DISTRIBUTION OF THE RATIO OF TWO
POISSON RANDOM VARIABLES

by

TRALISSA F. GRIFFIN, B.A.

A THESIS

IN

STATISTICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

MASTER OF SCIENCE

Approved

Accepted

August, 1992

ACKNOWLEDGEMENTS

I would like to express my sincerest appreciation to my advisor, Dr. Benjamin Duran, for the support he has offered during a most hectic year. He has been a great encouragment throughout this research. I also wish to thank my committee, Drs. Yang and Mansouri, for being wonderful teachers as I prepared for this experience. Most of all, I want to thank my husband, David, and my family for being patient and supportive of me during this long and difficult year. This is the culmination of much work by not only myself, but everyone I have mentioned.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The ratio of two Gaussian (normal) random variables has been the topic of numerous reports, and the information concerning this ratio is abundant. Since the normal distribution stems from a continuous random variable, the maneuverability is somewhat elementary and is preferred over such cases as a discrete random variable. On the other hand, the amount of information known about a ratio of discrete random variables, in particular those having a Poisson distribution, is very limited. The reason could possibly be due to the greater complexity of this type of discrete random variable or maybe the ability to utilize the more familiar normal approximations. Common knowledge leads one to note if the mean of the Poisson random variable is *sufficiently large*, say five or more, the normal approximation is quite good.

The idea of approximating each of the Poisson random variables in a ratio directs one to the question: "How is the distribution of the ratio affected by the use of the normal approximations for each Poisson random variable X and Y in the ratio $R = X/Y$?" In a report by Duran and McCready [1991] this question arose; however, the actual distribution was not deemed necessary as the assumption of sufficiently large means allowed for the use of the normal approximations. Therefore, the results to follow shall begin with the Poisson distribution. In addition to the important formulations of this distribution, some historical remarks will be noted. The distribution of the ratio, R, of two random variables each having the Poisson distribution will be determined along with that of the truncated Poisson, truncated at $Y = 0$. Following the actual distribution of the aforementioned ratio, the normal approximations of the Poisson random variables will be considered as well as the ratio of these normal random variables. Next a method of estimation will be derived using the distribution of the normal approximations to the Poisson random variables. Finally, a brief discussion of further applications will be provided which includes those in fields other than mathematics and topics for further research.

CHAPTER II

DISTRIBUTION OF THE RATIO OF RANDOM VARIABLES

Poisson's Distribution

Simeon D. Poisson (1781-1840), the noted scientist and mathematician, is attributed with the discovery of the distribution bearing his name. Even though some mathematical historians, such as Hald [1990], note that Abraham de Moivre (1667-1754) actually derived the result prior to Poisson. However, de Moivre never explicitly stated the exact formulation of the limiting distribution for Bernoulli's binomial distribution.

Poisson's main work on probability theory, containing this well-known theorem, was published in 1837: "Recherches sur la probibilité des jugements en matière criminelle et en matière civile." (Note: Haight [1967] mentioned that the date 1832 had been given by some authors for the publication of this work by Poisson; however, he continued the Catalogue Generale des Livres Imprimes de la Bibliotheque Nationale (Vol. 193, p. 982) gave the date 1837 and prior to this date there was no mention of Poisson's "Recherches sur ...") In this book, he posed the problem of applying probability theory to court decisions. Not concerned with the probabilities pertaining to any one trial; but considering a series of trials, he found that a relationship existed and drafted a formula for the

Poisson distribution known today. His investigation of limiting theorems of probability theory produced what he called the "law of large numbers."

Adolphe Quetlet (1796-1874) followed the same route in his research, minus the mathematical probability, with similar results provided in his book of 1835: "Sur l'homme et le développement de ses facultés, ou essai de physique sociale." Stigler [1986] made mention of the acquaintance between Poisson and Quetlet and the fact that Poisson's data very closely resembles that of Quetlet just two years prior. Based on their common idea that moral as well as physical events are subject to a universal law, Poisson composed his theorem. Poisson's theorem is stated concisely by Maistrov [1974] as follows:

> If n independent trials are performed, resulting in the occurrence or non-occurrence of an event A, and the probability of occurrences of events is not the same in each one of the trials, then, with probability as close to unity as desired (in other words, as close to certaintity as desired), one can assert that the frequency $\frac{m}{n}$ of the occurrence of event A will deviate arbitrarily little from the arithmetic mean $\tilde{p}$ of probabilities of occurrences of events in the individual trials. This theorem is formulated in modern notation as follows:
>
> $$\lim_{n \to \infty} (|\frac{m}{n} - \tilde{p}| < \epsilon) = 1.$$
>
> If, however, the probability of the occurrences of events remain constant from trial to trial, then $\tilde{p} = p$, and Poisson's theorem in this case reduces to Bernoulli's theorem. (p. 159)

