9500

59244

THE DIRICHLET-GAMMA-POISSON MODEL OF REPEATED EVENTS: A MULTIVARIATE DESCRIPTION OF CRIMINAL VICTIMIZATION IN AMERICAN CITIES*

by

James P. Nelson, Ph.D.

The Michael J. Hindelang
Criminal Justice Research Center, Inc.
One Alton Road
Albany, New York 12203

September, 1982

*Support for this project was provided, in part, by grant number 81-IJ-CX-0045, awarded to the Criminal Justice Research Center, Inc. under the Methodology Development Program by the National Institute of Justice, United States Department of Justice. Points of view or opinions expressed herein are those of the author, and do not necessarily represent the official position of the Department of Justice.

I would like to thank Melvin Katz for help in conceptualizing the models and in developing maximum likelihood solutions, Steven Greenstein for computer programming, and Betsy Colvin, for editorial assistance.

Abstract

A new multivariate statistical model of repeated events, the Dirichlet-gamma-Poisson model, is shown to accurately account for the multivariate distribution of four types of victimizations reported in city samples of the National Crime Survey. The life-style theory of victimization is used to interpret the compounding that defines the model. Parameter estimation, interpretation, and the prediction of future events based on past events are discussed. The model appears to be applicable to a variety of repeated events data.

THE DIRICHLET-GAMMA-POISSON MODEL OF REPEATED EVENTS:
A MULTIVARIATE DESCRIPTION OF CRIMINAL VICTIMIZATION IN AMERICAN CITIES

This paper develops a new model of repeated events, the Dirichlet-gamma-Poisson model, as a means of understanding how the multivariate distribution of crimes reported in city samples of the National Crime Survey (NCS) can be used to make inferences about exposure to high crime situations. The model is based upon the assumption that persons have a constant chance of being victimized over time, but that not all persons have the same chance.

Differences in the chances of being victimized are hypothesized by a number of researchers (Cohen and Felson, 1979; Hindelang, Gottfredson and Garofalo, 1978; the National Research Council, 1976; Skogan, 1980; Sparks, Genn and Dodd, 1977; and Sparks, 1980), to be largely due to differences in exposure to high crime situations, which in turn, are hypothesized to be largely due to differences in life-styles. For example, males are thought to be more exposed to crime than females because they spend more time away from home and are more likely to be in the company of potential offenders. Unfortunately, this theory is difficult to evaluate because exposure is hard to measure. Other than needing to know how often persons are in the presence of potential offenders, most researchers agree that one must also know how often potential victims represent vincible and desirable targets to potential offenders. The present research shifts the emphasis from asking what constitutes exposure, to asking how the multivariate distribution of various types of crimes reported in one time period can be used to make inferences about victim liability, which presumably, corresponds closely to victim exposure.

The discussion will begin by reviewing the simple Poisson model and showing how it can be generalized into the univariate gamma-Poisson model. This model has been shown by Nelson (1980a) to be compatible with the life-style/exposure theory of victimization and to be capable of generating the univariate distribution of many different types of victimizations. Three multivariate gamma-Poisson models will then be developed and fitted to distributions of four specific types of victimizations reported in the NCS city samples. The Dirichlet-gamma-Poisson model is the most general of these models. The discussion will show how the model can be used to estimate individual liability rates of specific types of crimes and to predict chances that specific types of crimes will occur in the future. The model is expected to be useful in describing many different kinds of social phenomena.

UNIVARIATE MODELS

The Poisson Model

The Poisson model is based upon the assumptions that (1) the probability of being victimized is the same for all persons, and (2) that it does not vary over time. This model has frequently been used to evaluate whether there are more persons reporting two or more victimizations than would be expected if all persons had the same chance of being victimized. Some persons are expected to be multiply victimized under Poisson models and such misfortune is assumed to represent bad luck rather than victim liability. To the extent that the data show more multiple victims than expected, one tends to reject the hypothesis of equal victim liability in favor of stating that some persons are more liable of being victimized than others. Research (Hindelang, et al., 1978; Nelson, 1980a; Sparks, et al., 1977) has shown that there are more persons reporting multiple victimizations than are expected under Poisson models.

Under the Poisson model, the probability of experiencing x victimizations during some period of time may be expressed as:

$$P(x) = e^{-\lambda}\lambda^{x}/ x!, \qquad (1)$$

where λ is the Poisson parameter for this time period. The maximum likelihood estimate of λ is the mean or average rate.[1]

---

[1] In general, the parameter can be expressed as λt, where t measures the number of time units that λ is based upon. Here, t equals one to simplify various equations.

The inability of the Poisson model to account for multiple victimizations is illustrated in Table 1, which displays the observed and the expected number of personal contact victimizations (excluding rape) recorded in the National Crime Survey (NCS) made in Baltimore, 1975. The NCS city data are based upon interviews of all persons living in approximately 10,000 randomly selected households in each city. Persons aged 12 and over were asked to report their victim experiences for the year preceding the interview regardless of whether they reported the crimes to the police. Table 1 shows that the Poisson model predicted far fewer multiple victimizations than were reported in the survey. This suggests that either or both assumptions of the Poisson model are inconsistent with the data. In sharp contrast, the table shows that the gamma-Poisson model was very consistent with the observed data.

The unweighted number of personal victimizations reported in city samples of the NCS will be used to develop models in this paper. The data analysis will be limited to interviews made in the five largest cities of the United States and in the eight cities that participated in the Law Enforcement Assistance Administration's High Impact Crime Reduction Program. These interviews occurred during the first quarter of 1975 so that the victimizations correspond to crimes that occurred during most of 1974 and part of 1975. The NCS program is described by Garofalo and Hindelang (1978). Comparisons of NCS and Uniform Crime Report data can be found in Nelson (1980b).

Table 1: Observed and Expected Number of Personal Contact Victimizations Under a Poisson and a Gamma-Poisson Model in Baltimore.
(National Crime Survey Data, 1974-1975)

Number of Victimizations	Observed Frequency	Expected Frequency Under Two Models: Poisson Model	Gamma-Poisson Model
0	21,511	21,213.9	21,512.8
1	1,494	1,995.2	1,478.3
2	231	93.8	249.9
3	52	2.9	50.6
4	11	.1	11.1
5	6	.0	2.5
6	1	.0	.6

The Gamma-Poisson Model

Greenwood and Woods (1919) and Greenwood and Yule (1920) expanded the Poisson model by compounding it with a gamma distribution. Applied to victimization studies, the model suggests that persons have a constant chance of being victimized over time, but that not all persons have the same chance. Individual victimization rates are treated as random variables from a gamma distribution. The probability density function for a gamma distribution may be expressed as:

$$f(\lambda) = (k/m)^k \lambda^{k-1} e^{-(k/m)\lambda} / \Gamma(k) \qquad (2)$$

where m is the mean victimization for the population, k is the exponent and in conjunction with m defines the shape of the gamma distribution, Λ (which is not directly represented in this equation) is the random variable representing individual victimization rates with density function $f(\lambda)$, and $\Gamma(k)$ is the gamma function of k. Graphs of various gamma density functions are presented in Nelson (1980a).

