, "Analyzing the Demand for Fire Department Services," Chap. 8 in Rand Fire Project, *Fire Department Deployment Analysis*, Elsevier-North Holland, 1979.

Chaiken, J. M., "Models Used for Estimating Crime Rates," in Mark A. Peterson et al., *Doing Crime: A Survey of California Prison Inmates*, The Rand Corporation, R-2200-DOJ, April 1980, pp. 224-252.

Chaiken, J. M., and J. E. Rolph, "Selective Incapacitation Policies Based on Estimated Crime Rates," *Operations Research*, Vol. 28, No. 6, 1980, pp. 1259-1274.

Clevenson, M. L., and J. V. Zidek, "Simultaneous Estimation of the Mean of Independent Poisson Laws," *Journal of the American Statistical Association*, Vol. 70, 1975, pp. 698-705.

Cox, D. R., and P.A.W. Lewis, *The Statistical Analysis of Series of Events*, Methuen and Company, London, 1966.

Dahmann, J., and J. L. Lacy, *Criminal Prosecution in Four Jurisdictions: Departures from Routine Processing in the Career Criminal Program*, METREK-MITRE, Mitre Technical Report MTR-7550, 1977.

DeGroot, M. H., *Optimal Statistical Decisions*, McGraw-Hill Book Co., Inc., New York, 1970.

Efron, B., and C. Morris, "Combining Possibly Related Estimation Problems (with discussion)," *Journal of the Royal Statistical Society*, Series B, Vol. 35, No. 3, 1973a, pp. 379-421.

Efron, B., and C. Morris, "Data Analysis Using Stein's Estimator and Its Generalizations," *Journal of the American Statistical Association*, Vol. 70, 1975, pp. 311-319.

Efron, B., and C. Morris, "Stein's Estimation Rule and Its Competitors--an Empirical Bayes Approach," *Journal of the American Statistical Association*, Vol. 68, 1973b, pp. 101-108.

Fay, R. E., and R. A. Herriot, "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, Vol. 74, 1979, pp. 269-277.

Fisher, R. A., "On the Interpretation of χ^2 from Contingency Tables and Calculation of P," *Journal of the Royal Statistical Society*, Series A, Vol. 85, 1922, pp. 87-94.

Gottfredson, D. M., et al., "Classification for Parole Decision Policy," National Institute of Law Enforcement and Criminal Justice, Washington, D.C., 1978.

Greene, M. A., "The Incapacitative Effect of Imprisonment Policy on Crime," Ph.D. Thesis, Carnegie-Mellon University, 1977.

Greenwood, P. W., *Rand Research on Criminal Careers: An Update on Progress to Date*, The Rand Corporation, N-1572-DOJ, October 1980.

Hartigan, J. A., "Linear Bayesian Methods," *Journal of the Royal Statistical Society*, Series B, Vol. 31, No. 3, 1969, pp. 446-454.

Hartley, H. O., "Maximum Likelihood Estimation from Incomplete Data," *Biometrics*, Vol. 14, 1958, pp. 174-194.

Hoffman, P., and L. K. DeGostin, *Parole Decision-Making: Structuring Discretion*, U.S. Board of Parole Research Unit, Report 5, 1974. Reprinted in L. Radzinowicz and M. E. Wolfgang, *Crime and Justice*, Vol. II, Basic Books, New York, 2d ed., 1977.

Hudson, H. M., "Adaptive Estimators for Simultaneous Estimation of Poisson Means," School of Economic and Financial Studies, Macquarie University, New South Wales, Australia, 1980.

Hudson, H. M., "A Natural Identity for Exponential Families with Applications in Multi-Parameter Estimation," *Annals of Statistics*, Vol. 6, 1978, p. 478.

Hudson, H. M., and K. Tsui, "Simultaneous Poisson Estimators for a Priori Hypotheses About Means," *Journal of the American Statistical Association*, Vol. 76, 1981, pp. 182-187.

Institute for Law and Social Research, "Curbing the Repeat Offender: A Strategy for Prosecutors," PROMIS Research Project, Publication 3, 1977.

Irwin, J. O., "The Generalized Waring Distribution," *Journal of the Royal Statistical Society*, Series A, Vol. 138, 1975, Part I, pp. 18-31; Part II, pp. 205-225; Part III, pp. 374-384.