Poisson also wrote, in the same book, the "law of small numbers." He discovered as p, else $q = 1 - p$, deviated greatly from $\frac{1}{2}$ and as n increased asymptotically, the normal approximation for the binomial distribution derived by Bernoulli

$$(2npq\pi)^{-1/2} exp \left\{ \frac{-(x - np)^2}{2npq} \right\}$$

became less accurate for the probability of m occurrences of event A in n trials, $P_{m,n}$. He then showed as $n \to \infty$ and $p_n \to 0$, this probability took the form $P_{m,n} = \frac{\lambda^m}{m!}e^{-\lambda}$ where $\lambda = np_n$, a constant. Haight [1967] noted that "the value of p_0 is singled out for special mention, and the cumulative form is observed to converge to unity" (p. 113).

Although Poisson made this discovery, his further works were not concerned with this formula. It was Laudislaus Von Bortkiewicz (1868-1931) who found the existence of a correspondence between Poisson's formula and certain types of countable (discrete) populations. Bortkiewicz, in studying such events as suicides of Prussian children, births of triplets, and (perhaps his most famous study) deaths by horse kick in the Prussian army, discovered these types of data worked well in Poisson's formula. Since the probabilities of these events are very small in a large number of occurrences, the "law of small numbers" can easily be applied. In his monograph from 1898 entitled "Das Gesetz der kleinen Zahlen," Bortkiewicz developed the actual Poisson distribution using the above studies. Included in this work was a myriad of information on the Poisson distribution such as the limiting distribution to the binomial as derived previously by Poisson, with credit to Poisson; the mean, variance, mean deviations, and other moments of this distribution were computed. Differential and difference relations for the probabilities were considered in "Das Gesetz ..." along with the normal limit to this distribution and other relationships, applications and tables were included. Altogether this was a concise and complete work which enhances the importance of Bortkiewicz's contribution "by the fact that it came about quite abruptly as the work of a single individual" (Haight [1967], p. 115).

Poisson's "law of small numbers" and the distribution derived by Bortkiewicz may now be employed in the following problem: Define the random variables X and Y by their probability density functions (p.d.f.)

$$X \sim p(x, \lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \text{ and } Y \sim p(y, \mu) = \frac{\mu^y}{y!}e^{-\mu}$$

where $\sim$ means "distributed as" and $p(a, b)$ is the Poisson formula for the random variable a with expected value b. The ranges of X and Y are each a discrete set. Without loss of generality, consider the range of X to be $x = 0, 1, 2, \ldots$ and $y = 0, 1, 2, \ldots$ for Y. As a matter of notation, let $x \in \mathbf{Z}_0^+$, defined to be the set of non-negative integers, likewise $y \in \mathbf{Z}_0^+$. Assuming the random variables X and Y to be independent and making a transformation, the ratio $R = X/Y$ will be of major interest at this time.

Truncated Poisson Distribution

Notice, however, when $Y = 0$ the denominator vanishes, hence R is undefinable. To render this problem, one must truncate the range of Y in such a way that $P(Y = 0) = 0$, the probability that Y takes on the value zero is zero; and consequently $P(Y \geq 1) = 1$, or the probability Y takes a value larger than zero is one.

The truncated probability density function (p. d. f.) of a Poisson random variable may be derived as follows:

Considering the p. d. f. of a Poisson random variable, Z,

$$p(z, \mu) = \frac{\mu^z}{z!}e^{-\mu} \text{ for } z \in \mathbf{Z}_0^+.$$

Note that

$$\begin{aligned} P(Z \geq 0) = 1 &= P(Z = 0) + P(Z > 0) \\ &= e^{-\mu} + P(Z > 0) \end{aligned}$$

which implies

$$P(Z > 0) = 1 - e^{-\mu}.$$

Then by dividing both sides by $1 - e^{-\mu}$, the truncated p. d. f. is produced

$$p(z, \mu) = \frac{\mu^z e^{-\mu}}{z!(1 - e^{-\mu})} \text{ for } z \in \mathbf{Z}^+.$$

And hence (as hoped) the probability $P(Z > 0) = P(Z \geq 1) = 1$, along with the fact that $P(Z = z) > 0$ for every $z \in \mathbf{Z}^+$ satisfying the criterion for a function to be a probability density function.