Under the gamma-Poisson model, the probability of experiencing x victimizations is a Poisson random variable conditional upon the value of λ. If everyone in the population had exactly the same rate, then the model would be the Poisson model. The unconditional probability of reporting x victimizations is found by multiplying equation (1) by the probability density function for λ, equation (2), and then integrating λ from zero to infinity. This results in a compound Poisson model which may be expressed as:

$$\begin{aligned} P(x) &= \int_0^{\infty} P(x|\lambda)\, f(\lambda)\, d\lambda, \\ &= \left[\frac{k}{k+m}\right]^k \frac{\Gamma(x+k)}{\Gamma(k)x!} \left[\frac{m}{k+m}\right]^x . \end{aligned} \qquad (3)$$

This model will be called the univariate gamma-Poisson model to emphasize that it is a Poisson model compounded with a gamma distribution of victim liability. It is identical in form to a negative binomial distribution and can be generated in a number of other ways. A maximum likelihood procedure is presented in Appendix A to estimate the parameters.

A summary of the fitting of univariate gamma-Poisson models to the number of persons reporting robberies, aggravated assaults, simple assaults, and larcenies with contact (purse snatching and pocket picking) and to the total of these four crimes for 13 cities in the NCS is presented in Table 2. The p values are large suggesting close correspondence between the model and the observed data. Half of the samples that were testable had p values in excess of .47. Not one city had a p value below .01.

Table 2: Parameter Estimates and P Values for Univariate Gamma-Poisson Models Fitted to Various Types of Personal Crime in 13 Cities (National Crime Surveys, 1974-75)

City[a]	Parameter estimate and P value[b]	Type of personal crime: Robbery	Aggravated assault	Simple assault	Larceny with contact	All four combined
Newark	m	.0229	.0076	.0057	.0105	.0467
	k	.3144	.1323	.5124	1.3580	.3606
	p value	.91	N.T.	N.T.	N.T.	.75
Atlanta	m	.0174	.0124	.0114	.0093	.0504
	k	.1636	.0664	.0965	.2435	.2249
	p value	.64*	.44*	N.T.	N.T.	.47
Dallas	m	.0123	.0175	.0169	.0063	.0531
	k	.1044	.1093	.0800	.0897	.1709
	p value	.89	.72*	.88	N.T.	.08
St. Louis	m	.0189	.0143	.0139	.0091	.0562
	k	.2369	.0886	.0689	.1917	.2147
	p value	N.T.	.26*	.40*	N.T.	.23
New York	m	.0236	.0085	.0096	.0148	.0565
	k	.1400	.0597	.2294	.1801	.2717
	p value	.94	N.T.	N.T.	N.T.	.30
Philadelphia	m	.0205	.0133	.0133	.0124	.0595
	k	.0778	.0791	.0759	.1188	.1923
	p value	.65	.85*	.74*	N.T.	.45
Los Angeles	m	.0177	.0165	.0222	.0079	.0643
	k	.1063	.1032	.0925	.1200	.2136
	p value	.60*	.05+	.62	N.T.	.38
Portland	m	.0157	.0217	.0296	.0052	.0725
	k	.0553	.1395	.1271	.1784	.1858
	p value	.01	.20	.05	N.T.	.88
Denver	m	.0188	.0224	.0271	.0058	.0741
	k	.0715	.0975	.1425	.2001	.1837
	p value	.89	.56	.64	N.T.	.02
Cleveland	m	.0270	.0202	.0175	.0095	.0742
	k	.1219	.0993	.1174	.1098	.2376
	p value	.23	.33	.76*	N.T.	.02
Chicago	m	.0286	.0156	.0138	.0167	.0748
	k	.1614	.0904	.0713	.4095	.2573
	p value	.88	.58*	.20	N.T.	.44
Detroit	m	.0368	.0210	.0176	.0082	.0835
	k	.1552	.0931	.0763	.2871	.2389
	p value	.39	.15	.82	N.T.	.01
Baltimore	m	.0346	.0205	.0205	.0185	.0941
	k	.1450	.0925	.0754	.1911	.2551
	p value	.49	.23	.51	.65*	.13

*These p values were calculated allowing a minimum of at least one observation for each expected value in the chi-square test. A chi-square test could not have been made for these models if the expected value in each cell had to equal three or more.

[a]Cities are listed in ascending order by their overall victimization rate.

[b]The p values were based upon comparing observed and expected frequencies with the Pearson chi-square test.

[c]N.T. signifies the model was not testable because all the degrees of freedom were used to estimate the parameters.

P values could not be estimated in 18 samples because there were not enough multiple victimizations to both estimate the parameters and to test them on the same data. This situation occurred in 12 out of the 13 analyses of larceny with contact. Only five persons in all 13 cities reported more than two larcenies with contact. This means that almost every city analysis was made on the frequency of persons reporting zero, one and two larcenies with contact. While it was possible to estimate the two parameters of the gamma-Poisson model, it was not possible to test the fit because there were no degrees of freedom left after the parameters were estimated.[2]

Table 2 demonstrates that the gamma-Poisson model is capable of generating the univariate distribution of specific as well as aggregated types of personal victimizations reported in the NCS data. Under the model, each person can be thought of as having an unique liability rate for each specific type of crime that is stable over time. This liability rate is hypothesized to be largely a function of exposure to high crime situations, wherein exposure refers to the frequency that offenders come into contact with victims who are judged to be desirable and vincible targets of their actions.

The question raised is: Can the same liability rate account for the distribution of all four types of crime analyzed thus far, or is a multivariate conceptualization needed to study victim liability? If a multivariate model were needed, would the dimensions be related to or independent of each other? These questions can be answered by comparing various multivariate models based upon different assumptions about how liability is related to reported victimizations. The Dirichlet-gamma-Poisson was developed by comparing various multivariate models.

[2]Some of the p values listed in Table 2 would be reclassified as not testable if different criteria for aggregating expected values were used. If the nine starred p values were based on chi-square tests wherein expected values were aggragated to produce an expected value of at least three, then these nine tests would be classified as not testable. All chi-square tests are based upon aggregating expected counts to at least three in other tables.