Irwin, J. O., "The Generalized Waring Distribution Applied to Accident Theory," *Journal of the Royal Statistical Society*, Series A, Vol. 131, 1968, pp. 205-225.

James, W., and C. Stein, "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium, Math. Stat. Prob.*, Vol. 1, 1961, pp. 361-379.

Johnson, N. L., and S. Kotz, *Discrete Distributions*, John Wiley & Sons, New York, 1972.

Johnson, N. L., and S. Kotz, *Multivariate Distributions*, John Wiley & Sons, New York, 1970.

Maltz, M. D., *On Recidivism: Exploring Its Properties as a Measure of Correctional Effectiveness*, Center for Research in Criminal Justice, University of Illinois at Chicago Circle, Chicago, 1980.

Maltz, M. D., and S. M. Pollock, "Artificial Inflation of a Delinquency Rate by a Selection Artifact," *Operations Research*, Vol. 28, No. 3, Part I, May-June 1980, pp. 547-559.

Morris, N., *The Future of Imprisonment*, University of Chicago Press, Chicago, 1974.

Muñoz, A., *Nonparametric Estimation from Censored Bivariate Observations*, Biostatistics Technical Report No. 60, Stanford University, Stanford, August 1980.

Pate, T., R. A. Bowers, and R. Parks, *Three Approaches to Criminal Apprehension in Kansas City: An Evaluation Report*, Police Foundation, Washington, D.C., 1976.

Peng, J., *Simultaneous Estimation of the Parameters of Independent Poisson Distributions*, Technical Report No. 78, Department of Statistics, Stanford University, Stanford, 1975.

Peterson, M., and H. Braiker, with S. Polich, *Doing Crime: A Survey of California Prison Inmates*, The Rand Corporation, R-2200-DOJ, April 1980.

Peterson, M., J. Chaiken, and P. Ebener, *Survey of Prison and Jail Inmates: Background and Methods*, The Rand Corporation, N-1635-NIJ (forthcoming).

Sampford, M. R., "The Truncated Negative Binomial Distribution," *Biometrika*, Vol. 42, 1955, pp. 58-69.

Sanathanan, L., "Estimating the Size of a Truncated Sample," *Journal of the American Statistical Association*, Vol. 72, 1977, pp. 669-672.

Särndal, C.-E., "Estimation of the Parameters of the Gamma Distribution by Sample Quantiles," *Technometrics*, Vol. 6, No. 4, 1964, pp. 405-414.

Sibuya, M., "Generalized Hypergeometric, Digamma, and Trigamma Distributions," *Annals of Mathematical Statistics*, Vol. 31, Part A, 1979, pp. 373-390.

Sibuya, M., "Multivariate Digamma Distribution," *Annals of Mathematical Statistics*, Vol. 32, Part A, 1980, pp. 25-36.

Stephens, M. A., "Use of the Kolmogorov-Smirnov, Cramer-Von Mises and Related Statistics Without Extensive Tables," *Journal of the American Statistical Association*, Vol. 69, 1974, p. 730.

Stollmack, S., and C. M. Harris, "Failure-Rate Analysis Applied to Recidivism Data," *Operations Research*, Vol. 22, No. 6, 1974, pp. 1192-1205.

Tsui, K., *Simultaneous Estimation of Several Poisson Parameters Under Squared Error Loss*, Technical Report No. 37, Department of Statistics, University of California, Riverside, 1978.

Tsui, K., and S. J. Press, *Simultaneous Bayesian Estimation of the Parameters of Independent Poisson Distributions*, Technical Report No. 33, Department of Statistics, University of California, Riverside, 1977.

Turnbull, B. W., "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Series B, Vol. 38, No. 3, 1976, pp. 290-295.

von Hirsch, A., *Doing Justice: The Choice of Punishment*, Hill and Wang, New York, 1976.

Wilson, E. B., and M. M. Hilferty, "The Distribution of Chi-square," *Proceedings of the National Academy of Sciences*, Washington, D.C., Vol. 17, 1931, pp. 684-688.

Wolfgang, M., R. Figlio, and T. Sellin, *Delinquency in a Birth Cohort*, University of Chicago Press, Chicago, 1972.

END