Considering the effect of truncation on the mean, the expected value of Z may be determined.

$$\begin{aligned} E(Z) &= \sum_{z=1}^{\infty} z p(z, \mu) \\ &= \sum_{z=1}^{\infty} \frac{z \mu^z e^{-\mu}}{z!(1 - e^{-\mu})} \\ &= \frac{\mu}{1 - e^{-\mu}}. \end{aligned}$$

The variance of Z may also be developed in the light of this truncated p. d. f. using the moment generating function derived as follows:

$$\begin{aligned} M_z(t) = E(e^{zt}) &= \sum_{z=1}^{\infty} e^{zt} \frac{\mu^z e^{-\mu}}{z!(1 - e^{-\mu})} \\ &= \frac{e^{-\mu}}{1 - e^{-\mu}} \sum_{z=1}^{\infty} \frac{(e^t \mu)^z}{z!} \\ &= \frac{e^{-\mu}}{1 - e^{-\mu}} \left[e^{\mu e^t} - 1 \right]. \end{aligned}$$

So the variance of Z will be

$$\begin{aligned} Var(Z) &= M_z''(0) - (M_z'(0))^2 \\ &= \frac{\mu(1+\mu)}{1-e^{-\mu}} - (\frac{\mu}{1-e^{-\mu}})^2 \\ &= \frac{-\mu(\mu e^{-\mu} - (1-e^{-\mu}))}{(1-e^{-\mu})^2} \end{aligned}$$

where $M_z'(0) = \frac{d}{dt}M_z(t)|_{t=0}$ and $M_z''(0) = \frac{d^2}{dt^2}M_z(t)|_{t=0}$.

Much work has been reported concerning the truncated Poisson distribution, especially in the area of the estimation of the mean. Some of these are David and Johnson [1952], Plackett [1953], Cohen [1954], Moore [1954], Irwin [1959], and Doss [1963]. The most general of these, and most practical, is that of Cohen [1954] on "Estimation of the Poisson parameter from truncated samples or censored samples." For the sample truncated on the left at zero, the log-likelihood function used in estimating the parameter (the mean), μ is

$$\begin{aligned} L &= -n\ln(P_z(1)) - n\mu + n\bar{z}\ln\mu - \ln(\Pi_{i=1}^n z_i!) \\ &= -n\ln(1-e^{-\mu}) - n\mu + n\bar{z}\ln\mu - \ln(\Pi_{i=1}^n z_i!) \end{aligned}$$

where $P_z(c)$ is the cumulative probability of $z \geq c$ occurrences, and is written

$$P_z(c) = \sum_{z=c}^{\infty} \frac{e^{-\mu}\mu^z}{z!}.$$

To obtain an estimate for μ, differentiate the log-likelihood function and equate to zero as follows:

$$\frac{1}{n}\frac{d}{d\mu}L = \frac{\bar{z}}{\mu} - 1 - \frac{e^{-\mu}}{1-e^{-\mu}} = 0.$$

Hence, by solving, it may be concluded that $\bar{z}$ is an estimate for the mean; which is true in general since $\bar{z}$ is the minimum variance unbiased estimator for the mean.

Ratio of Two Poisson Random Variables

Utilizing the truncated density for Y, a transformation may be made as follows: $R = X/Y$ and $S = Y$. Then obtaining the inverse transform: $x = rs$ and $y = s$, the joint p. d. f. of R and S may be obtained from the p. d. f.'s of X and Y as

$$f(r,s) = p(x(r,s),\lambda)p(y(r,s),\mu) = \left[\frac{\lambda^{rs}e^{-\lambda}}{(rs)!}\right]\left[\frac{\mu^{s}e^{-\mu}}{s!(1-e^{-\mu})}\right] \text{ for } rs \in \mathbf{Z}_0^+, s \in \mathbf{Z}^+$$

by the assumption of the independence of X and Y.

The random variable of interest here is R, therefore, it becomes necessary to find the marginal density of R. This marginal density may be formulated from the joint distribution as follows:

$$f_1(r) = \sum_{s \in A} f(r,s) = \sum_{s \in A} \frac{(\lambda^r \mu)^s e^{-(\lambda+\mu)}}{(rs)!s!(1-e^{-\mu})} \text{ for } r \in \mathbf{Q}_0^+$$

where $A = \{s | rs \in \mathbf{Z}_0^+\}$. This distribution is not of a well-known form; however, it may be computed by the expansion of the infinite sum.

Difficulties Arise

When considering the actual distribution of the ratio of two Poisson random variables with the denominator truncated on the left at zero, one may employ the computer to calculate the exact probabilities of $R = X/Y$. The simplicity of the formula for the probability function of R leads one to believe that numerical calculations may be straightforward. However, in this instance many difficulties are faced in computing the desired probabilities.