MULTIVARIATE MODELS

The Independence Gamma-Poisson Model

One of the first models to be tested in almost any multivariate analysis is the independence model. Under this model, crime rates are represented by four dimensions, each of which provides no information about the other. The model can be efficiently estimated by first fitting univariate gamma-Poisson model to each of the four types of crimes, and then by multiplying the probability of each separate crime to get the joint probability of all four types. Table 2 shows that univariate gamma-Poisson models are consistent with the univariate distribution of all four crimes across all 13 cities. The m and k parameter estimates for each crime and city are also listed in this table.

The Fixed Gamma-Poisson Model

The independence model is not expected to accurately describe the data because the specific types of crime are usually thought to be related to each other. A simple multivariate generalization of the gamma-Poisson model that allows the crime types to be related to each other can be developed by assuming that the joint probability of all four crimes is a product of independent Poisson probabilities conditional upon λ, that each type of crime has a mean equal to $p_i\lambda$, and that λ is a random variable from a gamma distribution. In this model, λ represents each person's liability of reporting a victimization, and p_i represents the probability that a victimization is of type i. Note that p_i represents the conditional probability that a victimization is of type i given that a victimization has occurred. The model is called fixed

because all persons are hypothesized to have exactly the same set of conditional probabilities. For example, if 40 percent of all victimizations were robberies, than p_i would equal .40 for robbery for all victims. The model may be written as:

$$P(x_1,x_2,x_3,x_4) = {}_0\int^{\infty} \prod_{i=1}^{4} (P(x_i|\lambda)\ f(\lambda)\ d\lambda,$$

$$= \{[\frac{k}{k+m}]^k \frac{\Gamma(x_T+k)}{\Gamma(k)x_T!} [\frac{m}{k+m}]^{x_T}\}$$

$$x_T! \prod_{i=1}^{4} p_i^{x_i} / x_i! \ , \qquad (4)$$

where $x_T = x_1 + x_2 + x_3 + x_4$. The integration in equation (4) shows that crime types appear to be related to each other because they are related to one liability dimension. For example, reporting a robbery would be associated with reporting an assault if both events were indicators of high exposure to crime. This model is analogous to a one dimensional factor analysis model.

The fixed gamma-Poisson model is simple to estimate because it can be broken down into a univariate gamma-Poisson model for the total number of reported victimizations (the part within braces in equation 4) multiplied by an independent multinomial model that distributes the total number of victimizations into combinations of crime types. This form of the model was introduced by Patil (1964). Maximum likelihood estimates can be found by estimating m and k in an univariate gamma-Poisson model fitted to the total number of victimizations, and by estimating p_i from the observed proportion of victimization of each type. The model has been developed in some detail by Bates and Neyman (1952) and by Arbous and Kerrich (1951).

The Dirichlet-Gamma-Poisson Model

The assumption that all victims have the same conditional probability of each type of victimization in the fixed gamma-Poisson model appears restrictive. From a life-style/exposure perspective, it seems more likely that certain life-styles will be associated with certain types of crime. For example, the NCS data show that younger males have a greater tendency to be assaulted than to have their wallets picked, whereas older males have a greater tendency to have their wallets picked than to be assaulted.

One way to introduce victim "specialization" is to treat the conditional probability of each type of crime as a random variable. If the conditional probability that a crime was of a particular type were a random variable from a Dirichlet distribution, then some persons would be more likely to experience various types of victimizations than others, presumably due to differences in exposure to each type of crime.

Let P_i represent the random variable measuring the conditional probability that a crime is of type i, and let p_i represent its particular value for some person. The Dirichlet distribution for four types of crime may be written as:

$$P(p_1, p_2, p_3, p_4) = \frac{\Gamma(\theta_T) \prod_{i=1}^{4} p_i^{\theta_i - 1}}{\prod_{i=1}^{4} \Gamma(\theta_i)}, \qquad (5)$$

where $\theta_T = \theta_1 + \theta_2 + \theta_3 + \theta_4$, $\theta_T > 0$, and $p_1 + p_2 + p_3 + p_4 = 1$. The parameters to be estimated are θ_1 to θ_4, one for each crime. The Dirichlet distribution is discussed by Johnson and Kotz (1972).

The θ_i's are related to the p_i's in the following manner:

$$E(p_i) = \theta_i / \theta_T, \qquad (6)$$

and

$$V(p_i) = \theta_i(\theta_T - \theta_i)/[\theta_T^2 (\theta_T + 1)]. \qquad (7)$$

The Dirichlet-gamma-Poisson model is formed by assuming that the fixed gamma-Poisson model is defined conditionally for a set of p_i values, by multiplying it by the probability density function of the Dirichlet distribution, and then by integrating the product over all possible p_i values. For four types of crimes, the model may be written as:

$$P(x_1, x_2, x_3, x_4) = \left\{ \left[\frac{k}{k+m}\right]^k \frac{\Gamma(x_T + k)}{\Gamma(k) x_T!} \left[\frac{m}{m+k}\right]^{x_T} \right\} \cdot \frac{x_T! \, \Gamma(\theta_T) \prod_{i=1}^{4} \Gamma(x_i + \theta_i)}{\Gamma(\theta_T + x_T) \prod_{i=1}^{4} \Gamma(\theta_i) x_i!}. \qquad (8)$$

This equation shows that the model can be thought of as a univariate gamma-Poisson model (the part in braces) that generates the distribution of the total number of victimizations (x_T), times a Dirichlet part that allocates the total to the multivariate distribution of the various combinations corresponding to this total.

Although the Dirichlet part of equation (8) may look formidable, differences between it and the probability density function for the fixed gamma-Poisson model presented in equation (4) can be readily understood by noting that both models can be divided into 1) a part that generates the probability of observing the total number of victimizations under consideration, 2) a part that counts the number of ways or permutations in which the particular outcome could have occurred, and 3) the conditional probability of one of those ways given that the total number of victimizations corresponding to this event occurred. The first part is generated by the same univariate gamma-Poisson model under both models. Therefore, both models predict the same number of persons to not be victimized, as well as the same number of persons to experience a total of one, two, three, etc. victimizations.

The count of the number of permutations in which an event can occur is also identical in each mode. It is represented by the $x_T!/(x_1!x_2!x_3!x_4!)$ term. Thus, differences between the models lie only in the estimation of the conditional probability of the permutations making up the event under consideration. These differences can best be understood by considering a permutation as if the order of the victimizations were known. Of course, all permutations have the same conditional probability so that it is not necessary to consider all of them.

