58134

# An Approach to Interval Estimation In Crime Analysis

by

Yoshio Akiyama, Ph. D.

Uniform Crime Reports
Federal Bureau of Investigation
Washington, D. C. 20535

Abstract

This paper presents a probability model that refines the conventional percent change analysis of crimes by introducing an interval estimation. The model assumes that crime incidents are Poisson distributed with some unknown parameter viewed as a measure of criminality. Using a Bayesian approach, criminality is then expressed by a Gamma distribution with observed volumes of crime serving as a parameter. Under the assumption that the level of criminality in two consecutive periods are independent, an upper and a lower bound (for percent changes) are expressed using a Beta function. These upper and lower bounds which give rise to an interval estimation for percent changes are functions of crime volume, percent change, and the prescribed level of confidence. An asymptotic approximation through standard normal is introduced for purposes of computational simplicity. An example of application is found in smaller Uniform Crime Reports (UCR) contributing agencies where instability of percent change variable often presents a problem of interpretation. The technique promulgated is considered useful in this situation. The concept presented in this paper has also been applied to the development of a UCR data quality control model.

## I. INTRODUCTION



In terms of percent changes, crime volumes for larger agencies tend to be more stable than those for smaller agencies. This means that an agency's crime conditions would be better understood or described if the conventional percent change analysis is augmented by additional information.

The approach taken in this paper, in coping with the above problem, is the interval estimation of percent change variable. Given a percent change value C, a lower limit of change A and an upper limit of change B will be computed for C. The size B-A of the interval [A;B] will depend on the observed volumes of crime. Furthermore, the size of the interval will serve as an indicator of the reliability of the given percent change figure C. To be more specific, a large interval size B-A (which happens for smaller volumes of crime) indicates a wide variation or a low reliability of C. On the other hand, a small interval size B-A (which happens for larger volumes of crime) indicates a narrow variation or a high dependability of the value C.

The lower and upper limits, A and B, will depend on the prescribed level of confidence, the degree of accuracy that a person desires to maintain. The interval [A;B] indicates the estimated range of variation that the percent change variable could have taken. In other words, instead of taking one percent change number as a rigid basis of analysis, this paper attempts to estimate an interval or a range in which the value C could have fallen under a prescribed level of confidence.

## II. INTERVAL FOR CRIMINALITY TREND

Let X be a random variable which is Poisson distributed with some unknown parameter $\theta > 0$, i.e.,

$$p(x) = \frac{e^{-\theta}\theta^{x}}{x!}, \quad x = 0, 1, 2, \ldots.$$

Using the prior density function $p(\theta) = 1/\theta$, $\theta > 0$, in Bayesian procedure, the posterior density function given $X = x$ is:

$$p(\theta|x) = \frac{p(\theta)\, p_\theta(x)}{\int_0^\infty p(\theta) p_\theta(x)\, d\theta}$$

$$= \frac{e^{-\theta}\,\theta^{x-1}}{\int_0^\infty e^{-\theta}\theta^{x-1} d\theta}$$

$$= \frac{e^{-\theta}\theta^{x-1}}{\Gamma(x)};$$



where $\Gamma(x) = \int_0^\infty e^{-\theta}\,\theta^{x-1} d\theta$ is the Gamma function. $\Gamma(x) = (x-1)!$ if x is an integer. This means that $\theta$ is Gamma distributed with parameter x. The expectation $E(\theta|x)$ and the variance of $\theta$ are both x.

Let X and Y be both Poisson distributed with parameter $\theta$ and $\xi$, respectively. If $\theta$ and $\xi$ are independently Gamma distributed with parameters x and y, then the variable $u = \theta/(\theta+\xi)$ is a Beta distribution with parameter (x,y), i.e.,

$$p(u) = \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)} u^{x-1}\cdot(1-u)^{y-1}$$

$$= \frac{(x+y-1)!}{(x-1)!\,(y-1)!} u^{x-1}\cdot(1-u)^{y-1}, \quad 0<u<1.$$

Note that the independence of $\theta$ and $\xi$ is different from that of X and Y.