The procedure utilized in computing the probabilities here involves inputting the means for X and Y and calculating an estimate for the mean of R. Then the marginal p. d. f. of R

$$f_1(r) = \sum_{s \in A} \frac{(\lambda^r \mu)^s e^{-(\lambda+\mu)}}{(rs)!s!(1-e^{-\mu})}$$

is computed by summing terms until the absolute difference between two consecutive terms is small, say less than 10^{-6}. The natural log of each term was utilized to speed computation and eliminate errors, while maintaining accuracy. The factorials in the denominator of this equation dictate the method of summation in that only those values of s that make $rs \in \mathbf{Z}_0^+$ create a corresponding term in this series. Thus arises the first difficulty.

For any rational number, r, there exists an integer s, such that rs is also an integer, namely $s = ny$ for any number $n = 0, 1, 2, \ldots$ and y from the range of the random variable Y. Considering this fact, possible values of R may be constructed in the following manner:

Begin by letting $x = 0$ and y taking any value in the range $1, 2, 3, \ldots$; then $r = 0$. Let $x = 1$ and $y = 1, 2, 3, \ldots$; then $r = 1, \frac{1}{2}, \frac{1}{3}, \ldots$ are the values obtained. For $x = 2$ and $y = 1, 2, 3, \ldots$; $r = 2, 1, \frac{2}{3}, \ldots$ are yielded. Continuing in this manner, values are constructed as follows:

$$r = 0, \frac{1}{2}, \frac{1}{3}, \ldots, 1, \frac{3}{2}, \frac{4}{3}, \ldots, 2, \ldots$$

Since the range of R is the set of all non-negative rational numbers, $\mathbf{Q}_0^+$, it is countable; countable meaning it is isomorphic to the natural numbers or there exists a one-to-one correspondance between the range of R and the natural numbers. This correspondance may be seen by letting the numerator and the denominator denote the values of an ordered pair (i.e., $\frac{a}{b}$ becomes (a,b), a point in the Cartesian plane). By allowing each point to correspond to one of the natural numbers, the isomorphism is established.

Furthermore, there exists an infinite number of ways to create any one of these values. For example,

$$r = \frac{1}{2} \Longrightarrow \frac{x}{y} = \frac{1}{2} = \frac{2}{4} = \frac{3}{6} = \cdots = \frac{n}{2n} = \cdots \quad \text{for } n = 1, 2, 3, \ldots$$

In computing $P(R = \frac{1}{2})$, one must realize that for any of the above values of r there exists a positive probability, however small that probability may be. Since $\sum_r P(R = r) = 1$ and there are infinitely many values of r each with positive probability, the value of $P(R = \frac{1}{2})$ is difficult to actually calculate, even with the aid of the computer.

For computer calculation, a typical beginning point in determining probabilities would be the mean of the distribution. Since the mean of R is very complicated, an estimate was used, namely the ratio of the means of X and Y. Logically, the probability that R equals the ratio of λ and $\gamma = \frac{\mu}{1-e^{-\mu}}$ should be of significance. However, many times during the computation process, probabilities very near zero appeared in unexpected places. Hence, another difficulty arises even though computer calculations prove to be correct. As an example, choosing the mean of X to be large (i.e., large enough to consider the normal approximation), say $\lambda = 5$ and choosing the mean of Y also large, $\mu = 25$, so that $\gamma = 25$; the probability determined via computer is as follows:

$$P(R = \frac{5}{25}) = \sum_{s \in A} \frac{(25\sqrt[5]{5})^s e^{-30}}{(\frac{s}{5})! s! (1 - e^{-25})^{s+1}} = 0.032406788$$

where $A = \{s | \frac{s}{5} \in \mathbf{Z}_0^+\}$. However, at one standard deviation away from the mean on either side the probability was found to be zero. That is, the standard deviation for R was approximated by the ratio

$$\frac{(\lambda\gamma)(\lambda+\gamma)}{(1-e^{-\mu})^4}$$

and the probabilities produced are $P(R = \frac{5}{25} + \frac{2}{125}) = 0$ and $P(R = \frac{5}{25} - \frac{2}{125}) = 0$. Since for each rational value between $\frac{5}{25}$ and $\frac{5}{25} + \frac{2}{125}$, there exists a positive probability, the computer calculations are plausible. Thus, the actual distribution cannot be easily computed for all values of R. For this reason, the normal approximation is used as a general rule.

Normal Distribution

The difficulty in calculating the distribution of R prompts the use of the normal approximations to X and Y in approximating the distribution of the ratio, R. It is known that the Poisson distribution may be approximated by the normal distribution under the provision of *sufficiently large* means. Then conceivably, the ratio of two independent Poisson random variables may be approximated by the ratio of two independent normal random variables.