First, consider the conditional probability of reporting exactly one victimization of type i given that at least one victimization was reported. It equals p_i under the fixed model and θ_i/θ_T under the Dirichlet model. The expression for the Dirichlet model was derived from equation (8) by noting that $\Gamma(x+1) = x\Gamma(x)$. These conditional probabilities are expected

to be very close to each other in most data analyses. Under the fixed model, p_i is estimated by the proportion of victimizations of type i. Under the Dirichlet model, θ_i/θ_T is equal to the expected value of the random variable P_i, which measures each person's conditional probability of a type i crime given that he or she was victimized.

Second, consider the conditional probability of reporting two crimes of type i for persons who experienced at least two crimes. This can be calculated by multiplying the conditional probability that the crime was of type i for persons who reported at least two crimes, times the conditional probability that the second crime was of type i for persons who reported at least two crimes and who reported a type i first crime. This can be expressed as p_i^2 under the fixed model and as (θ_i/θ_T) times $[(\theta_i+1)/(\theta_T+1)]$ under the Dirichlet model. Note that the conditional probability that the second crime is of type i (listed within brackets for the Dirichlet model) is larger than the conditional probability that the first crime is of type i for the Dirichlet but not for the fixed model. Likewise, the conditional probability of reporting three type i crimes equals p_i^3 under the fixed model and $(\theta_i/\theta_T)\ [\theta_i+1)/(\theta_T+1)]\ [\theta_i+2)/(\theta_T+2)]$ under the Dirichlet model. In general, the conditional probability that the next crime is the same as the last crime increases under the Dirichlet model but not under the fixed model.

Conversely, the conditional probability of reporting different types of crime decreases in the Dirichlet but not in the fixed model. For example, the conditional probability of reporting a type i followed by a type j crime equals $p_i p_j$ in the fixed model and

and (θ_i/θ_T) $[\theta_j/(\theta_T+1)]$ in the Dirichlet model. This ability to modify the conditional probability of the next crime type is what allows victim specialization to be incorporated into the Dirichlet model. It does not suggest that a victim's chances of experiencing a particular type of crime change, though. Rather, it shows how the model's estimation of a person's chances of reporting a particular type of crime can change depending upon the person's victimization history. This will be illustrated again in a later section.

The extent of the differences between the fixed and the Dirichlet models depends upon the size of the θ_i parameters. If θ_T were to approach infinity such that the expected value of the random variable p_i equalled θ_i/θ_T for all i, then the conditional probability of crime i would remain constant over repeated victimizations. In other words, the Dirichlet model would degenerate into the fixed model. If θ_T were to approach zero, then the conditional probability that the second crime were the same as the first would approach one. Here the Dirichlet model would represent a model of mutually exclusive types of victimizations in which a victim could experience at most one type of crime.

A number of models are special cases of the Dirichlet-gamma-Poisson model. If θ_T becomes very large, then the model degenerates into the fixed gamma-Poisson model. If θ_T becomes very small, then the model becomes a mutually exclusive gamma-Poisson model. If the parameter k becomes very large, then the model degenerates into a Dirichlet-Poisson model. In this model, all persons have the same chance of being victimized, but the conditional probability of any specific type of victimization given that a victimization occured differs by person. If θ_T as well as k become very large, then the model

degenerates into a multivariate independence Poisson model. Also note that the univariate gamma-Poisson model degenerates into a Poisson model when k or when the ratio of k to m becomes very large.

Maximum likelihood estimates of the m and k parameters can be easily obtained by fitting a univariate gamma-Poisson model to the total number of victimizations. These estimates of m and k are independent of the Dirichlet parameters. Maximum likelihood estimates of the Dirichlet parameters are presented in Appendix B.

Comparisons of the Independence, the Fixed and the Dirichlet-Gamma-Poisson Models

Pearson chi-square goodness-of-fit test statistics for the independence, the fixed, and the Dirichlet-gamma-Poisson models are presented in Table 3 for the 13 NCS cities. The procedures used to estimate the chi-square values are discussed in the next paragraph. The large chi-square values for the independence model suggest that it is unreasonable to assume that the four types of crime are unrelated to each other. The fixed model fit the data better than did the independence model, but not as well as the Dirichlet model. The fixed model accurately described the multivariate distribution of crime in only one city, Newark. The Dirichlet-gamma-Poisson model accurately reproduced the multivariate distribution of the crime types in at least nine of the other twelve cities.

Degrees of freedom were derived by subtracting the number of independent parameters estimated in each model from the number of cells used in the chi-square calculation minus one. One degree of freedom was lost because the models were conditioned upon the total number of persons interviewed. Note that two models could differ by one parameter but their chi-square tests would not necessarily differ by one degree of freedom because the expected values determined the number of cells to be used in the chi-square test. For example, a cell could have an expected value greater than three under one model and therefore be counted in the total number of cells for the test, but it could have an expected value less than three under another model and therefore not be counted as a separate cell.

Table 3 Chi-Square Goodness of Fit Statistics of the Multivariate Distribution of Four Types of Crimes Under Three Gamma-Poisson Models in 13 Cities
(National Crime Surveys, 1974-75)

	MODEL					
	Independence Gamma-Poisson		Fixed Gamma-Poisson		Dirichlet-Gamma-Poisson	
City[a]	Pearson Chi-Square	Degrees of Freedom	Pearson Chi-Square	Degrees of Freedom	Pearson Chi-Square	Degrees of Freedom
Newark	83.6	2	9.7**	7	9.7**	7
Atlanta	119.6	3	36.4	10	23.2*	10
Dallas	199.7	5	50.2	10	16.2*	10
St. Louis	161.3	4	55.4	10	39.5	10
New York	77.4	4	36.0	8	17.1*	10
Philadelphia	153.4	6	110.7	10	23.5*	11
Los Angeles	155.1	5	83.6	11	22.5*	10
Portland	268.2	5	76.4	14	19.0*	12
Denver	314.0	6	46.0	14	22.1*	13
Cleveland	137.9	8	77.0	12	11.9**	11
Chicago	323.0	9	105.2	12	47.8*	11
Detroit	159.7	7	98.8	14	16.9**	14
Baltimore	332.4	11	181.9	17	63.5	16

[a] Cities are listed in ascending order by their overall victimization rate.

* $p > .01$

** $p > .10$

The techniques used to calculate the chi-square test statistics as well as differences between the three models are illustrated in Table 4 for Baltimore, the city with the worst fit of all Dirichlet models. Table 4 displays combinations of zero, one, two and three victimizations wherein the expected value for each combination under the Dirichlet model exceeded three, and two aggregated cells containing combinations whose expected values before aggregation were less than three. One aggregated cell contains combinations of multiple victimizations of only one type, such as four robberies, and the other contains combinations of multiple victimizations of more than one type, such as four robberies and one aggravated assault.