Given $\varepsilon>0$. Define $a=a(\varepsilon;x,y)$ and $b=b(\varepsilon;x,y)$ by the relation $P(u<b) = P(u>a) = \varepsilon/2$, so that

$$P(b \leqq u \leqq a) = P(-2 + 1/a \leqq (\xi-\theta)/\theta \leqq -2 + 1/b) = 1-\varepsilon.$$

In application, $\theta$ and $\xi$ are viewed as criminality measures that have yielded the number of crimes $x$ for the past year and $y$ for the current year. Note that the percent change in criminality is then $100(\xi-\theta)/\theta$ percent. According to the above paragraph, this percent change in criminality lies in the interval $[-200 + 100/a; -200 + 100/b]$ for the level $\varepsilon$.

$P(u<b)$ is an incomplete Beta function, which is related to Binomial distribution. More precisely, it can be shown that

$$\begin{aligned}
&P(u<b) \\
&= \frac{\Gamma(x+y)}{\Gamma(x)\cdot\Gamma(y)} \int_0^b u^{x-1}(1-u)^{y-1}du \\
&= \sum_{i=x}^{x+y-1} \binom{x+y-1}{i} b^i (1-b)^{x+y-1-i} \\
&= 1 - \sum_{i=0}^{x-1} \binom{x+y-1}{i} b^i \cdot (1-b)^{x+y-1-i} \\
&= 1 - \sum_{i=0}^{x'} \binom{x'+y}{i} b^i \cdot (1-b)^{x'+y-i}, \quad \text{where } x' = x-1.
\end{aligned}$$

The equation $P(u<b)=\varepsilon/2$ then becomes

$$\sum_{i=0}^{x'} \binom{x'+y}{i} b^i \cdot (1-b)^{x'+y-i} = 1-\varepsilon/2.$$

When $x'+y$ is sufficiently large, the central limit theorem applies. If $x'+y \geqq 20$, for instance, set

$$\frac{x' - b(x'+y)}{\sqrt{(x'+y)b(1-b)}} \doteq z,$$

where $z$ is the $(1-\varepsilon/2)$-percentile for the standard normal distribution. This equation can be solved for $b$. Likewise, $a$ can be obtained by solving the equation:

$$\frac{x' - a(x'+y)}{\sqrt{(x'+y)a(1-a)}} \doteq -z.$$

Such $a$ and $b$ determine, for the level $\varepsilon$, the interval $[A;B] = [-200+100/a;\ -200+100/b]$ for the percent change $C=100(y-x)/x$. Finally,

$$a = \frac{(x'+y)(2x'+z^2) + \sqrt{(x'+y)^2 z^4 + 4(x'+y)x'yz^2}}{2(x'+y)(x'+y+z^2)}, \text{ and}$$

$$b = \frac{(x'+y)(2x'+z^2) - \sqrt{(x'+y)^2 z^4 + 4(x'+y)x'yz^2}}{2(x'+y)(x'+y+z^2)}.$$

TABLE 1. ILLUSTRATION

| Agency | Number of Crimes Last Year | This Year | Percent Change | Intervals [A;B] ε=.2 | ε=.3 |
|---|---|---|---|---|---|
| 1 | 10 | 12 | +20.0 | [-27;126] | [-14;110] |
| 2 | 100 | 120 | +20.0 | [ 2; 44] | [ 5; 39] |
| 3 | 1000 | 1200 | +20.0 | [ 14; 27] | [ 15; 26] |
| 4 | 5000 | 6000 | +20.0 | [ 17; 23] | [ 18; 22] |
| 5 | 500 | 480 | - 4.0 | [-11; 4] | [-10; 3] |
| 6 | 500 | 510 | + 2.0 | [- 6; 11] | [- 4; 9] |
| 7 | 500 | 532 | + 6.4 | [- 2; 15] | [ 0; 14] |
| 8 | 193 | 180 | - 6.7 | [-18; 7] | [-16; 4] |
| 9 | 165 | 180 | + 9.1 | [- 4; 26] | [- 2; 23] |
| 10 | 171 | 180 | + 5.3 | [- 7; 21] | [- 5; 18] |
| 11 | 15 | 30 | +100.0 | [ 42;223] | [ 54;199] |
| 12 | 47 | 55 | +17.0 | [- 7; 54] | [- 3; 46] |
| 13 | 145 | 138 | - 4.8 | [-17; 12] | [-15; 8] |
| 14 | 760 | 725 | - 4.6 | [-11; 2] | [- 9; 1] |
| 15 | 1550 | 1600 | + 3.2 | [- 1; 8] | [ 0; 7] |

END