The ratio of two normal random variables is found in many works such as Geary [1930], Fieller [1932], Marsaglia [1965], and Hinkley [1969]. Of these papers, Hinkley's is the most appropriate in this instance. Hinkley formulated the exact distribution for the ratio of two correlated normal random variables and also considered an approximation in comparison with this distribution. His results may be stated as follows:

Let X and Y be normally distributed with respective means, μ_x and μ_y, and variances, σ_x^2 and σ_y^2. Let ρ be the correlation coefficient of X and Y. Then the joint p. d. f. of X and Y may be represented by the bivariate normal of the form

$$\begin{aligned} f(x,y) &= \frac{1}{\sigma_x\sigma_y\sqrt{2\pi(1-\rho^2)}} exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right.\right. \\ &\left.\left. -2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}, \text{ for } x,y\in\mathbf{R}. \end{aligned}$$

By a similar transformation as performed previously, let $R = X/Y$ and $S = Y$. So the inverse transform yields $x = rs$ and $y = s$. Since X and Y are continuous random variables (i.e., have a continuous population), the probability $P(Y = 0) = 0$ implies that there exists no reason for truncation in this formulation.

Then the joint p. d. f. of R and S may be derived from the joint p. d. f. of X and Y as follows:

$$g(r,s) = f(x(r,s), y(r,s))|J|$$

where $|J|$ is the absolute value of the determinant of the Jacobian matrix, the matrix of partial derivatives, that is,

$$|J| = abs\left(\left|\frac{\partial(x,y)}{\partial(r,s)}\right|\right) = abs\left(\begin{vmatrix} s & r \\ 0 & 1 \end{vmatrix}\right) = |s|.$$

Then

$$\begin{aligned} g(r,s) &= \frac{|s|}{\sigma_x\sigma_y\sqrt{2\pi(1-\rho^2)}} exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{rs-\mu_x}{\sigma_x}\right)^2\right.\right. \\ &\left.\left. -2\rho\left(\frac{rs-\mu_x}{\sigma_x}\right)\left(\frac{s-\mu_y}{\sigma_y}\right)+\left(\frac{s-\mu_y}{\sigma_y}\right)^2\right]\right\}, \text{ for } r,s\in\mathbf{R}. \end{aligned}$$

Now to find the p. d. f. of the random variable of concern, R, integrate the joint p. d. f. with respect to S, obtaining the marginal p. d. f. of R.

$$g_1(r) = \int_{-\infty}^{\infty} g(r,s)ds.$$

By completing the square in the exponential and making proper substitutions, one may conclude, as did Hinkley, that

$$\begin{aligned}
g_1(r) &= \frac{b(r)d(r)}{\sigma_x\sigma_y a^2(r)\sqrt{2\pi}}\left[\Phi\left\{\frac{b(r)}{a(r)\sqrt{1-\rho^2}}\right\} - \Phi\left\{\frac{-b(r)}{a(r)\sqrt{1-\rho^2}}\right\}\right] \\
&\quad + \frac{\sqrt{1-\rho^2}}{\pi\sigma_x\sigma_y a^2(r)} exp\left\{\frac{-c}{2(1-\rho^2)}\right\},
\end{aligned}$$

where

$$\begin{aligned}
a^2(r) &= \frac{r^2}{\sigma_x^2} - \frac{2r\rho}{\sigma_x\sigma_y} + \frac{1}{\sigma_y^2}, \\
b(r) &= \frac{r\mu_x}{\sigma_x^2} - \frac{\rho(\mu_x + r\mu_y)}{\sigma_x\sigma_y} + \frac{\mu_y}{\sigma_y^2}, \\
c &= \frac{\mu_x^2}{\sigma_x^2} - \frac{2\rho\mu_x\mu_y}{\sigma_x\sigma_y} + \frac{\mu_y^2}{\sigma_y^2}, \\
d(r) &= exp\left\{\frac{b^2(r) - ca^2(r)}{2(1-\rho^2)a^2(r)}\right\} \\
\text{and } \Phi(\tau) &= \int_{-\infty}^{\tau} \phi(u)du, \text{ where } \phi(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}.
\end{aligned}$$

CHAPTER III

COMPARISON

Normal Approximations

Considering X^* and Y^* to be the normal approximations of the Poisson random variables X and Y, Hinkley's results may be applied to the case of independent normal random variables. In this instance, the parameters for X^* and Y^* will be

$$\mu_x = \lambda \quad , \quad \sigma_x^2 = \lambda$$
$$\mu_y = \mu \quad , \quad \sigma_y^2 = \mu$$
$$\text{and} \quad \rho = 0.$$