Table 4 shows that the Dirichlet-gamma-Poisson model did a good job of fitting the observed frequencies for nearly all combinations of crimes not involving larceny with contact.[3] It underestimated the number of persons reporting exactly one larceny, but it overestimated the number reporting exactly two larcenies as well as the number of persons reporting one larceny and one other crime.[4]

---

[3] The number of persons reporting 0 robberies, 1 aggravated assault, 1 simple assault, and 0 larcenies with contact were underestimated by the model. This occurred in several of the cities.

[4] This pattern suggests that the data for Baltimore might be better modeled by fitting a Dirichlet-gamma-Poisson model to the trivariate distribution of robbery, aggravated assault and simple assault, by fitting a gamma-Poisson model to larceny, and then by fitting all four crimes by assuming independence between these two models. This model reduced the chi-square to 43.4 on 16 degrees of freedom. This independence model was also fitted to the other 12 cities. It improved the fit to the St. Louis data, but failed to improve the fit or made it considerably worse in the other cities.

Table 4 Observed and Expected Frequencies for the Dirichlet-Gamma-Poisson Model of Personal Victimization in Baltimore (National Crime Survey, 1975)

Number and Type of Victimization* R A S L	Observed Frequency	Expected Frequency
0 0 0 0	21,511	21,512.8
One Victimization:		
1 0 0 0	551	542.8
0 1 0 0	299	319.7
0 0 1 0	295	315.5
0 0 0 1	349	300.2
Two Victimizations:		
2 0 0 0	58	53.3
0 2 0 0	25	26.0
0 0 2 0	29	25.6
0 0 0 2	15	24.0
1 1 0 0	27	26.3
1 0 1 0	21	25.9
1 0 0 1	12	24.7
0 1 1 0	32	15.3
0 1 0 1	6	14.5
0 0 1 1	6	14.4
Three Victimizations:		
3 0 0 0	4	7.4
0 3 0 0	4	3.2
0 0 3 0	6	3.2
2 1 0 0	4	3.5
2 0 1 0	0	3.4
2 0 0 1	2	3.3
Victimizations of One Type Collapsed into a Single Cell	2	5.6
Victimizations of More than One Type Collapsed into a Single Cell	46	34.7

*The abbreviations are:
R Robbery
A Aggravated Assault
S Simple Assault
L Larceny with Contact

Because the Dirichlet compounding was motivated by noting that the conditional probability of specific types of crime differ by sex and age categories, one might ask if the Dirichlet distribution would be needed if age and sex were held constant. A comparison of the fixed versus the Dirichlet model across eight combinations of age and sex for the 13 NCS cities combined into one data set showed that the Dirichlet model provides a far better description of the data than does the fixed model. The chi-square test statistics were reduced by from 50 to 80 percent under the Dirichlet compared to the fixed gamma-Poisson model. In other words, the conditional probability of each specific type of crime varies within as well as across age and sex categories.

Thus, the Dirichlet-gamma-Poisson model provides an excellent description of the multivariate distribution of crimes reported in the NCS. The underlying assumptions, namely that liability differs by person and that not all persons have the same conditional probability of each type of event, seem to describe a number of situations. Partially as a test of this hypothesis, the model was fitted to the bivariate distribution of major and minor disciplinary infractions reported in a year for 1,825 inmates in a Northeastern prison, as well as to the bivariate distribution of the number of episodes of respiratory and digestive illnesses of a group of office workers reported by Bates and Neyman (1952). The observed and expected number of disciplinary infractions are presented in Table 5, and the observed and expected number of illnesses are presented in Table 6. The mean number of infractions was about 2 per year, and the mean number of illnesses was about 6. The fit to both data sets is remarkable. Obviously, the Dirichlet-gamma-Poisson model shows potential for understanding far more than just criminal victimization.

Table 5 The observed and expected number of major and minor disciplinary infractions for prisoners under a Dirichlet-gamma-Poisson model*

Minor Infractions	Major Infractions 0	1	2	3	4	5	6 or more
0	723 724.8	107 128.3	38 39.4	12 15.2	11 6.7	4 3.2	3 3.8
1	248 229.6	72 81.9	32 33.5	18 15.2	8 7.5	3 3.9	3 5.1
2	114 102.8	49 48.9	32 23.5	18 11.9	7 6.3	4 3.5	6 4.9
3	66 52.6	23 29.4	18 15.8	9 8.6	4 4.8	5 2.8	9 4.2
4	31 29.1	11 18.1	12 10.4	7 6.0	5 3.5	2 2.1	4 3.4
5	10 16.9	12 11.3	6 6.9	2 4.2	5 2.5	1 1.6	1 2.6
6 or more	19 28.2	20 21.3	10 14.3	3 9.4	5 6.1	5 4.0	9 6.7

Statistics:**

Pearson chi-square = 53.4
Degrees of freedom = 45

Gamma-Poisson Parameter Estimates:

$\hat{m}$ = 2.05
$\hat{k}$ = .65

Dirichlet Parameter Estimates:

Major Violations = 1.39
Minor Violations = 2.48

*Data are based upon following prisoners for one year in a Northeastern Prison.

**The Pearson chi-square for the fixed gamma-Poisson model is 328.7 on 46 degrees of freedom

Table 6 The observed and expected number of office workers reporting digestive and respiratory illness under a Dirichlet-gamma-Poisson model*

Respiratory Illness	Digestive Illness 0	1	2 or more
0	41 38.8	5 7.5	0 2.5
1	36 37.4	8 11.1	5 5.2
2	35 31.1	13 12.0	8 7.2
3	24 24.7	8 11.4	6 8.3
4	24 19.2	10 10.1	8 8.7
5	20 14.8	7 8.6	13 8.5
6	11 11.4	6 7.2	10 8.0
7	7 8.7	7 5.9	8 7.2
8	7 6.7	3 4.8	10 6.4
9	5 5.1	3 3.8	6 5.6
10	4 3.9	2 3.1	8 4.8
11	3 3.0	1 2.4	6 4.0
12 or more	8 9.8	7 8.8	15 17.4

Statistics:** Pearson chi-square = 22.1
Degrees of freedom = 35

Gamma-Poisson Parameter Estimates: $\hat{m}$ = 5.99
$\hat{k}$ = 1.43

Dirichlet Parameter Estimates: Respiratory Illness = 8.58
Digestive Illness = .38

*Data are from Bates and Neyman(1952), Table 2, pp 230-231.

**The Pearson chi-square for the fixed gamma-Poisson model is 44.0 on 36 degrees of freedom.

SOME USES OF THE DIRICHLET-GAMMA-POISSON MODEL

The number of victimizations observed during one period can be used to estimate individual victimization rates (the λ parameter in the gamma distribution), to estimate conditional probabilities of each type of victimization, and to estimate the multivariate distribution of victimizations expected in future periods. These estimates are based upon definitions of conditional probability using the equations already introduced.

The liability rate for persons who experienced x_1 victimizations of type 1, x_2 of type 2, etc. may be expressed as:

$$P(\lambda|x_1,x_2,x_3,x_4) = ((k + m)/m)^{k+x_T}\, \lambda^{x_T+k-1}\, e^{-\lambda(k+m)/m} / \Gamma(k+x_T) \qquad (9)$$

which is itself a gamma distribution with mean $m(k+x_T)/(k+m)$ and exponent $k+x_T$. In other words, the expected liability rate for persons reporting a total of x_T victimizations is the mean of this conditional distribution. Confidence intervals for each person's λ parameter can be easily constructed (see Arbous and Kereich, 1951).

The conditional probability of each type of crime for a person with x_1 victimizations of type 1, x_2 of type 2, etc. may be written as:

$$P(p_1,p_2,p_3,p_4|x_1,x_2,x_3,x_4)$$

$$= \frac{\Gamma(\theta_T+x_T)\prod_{i=1}^{4} p_i^{\theta_i+x_i-1}}{\prod_{i=1}^{4}\Gamma(\theta_i+x_i)} \qquad (10)$$

This probability density function is a Dirichlet distribution with parameters $\theta_i + x_i$. Thus, the conditional probability that the next reported crime is type k can be estimated as $(\theta_k + x_k)/(\theta_T + x_T)$. This estimation only depends on the number and the type of crimes that have been reported in the past. It was used earlier to show how the model's estimate of a person's conditional probabilities can be interpreted as changing each time a new crime is reported.

The Dirichlet-gamma-Poisson model can be used to predict the multivariate distribution of victimizations in the future conditional on the number reported in the past by assuming that each person's rate λ as well as their conditional probability of each type of crime remain constant over time. Let the length of the observed time period equal one unit, and let the length of the future time period equal t units. Furthermore, let x_{ij} represent the number of victimizations of type j in period i, and let x_{iT} represent the total number of victimizations of all types observed in time period i. The bivariate probability of reporting x_{11}, x_{12}, x_{13}, x_{14} victimizations in the first period and x_{21}, x_{22}, x_{23}, x_{24} in the second period, conditional upon p_1, p_2, p_3 and p_4, may be expressed as:

$$P(x_{11},x_{12},x_{13},x_{14},x_{21},x_{22},x_{23},x_{24} \mid p_1,p_2,p_3,p_4)$$
$$= \int_0^{\infty} \prod_{i=1}^{4} P(x_{1j}|\lambda p_j)\, P(x_{2j}|\lambda t p_j)\, f(\lambda)d\lambda, \qquad (11)$$

where $P(x_{1j}|\lambda p_j)$ and $P(x_{2j}|\lambda t p_j)$ are Poisson random variables with means λp_j and $\lambda t p_j$, respectively; and where $f(\lambda)$ is the gamma density function. The unconditional bivariate distribution for the two periods is found by

multiplying this equation by the Dirichlet density function for p_1,p_2,p_3,p_4 and then integrating over all p_j values. The conditional probability of experiencing $x_{21},x_{22},x_{23},x_{24}$ victimizations in a future time period of length t, conditional upon experiencing $x_{11},x_{12},x_{13},x_{14}$ victimizations in a time period of length 1, is found by dividing the bivariate probability for two periods by the probability for the first period, as was given in equation (8). This may be expressed as:

$$P(x_{21},x_{22},x_{23},x_{24}|x_{11},x_{12},x_{13},x_{14}) =$$

$$\left[\frac{k+m}{k+m+mt}\right]^{k+x_{1T}} \frac{\Gamma(k+x_{1T}+x_{2T})}{\Gamma(k+x_{1T})} \left[\frac{tm}{k+m+mt}\right]^{x_{2T}}$$

$$\frac{1}{\prod_{j=1}^{4} x_{2j}!} \cdot \frac{\Gamma(\theta_T+x_{1T})}{\Gamma(\theta_T+x_{1T}+x_{2T})} \cdot \frac{\prod_{j=1}^{4}\Gamma(\theta_j+x_{1j}+x_{2j})}{\prod_{j=1}^{4}\Gamma(\theta_j+x_{1j})} . \qquad (12)$$

The probability of being victimized in the next period can be easily estimated by subtracting the probability of not being victimized from one.

The use of these equations for the 13 NCS city data set is illustrated in Tables 7 and 8. Table 7 shows the probability of reporting at least one victimization in the next year for persons reporting zero, one and two victimizations. Note that the pattern is quite similar across cities. About 4 to 5 percent of the persons who reported zero victimizations are expected to report one or more next year, about 20 to 25 percent of those who reported one are expected to report at least one next year, and about 30 to 40 percent of those persons reporting two victimizations are expected to report one or more victimizations next year. Only Newark differs considerably from this pattern.

Table 7 The Estimated Probability of Being Victimized at Least Once in the Next Year for Persons Who Reported Zero, One and Two Victimizations Under Dirichlet-Gamma-Poisson Models in 13 Cities
(National Crime Surveys, 1974-75)

	Number of Victimizations Reported		
City*	Zero	One	Two
Dallas	.036	.220	.370
Atlanta	.037	.186	.312
Newark	.038	.137	.226
St. Louis	.040	.205	.341
Philadelphia	.040	.223	.372
New York	.042	.183	.303
Los Angeles	.043	.223	.369
Portland	.045	.254	.418
Denver	.045	.258	.424
Cleveland	.049	.232	.379
Chicago	.051	.225	.368
Detroit	.054	.248	.403
Baltimore	.059	.259	.416

*Cities are ordered by the probability of being victimized next year for respondents who reported zero victimizations.

Table 8 displays the conditional probability that the next crime is a robbery for persons with a variety of victim histories across the 13 cities. The first column in the table, which displays the conditional probability that the next crime is a robbery for persons who did not report a victimization, is equivalent to the overall conditional probability of a robbery in each city under the Dirichlet model. Note that it ranges from .21 to .49 showing considerable variation in crime type by city. Ignoring Newark, Table 8 shows that this variability is reduced for persons reporting any combination of victimizations.