Then for $R = X^*/Y^*$, Hinkley's formula becomes

$$g_1(r) = \frac{b(r)d(r)}{a^2(r)\sqrt{2\lambda\mu\pi}}\left[\Phi\left\{\frac{b(r)}{a(r)}\right\} - \Phi\left\{\frac{-b(r)}{a(r)}\right\}\right] + \frac{1}{a^2(r)\pi\sqrt{\lambda\mu}} exp\left\{\frac{-c}{2}\right\},$$

where

$$a^2(r) = \frac{r^2}{\lambda} + \frac{1}{\mu},$$
$$b(r) = r + 1,$$
$$c = \lambda + \mu,$$
$$d(r) = exp\left\{\frac{b^2(r) - ca^2(r)}{2a^2(r)}\right\}$$
$$\text{and } \Phi(\tau) = \int_{-\infty}^{\tau} \phi(u)du, \text{ where } \phi(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}.$$

Estimation

In considering the expected value of R, the calculations are clearly more difficult than that of the p. d. f. shown previously. The expected value of R is of the form

$$E(R) = \sum_{r\in \mathbf{Q}_0^+} \sum_{s\in A} \frac{r(\lambda^r \mu)^s e^{-(\lambda+\mu)}}{(rs)!s!(1-e^{-\mu})}$$

where r is summed over the non-negative rational numbers. Therefore, it becomes necessary to find an appropriate estimate for $E(R)$.

One method for determining an approximation would be to expand $R = X/Y$ in a Taylor series about the point (λ, γ). Upon doing so, one arrives at

$$R = \frac{\lambda}{\gamma} + \frac{x-\lambda}{\gamma} - \frac{\lambda(y-\gamma)}{\gamma^2} - \frac{(x-\lambda)(y-\gamma)}{\gamma^2} + \frac{\lambda(y-\gamma)^2}{\gamma^3} + \cdots$$

The expected value may be calculated as

$$\begin{aligned} E(R) &= \frac{\lambda}{\gamma} + \frac{E(x)-\lambda}{\gamma} - \frac{\lambda E(y)-\lambda\gamma}{\gamma^2} - \frac{E(x-\lambda)(y-\gamma)}{\gamma^2} \\ &\quad + \frac{\lambda E(y-\gamma)^2}{\gamma^3} + \cdots \\ &= \frac{\lambda}{\gamma} + 0 - 0 - 0 + \frac{\lambda\sigma_y^2}{\gamma^3} + \cdots \\ &= \frac{\lambda}{\gamma} + \frac{\lambda}{\gamma^2} + \cdots \end{aligned}$$

Unless μ is very small, the first order approximation is a good one to use; this is the approximation utilized in computing the p. d. f. previously.

Given estimates $\hat{\lambda}$ and $\hat{\gamma}$ of λ and γ, respectively, one may consider $\frac{\hat{\lambda}}{\hat{\gamma}}$ to be a point estimate for $E(R)$. Having a point estimate, the next step to be concerned with is that of an interval estimate. Interval estimation uses the point estimate or information derived from prior knowledge. In this case, knowing the p. d. f.'s of X and Y, the point estimate may lead one to a confidence interval for the mean of R.

In lieu of the actual p. d. f. of R, one considers the approximation which is of the form

$$g_1(r) = \frac{(r+1)exp\left\{\frac{(r+1)^2-(\lambda+\mu)(\frac{r^2}{\lambda}+\frac{1}{\mu})}{2(\frac{r^2}{\lambda}+\frac{1}{\mu})}\right\}}{(\frac{r^2}{\lambda}+\frac{1}{\mu})\sqrt{2\lambda\mu\pi}}\left[\Phi\left\{\frac{\lambda\mu(r+1)}{\mu r^2+\lambda}\right\} - \Phi\left\{\frac{-\lambda\mu(r+1)}{\mu r^2+\lambda}\right\}\right] + \frac{exp\left\{\frac{-(\lambda+\mu)}{2}\right\}}{(\frac{r^2}{\lambda}+\frac{1}{\mu})\pi\sqrt{\lambda\mu}}.$$

When either λ or μ is a large value, the final term is negligible and may be disregarded for approximation. Let $g_1(r) = G_1 + G_2$ where G_1 represents the leading terms and G_2 the last term. The following table (Table 3.1) expresses the idea that $g_1(r) \approx G_1$.

Table 3.1: Calculated Values of $g_1(r)$ Given λ and μ

λ	μ	$g_1(r)$	G_1	G_2
0.5	0.5	0.3688861	0.2723537	0.09653235
0.5	10.0	1.789106	1.781992	7.1143×10^{-3}
0.5	25.0	2.820953	2.820947	6.4044×10^{-6}
10.0	0.5	0.08911754	0.08909976	1.7786×10^{-5}
10.0	10.0	0.3989467	0.3989395	7.2256×10^{-6}
10.0	25.0	0.6307824	0.6307824	9.0269×10^{-9}
25.0	0.5	0.0564190	0.0564190	2.5618×10^{-9}
25.0	10.0	0.2523129	0.2523129	1.4443×10^{-9}
25.0	25.0	0.3989435	0.3989435	2.2103×10^{-12}

If the value of λ or μ are large or especially if both are large, G_2 may be neglected and hence $g_1(r) = G_1$. Thus the p. d. f. of R is approximately normal and the standard normal distribution may be called upon for interval estimation.