Thus, Tables 7 and 8 suggest that being a victim in a variety of cities may represent a common experience in that the chances of being victimized in the future as well as the chances of specific types of victimizations are far more variable for non-victims than for victims under the Dirichlet models. If NCS data were similar to UCR data, then the estimates in Tables 7 and 8 might be applicable to interpreting victim patterns in police data across a variety of cities. Research into the role that the Dirichlet-gamma-Poisson model might play in analyzing police data appears warranted.

Table 8 The Estimated Conditional Probability that the Next Crime Reported is a Robbery for Persons with a Variety of Victimization Histories under Dirichlet-Gamma-Poisson Models in 13 Cities
(National Crime Surveys, 1974-75)

	Victimization History: Number of Robberies Reported								
	0	1	2	0	1	2	0	1	2
	Number of Other Crimes Reported								
City[a]	0	0	0	1	1	1	2	2	2
Portland	.21	.42	.54	.15	.33	.45	.12	.27	.38
Dallas	.23	.42	.54	.18	.34	.45	.14	.28	.38
Denver	.25	.42	.52	.19	.34	.44	.16	.29	.38
Los Angeles	.28	.51	.61	.19	.39	.51	.14	.31	.42
St. Louis	.34	.50	.59	.26	.40	.50	.21	.34	.43
Philadelphia	.34	.60	.72	.20	.43	.56	.15	.34	.46
Atlanta	.35	.53	.64	.25	.41	.52	.19	.34	.44
Cleveland	.36	.57	.67	.35	.43	.54	.19	.34	.45
Baltimore	.37	.58	.69	.24	.43	.55	.18	.35	.46
Chicago	.38	.54	.63	.29	.43	.52	.23	.36	.45
New York	.41	.58	.68	.29	.45	.55	.23	.37	.47
Detroit	.44	.63	.72	.29	.47	.58	.22	.38	.48
Newark	.49	.51	.53	.47	.49	.51	.46	.47	.49

[a]Cities are ordered by the conditional probability of a robbery for persons who reported zero victimizations.

SUMMARY

The Dirichlet-gamma-Poisson model did an excellent job of describing the multivariate distribution of the number of personal victimizations reported in city samples of the NCS. It is based upon assumptions that seem applicable to a variety of analyses, namely that persons have a constant chance of experiencing events over time, but that not all persons have the same chances. Applied to victimization surveys, the model suggests that exposure to high crime situations is multidimensional because being highly exposed to one type of crime does not necessarily imply high exposure to other types of crime.

The analyses of the NCS were interpreted as if liability remained constant over time. This assumption is not needed to generate data with a Dirichlet-gamma-Poisson distribution. The distribution can also be generated by compounding a Dirichlet distribution with a negative binomial model, and the negative binomial model can be generated in a variety of ways (see Anscombe, 1959; Eaton and Fortin, 1978; and Feller, 1943). Further research using longitudinal data is needed to verify the interpretation of constant liability for crime data.

Even if liability were constant only for short periods--as for 6 or 12 months--the model would be useful for simplifying the comparisons of large, multivariate data sets and for predicting what would happen if liabilities were to remain constant. The analysis of the NCS showed that fairly complex differences between victimization patterns in 13 cities could be simplified by comparing the Dirichlet-gamma-Poisson parameters. Somewhat surprisingly the model suggests that being victimized may be a common experience in that the chances of victims being repeatedly victimized were less variable across cities than were the chances of non-victims being victimized.

The model is expected to be useful for policy development and program evaluation because it provides a means of estimating what would happen if conditions were to remain the same. For example, the relative impact of victim assistance programs designed to reduce the liability of persons who reported a relatively high number of crimes could be evaluated by estimating what would happen if no such program existed. The analysis of major and minor disciplinary infractions in a group of prisoners suggests that the model could be used to identify persons most likely to commit serious violations in the future based solely on their history of disciplinary infractions.

Methodologically, the model is easy to interpret because it is hierarchical to a series of simpler models. By varying the size of the sum of the Dirichlet parameters, the Dirichlet-gamma-Poisson model can range from a fixed gamma-Poisson model that allows for no event specialization to a gamma-Poisson model that allows for complete event specialization in that different types of events are mutually exclusive of each other. By varying the size of the exponent parameter, the model can be simplified to a Dirichlet-Poisson model. The model is also easy to estimate because the parameters in the Dirichlet part are independent of those in the gamma-Poisson part.

Lastly, the model represents a new perspective on relating individual and group level data. For example, rates are frequently compared across groups to show that the rate in one group is higher than in another.

Yet, rate differences do not necessarily imply that all persons in the high group have a greater chance of experiencing the event than persons in the low group. In fact, two groups could have the same rate but the individual level chances of experiencing the event could be very different in both groups. The Dirichlet-gamma-Poisson model provides a technique of comparing distributions of individual rates across groups based on repeated events. The utility of making assumptions about the distributions of individual rates and then comparing distributions across groups of persons will be borne out by future research.

REFERENCES

Anscombe, F.J., 1959. "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions." Biometrika 37(December):246-54.

Arbous, A.G., and J.E. Kerrich, 1951. "Accident Statistics and the Concept of Accident Proneness." Biometrics 7(December):340-432.

Bates, G.E., and J. Neyman, 1952. "Contributions to the Theory of Accident Proneness: I. An Optimistic Model of the Correlation Between Light and Severe Accidents." University of California Publications in Statistics, 1:215-253.

Cohen, L., and M. Felson, 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." American Sociological Review 44(August): 588-607.

Eaton, W.M., and A. Fortin, 1978. "A Third Interpretation for the Generating Process of the Negative Binomial Distribution." American Sociological Review 43(April):264-267.

Feller, W., 1943. "On a General Class of 'Contagious' Distributions." Annals of Mathematical Statistics 14(December):389-400.

Garofalo, J., and M.J. Hindelang, 1978. An Introduction to the National Crime Survey. Analytic Report SD-VAD-4. Law Enforcement Assistance Administration, National Criminal Justice Information and Statistics Service. Washington, D.C.: Government Printing Office.

Greenwood, M., and H.M. Woods, 1919. A Report on the Incidence of Industrial Accidents upon Individuals with Special Reference to Multiple Accidents. Report of the Industrial Fatigue Research Board, no. 4, London: IFRB.

Greenwood, M., and G.U. Yule, 1920. "Miscellanea: An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents." Journal of the Royal Statistical Society 83(March):255-279.

Hindelang, M.J., M.R. Gottfredson, and J. Garofalo, 1978. Victims of Personal Crime: An Empirical Foundation for a Theory of Personal Victimization. Cambridge, Mass.: Ballinger.

Johnson, N.L., and S. Kotz, 1972. Distributions in Statistics: Continuous Multivariate Distributions. New York: Wiley.

National Research Council, 1976. Surveying Crime. Washington: National Academy of Sciences.

Nelson, J.F., 1980a. "Multiple Victimization in American Cities: A Statistical Analysis of Rare Events." American Journal of Sociology 85(January):870-891.

Nelson, J.F., 1980b. "Alternative Measures of Crime: A Comparison of the Uniform Crime Reports and the National Crime Survey in Twenty-Six American Cities," in Crime: A Spatial Perspective, by Daniel E. Georges-Abeyie and Keith D. Harries (eds.). New York: Columbia University Press.

Patil, G.P., 1964. "On Certain Compound Poisson and Compound Binomial Distributions," Sankhya, Series A, 26, 293-294.

Silvey, S.D., 1970. Statistical Inference. Baltimore: Penguin Books.

Skogan, W., 1980. "Victimization and Exposure to Risk." A paper presented at the 32nd Annual Meeting of the American Society of Criminology. San Francisco.

Sparks, R.F., 1980. "Multiple Victimization: Evidence, Theory and Future Research," in Victimology Research Agenda Development. Volume I: Invited Papers, by J.S. Dahmann and J.H. Sasfy (eds.). McLean, Virginia: The MITRE Corporation.

Sparks, R.F., H.G. Genn, and D.J. Dodd, 1977. Surveying Victims: A Study of the Measurement of Criminal Victimization. New York: Wiley.

Appendix A: Extimation of m and k in the Gamma-Poisson Model

Maximum likelihood procedures were used to estimate the parameters of the gamma-Poisson model. The maximum likelihood estimate of m, denoted $\hat{m}$, is the observed mean number of victimizations. The maximum likelihood estimate of k, denoted $\hat{k}$, was iteratively computed using Newton's method (Silvey, 1970). The estimate of k computed at step j+1, $\hat{k}_{j+1}$, equals:

$$\hat{k}_{j+1} = \hat{k}_j - \frac{\sum_{i=0}^{\infty} f_i[\psi(i+\hat{k}_j) - \psi(\hat{k}_j) + \log(\hat{k}_j/(\hat{k}_j+\hat{m}))]}{\sum_{i=0}^{\infty} f_i[\psi'(i+\hat{k}_j) - \psi'(\hat{k}_j) + \hat{m}/(\hat{k}_j(\hat{k}_j+\hat{m}))]}, \quad (13)$$

where f_i is the observed frequency of persons reporting i victimizations, $\psi(x)$ is the derivative of the gamma functions of x with respect to x, and $\psi'(x)$ is the derivative of $\psi(x)$. The iterations were continued until the difference between $\hat{k}_{j+1}$ and $\hat{k}_j$ was less than .00005. This usually occurred within three to five steps. The initial value of k, $\hat{k}_1$, was obtained by the method of moments (Anscombe, 1959) from:

$$\hat{k}_1 = \hat{m}^2 / (s^2 - \hat{m}) \quad (14)$$

where s^2 is the sample variance and $\hat{m}$ is the sample mean.

Appendix B: Estimation of the Dirichlet Parameters

Maximum likelihood estimates of the m and k parameters can be easily obtained by fitting a univariate gamma-Poisson model to the total number of victimizations. These estimates of m and k are independent of the Dirichlet parameters.

Maximum likelihood estimates of the Dirichlet parameters were interatively computed from the following equation:

$$\hat{D}_{j+1} = \hat{D}_j - (D2)^{-1}\, D1 \quad (15)$$

where $\hat{D}_j$ stands for the jth computed value of the vector of Dirichlet parameters, D2 stands for the second derivative of the natural logarithm of the likelihood function of the Dirichlet-gamma-Poisson model, and D1 stands for the first derivative of the natural logarithm of likelihood function.

---

[5]Although the values of D2 and D1 change at each iteration, subscripts indicating iteration cycle have been dropped to simplify notation.

The values for D1 and D2 used at calculation j+1 were estimated from the parameters in D estimated at calculation j. The ith row of the vector D1 was calculated from:

$$D1_i = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \sum_{x_3=0}^{\infty} \sum_{x_4=0}^{\infty} f(x_1,x_2,x_3,x_4) g(\hat{\theta}_T,x_T,\hat{\theta}_i,x_i), \quad (16)$$

where $g(\hat{\theta}_T,x_T,\hat{\theta}_i,x_i) = \psi(\hat{\theta}_T) - \psi(\hat{\theta}_T+x_T) + \psi(x_i+\hat{\theta}_i) - \psi(\hat{\theta}_i)$,

and where $f(x_1,x_2,x_3,x_4)$ is the observed frequency of persons reporting x_1 victimizations of type 1, x_2 of type 2, etc. The summations range from zero to the maximum number of each type of victimization reported in the data set.

The ith row and jth column elements of the matrix D2 for $i \neq j$ are all the same and were calculated from:

$$D2_{ij} = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \sum_{x_3=0}^{\infty} \sum_{x_4=0}^{\infty} f(x_1,x_2,x_3,x_4)[\psi'(\hat{\theta}_T) - \psi'(\hat{\theta}_T+x_T)] \quad (17)$$

where $\psi'(x)$ is the first derivative of $\psi(x)$. The elements on the main diagonal of D2 were calculated from:

$$D2_{ii} = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \sum_{x_3=0}^{\infty} \sum_{x_4=0}^{\infty} f(x_1,x_2,x_3,x_4) \; g'(\hat{\theta}_T,x_T,\hat{\theta}_i,x_i), \quad (18)$$

where $g'(\hat{\theta}_T,x_T,\hat{\theta}_i,x_i) = \psi'(\hat{\theta}_T) - \psi'(\hat{\theta}_T+x_T) + \psi'(x_i+\hat{\theta}_i) - \psi'(\hat{\theta}_i)$.

Initial estimates of the θ_i parameters were obtained by arbitrarily setting $\hat{\theta}_1$ to 1, by estimating $\hat{p}_1$ to $\hat{p}_4$, and by using equation (8) to estimate $\hat{\theta}_2$, $\hat{\theta}_3$, and $\hat{\theta}_4$. Note that setting $\hat{\theta}_1$ to 1 suggests that $\hat{\theta}_T = 1/\hat{p}_1$.

This procedure worked for most but not for all cities. In one city, the initial value of $\hat{\theta}$, had to be set to .5 for the iterative procedure to converge.

END