Concluding that $R \sim N(E(R), V(R))$, R is normally distributed with mean E(R) and variance V(R), and utilizing the approximations from the Taylor series expansion for $E(R)$ and $V(R)$, one may see that

$$R \sim N\left(\frac{\lambda}{\gamma}, \frac{\lambda}{\gamma^2} + \frac{\lambda^2}{\gamma^3}\right)$$

approximately. Thus an interval for $E(R)$ may be obtained as follows:

$$-z_{\frac{\alpha}{2}} \leq \frac{\frac{\hat{\lambda}}{\hat{\gamma}} - E(R)}{\sqrt{\frac{\hat{\lambda}(\hat{\gamma}+\hat{\lambda})}{\hat{\gamma}^3}}} \leq z_{\frac{\alpha}{2}}$$

implies

$$\frac{\hat{\lambda}}{\hat{\gamma}} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}(\hat{\gamma}+\hat{\lambda})}{\hat{\gamma}^3}} \leq E(R) \leq \frac{\hat{\lambda}}{\hat{\gamma}} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}(\hat{\gamma}+\hat{\lambda})}{\hat{\gamma}^3}}$$

which may be called a $z_{\frac{\alpha}{2}}$ standard deviation error bar for $E(R)$.

This idea of error bars stems from Duran and McCready [1991]. Two procedures of finding error bounds were used. One of these known as the one-sigma error bars, meaning $\mu_r \pm \sigma_r$, was found to have a smaller area and hence a smaller probability, namely $P = 0.68$, of containing the parameter in question. The other procedure found bounds on the parameter in a more statistical fashion. Since the distribution of the parameter was approximately normal, the upper and lower bounds could be determined more accurately. Even though the area of this interval was larger than that above, the probability of containing the parameter would be much greater than $P = 0.68$.

CHAPTER IV

CONCLUSIONS

Applications and Further Topics

The motivation for this research comes directly from Duran and McCready [1991]. In their report, two applications were considered. First, the measurement of the thickness of thin film coatings, also called thin foils, uses characteristic X-ray line ratio techniques. The characteristic X-rays are produced by the bombardment of a target with electrons. Two types of X-ray spectra are recorded, that of the substrate and of the overlay. After fitting Gaussian peaks to the X-ray lines contained in the spectra, the counts may be obtained for both the substrate X-ray line and the overlay X-ray line by considering the areas under the fitted Gaussian curve. Here, each of the counts for the X-ray lines has a Poisson distribution. The ratio of these X-ray lines is among the considerations by Duran and McCready.

The other application involved a similar process. The idea of irradiating a target with a neutral particle beam was considered to produce the X-ray spectrum. This spectrum was recorded and analyzed. In this case, the target could be determined to be either a reentry vehicle or a decoy upon application of this procedure. Since the reentry vehicle consists of an aeroshell of low atomic mass elements and a core of high atomic mass elements, high energy X-rays are produced when the beam reaches the core and do not diminish after leaving the target. On the other hand, a decoy containing a similar core would not have a massive aeroshell nor the weight. The X-ray's intensity would be greatly diminished upon leaving the target. The ratio of consideration here is that of one X-ray line to another each having a Poisson distribution, in determining a reentry vehicle or a decoy.

In both of these cases using the ratios of Poisson random variables, the means were large and normal approximations were utilized in the necessary testing.

Another work to consider is that of John Gart [1978] on "The Analysis of Ratios and Cross-Product Ratios of Poisson Variates with Application to Incidence Rates." The incidence rates of non-melanoma skin cancer was considered in two metropolitan areas over a six month period. Because the incidence of this disease is low for large populations, the number of new cases can be considered to have a Poisson distribution. Gart's study noted the variation due to age and sex and thus divided the data by area into age and sex stratifications. The ratios of consideration consist of the rates per 100,000 in six months according to age and sex of Area 1 divided by that of Area 2. Also, the ratios of these rates according to age and area of Male to Female were considered in a similar fashion. Another consideration was the cross-product ratios, the ratios of ratios (i.e., Area 1, ratio of Male to Female, divided by Area 2; or Male, ratio of Area 1 to Area 2, divided by Female). Significance tests were performed on each of these ratios and cross-product ratios using the normal distribution.

This research has raised many questions unanswerble at this time. One such question is "Exactly when is the ratio of the normal approximations an exceptable approximation for the ratio of the Poisson random variables?" This question stems from the Duran and McCready [1991] report and was considered during this research. However, the unobtainability of the actual distribution of the ratio R greatly hinders conclusions on the reliability of estimating this ratio by the ratio of the normal approximations to the Poisson random variables. Another possible consideration would be the ratio of two dependent Poisson random variables. In this work, the random variables were assumed to be independent; but there may be found a need for the ratio of two dependent Poisson random variables. The applications viewed here assume the use of normal approximations to the Poisson random variables. Each having very large means, the approximations would prove to be good. The assumption of independence was also noted in each example of the ratio of two Poisson random variables previously mentioned. This may imply that the need has not yet arisen for further research in these areas.

Closing Remarks

In considering the ratio of two Poisson random variables, one encounters many difficulties. The initial problem being the possibilty of the denominator taking the value zero. This minor set back has been resolved by a method of truncation, and the p. d. f. of the ratio obtained.

The intricacies of this p. d. f. create new troubles. One of these being the complexity of calculating the actual distribution of this ratio. As seen in Section 2.4 the range of values for R consists of all non-negative rational numbers, each having a positive probability of occurring. This leads one to the conclusion that the actual p. d. f. of R may not be calculable.

Turning to the employment of the normal approximations, one again finds slight difficulties. The p. d. f. of $R = X^*/Y^*$, where X^* and Y^* represent the normal approximations of the Poisson random variables X and Y, is not a precise normal distribution. Hence an approximation is again made. It is found that under conditions similar to those for the use of the normal approximations to the Poisson, R is approximately normally distributed.

The work provided during this research merely enhances that of previous mathematicians and statisticians. The normal approximations are widely utilized in dealing with the Poisson distribution. It has been concluded during the process of this investigation that the normal approximation may be applied when either the mean of the random variable in the numerator or that of the denominator, or both, is a large value. The question still remains,"What is large?" Greater depth of study is needed to quantify how "large" the means of the random variables should be for the normal approximation to be valid.

BIBLIOGRAPHY

Cohen, A.C., Jr. "Estimation of the Poisson Parameter From Truncated Samples and From Censored Samples." 49 Journal of the American Statistical Association, 1954, pp. 158-168.

David, F.N. and N.L. Johnson. "The Truncated Poisson." 8 Biometrics, 1952, pp. 275-285.

Doss, S.A.D.C. "On the Efficiency of Best Asymptotically Normal Estimates of the Poisson Parameter Based on Singly and Doubly Truncated or Censored Samples." 19 Biometrics, 1963, pp. 588-593.

Duran, Benjamin S. and Steven S. McCready. Ratio of Two Gaussian Random Variables with Applications. PL-TR-91-1008, Phillips Laboratory, Kirtland Air Force Base, NM, July 1991.

Fieller, E.C. "The Distribution of the Index in a Normal Bivariate Population." 24 Biometrika, 1932, pp. 428-440.

Gart, John J. " The Analysis of Ratios and Cross-Product Ratios of Poisson Variates with Application to Incidence Rates." A7(10) Commun. Statist. - Theor. Meth., 1978, pp. 917-937.

Geary, R.C. "The Frequency Distribution of the Quotient of Two Normal Variates." 93 Journal of the Royal Statistical Society, 1930, p. 442-446.

Haight, Frank A. Handbook of the Poisson Distribution. New York: Wiley, 1967.

Hald, Anders. A History of Probability and Statistics and their Applications Before 1750. New York: Wiley, c1990.

Hinkley, D.V. "On the Ratio of Two Correlated Normal Random Variables." 56 Biometrika, 1969, pp. 635-639.

Irwin, J.O. "On the Estimation of the Mean of a Poisson Distribution from a Sample with the Zero Class Missing." 15 Biometrics, 1959, pp. 324-326.

Maistrov, Leonid Efimovich. Probability Theory; A Historical Sketch. Translated and edited by Samuel Kotz. New York: Academic Press, 1974.

Marsaglia, George. "Ratios of Normal Variables and Ratios of Sums of Uniform Variables." 60 Journal of the American Statistical Association, 1965, pp. 193-204.

Moore, P.G. "A Note on Truncated Poisson Distributions." 10 Biometrics, 1954, pp. 402-406.

Plackett, R.L. "The Truncated Poisson Distribution." 9 Biometrics, 1953, pp. 485-488.

Stigler, Stephen M. The History of Statistics: The Measurement of Uncertainty Before 1900. Mass.: The Belknap Press of Harvard University Press, 1986.