CONFIDENCE INTERVALS FOR DISCRETE DISTRIBUTIONS

BU-839-M May 1984

by

George Casella and Charles E. McCulloch*
Biometrics Unit, Cornell University, Ithaca, NY 14853

*George Casella is Associate Professor of Biological Statistics and Charles E. McCulloch is Assistant Professor of Biological Statistics, both in the Biometrics Unit, Cornell University, Ithaca, NY 14853. Dr. Casella's Research was supported by National Science Foundation Grant No. MCS81-02541. This is technical report #BU-839-M in the Biometrics Unit series.

Abstract

Confidence intervals for the parameters of different discrete distributions can be derived through a similar underlying method, a method with roots in fiducial inference. This method is seen to be an application of the fundamental theorem of calculus, and thus is not tied to the theory of fiducial inference.

Some of the intervals derived through this method are well known (binomial, Poisson); however, the results for the negative binomial distribution are new. An advantage of the intervals is that they are given explicitly in terms of cut-off points of continuous distributions. Because of this, we are able to show that the intervals for the Poisson and negative binomial distributions are sharp (the infimum of the coverage probabilities is the nominal level). Optimality considerations, in terms of the size of the intervals, are also examined.

KEY WORDS AND PHRASES: Binomial; Poisson; Negative binomial; Coverage probability; Length optimality.

1. INTRODUCTION

Explicit formulas exist for nonrandomized confidence intervals for the parameters of the most commonly used discrete distributions (binomial, Poisson and negative binomial). These formulas require only standard (F and χ^2) statistical tables. Availability of the explicit formulas makes it easier to derive properties of the intervals.

The intervals presented here are not all new. Indeed, the binomial interval is that of Clopper and Pearson (1934), while the Poisson interval was first derived by Garwood (1936). The negative binomial interval, as far as we can tell, has not been derived by other authors.

The intervals can all be derived through a common methodology, which has its roots in fiducial inference. (Indeed, Clopper-Pearson and Garwood were deriving fiducial intervals.) Part of the purpose of this paper is to illustrate the common, underlying technique used in deriving the intervals, and to show that the technique is not dependent on fiducial theory, and may be applicable in other situations.

We will call a confidence interval sharp if the infimum (over the parameter space) of the coverage probability achieves the nominal level. We will call an interval conservative if the infimum of the coverage probability is strictly greater than the nominal level. The derivation used here is a conservative one, i.e., the intervals are guaranteed to have coverage probability at least $1-\alpha$. The surprising fact, however, is that for both the Poisson and negative binomial distributions, the intervals are in fact sharp.

In Section 2 we present the derivation of the intervals, and in Section 3 we prove sharpness for the Poisson and negative binomial intervals. Some aspects of length optimality are discussed in Section 4.

The case of the binomial distribution is somewhat different from that of the Poisson and negative binomial: in general, the binomial intervals are not sharp. These intervals are not discussed in detail here, the interested reader is referred to McCulloch and Casella (1983).

2. DERIVATION OF THE INTERVALS

A usual way (Lehmann, 1959) of deriving confidence intervals is to invert the acceptance region of an α-level test of the hypotheses

$$H_0: \theta=\theta_0 \qquad \text{versus} \qquad H_A: \theta\neq\theta_0 \quad .$$

A similar approach, called the "statistical method" by Mood, Graybill and Boes (1974) is to solve in each of the statements

$$P_\theta(X\geq x) \geq \alpha/2$$

and (2.1)

$$P_\theta(X\leq x) \geq \alpha/2$$

for regions in the parameter space that make the statements true. Each of the statements in (2.1) will give rise to sharp one-sided confidence intervals. If each of them is used to derive a 1-(α/2) one-sided interval, their intersection can be used as a conservative 1-α two-sided confidence interval. In general, these intervals are not sharp. In fact, when using this method to derive intervals for a binomial success probability (for any n), the actual coverage probability of the intervals will be as high as 1-(α/2) for a nominal 1-α interval (McCulloch and Casella, 1983).

The advantage of using the statistical method is that we can exploit the following identity for discrete, integer-valued distributions. First write

$$P_\theta(X\geq x) = \sum_{k=x}^{\infty} P_\theta(X=x) = \int_{\theta_{min}}^{\theta} \frac{\partial}{\partial t} \sum_{k=x}^{\infty} P_t(X=x)dt \quad , \tag{2.2}$$

where θ_{min} = smallest possible value of θ. After differentiation and integration with respect to the parameter, it is sometimes possible to simplify the expression inside the integral. Mechanistically, this is the same as the fiducial approach, and indeed, whenever the term inside the integral actually is a density, it is a fiducial density. In such cases, the integral is the fiducial cumulative distribution function of a continuous distribution and the solution to (2.1) can be written explicitly in terms of percentage points of that distribution. For the binomial, Poisson and negative binomial distributions, these are percentage points of well tabulated distributions (F and χ^2).

For example, for $X_1,\cdots,X_n$ iid Poisson (θ), let $Y = \sum_1^n X_i \sim$ Poisson ($n\theta$), and write

$$P_\theta(Y\geq y) = \sum_{k=y}^{\infty} e^{-n\theta} \frac{(n\theta)^k}{k!} = \sum_{k=y}^{\infty} \int_0^\theta \frac{\partial}{\partial t}\left[e^{-nt} \frac{(nt)^k}{k!}\right] dt \quad . \tag{2.3}$$

After differentiating and interchanging the order of summation and integration, one is left with a telescoping sum, thus,

$$\begin{aligned} P_\theta(Y\geq y) &= \int_0^\theta \left\{\sum_{k=y}^{\infty} \frac{e^{-nt}}{k!} \left(kn(nt)^{k-1} - n(nt)^k\right)\right\} dt \\ &= \int_0^\theta \frac{n(nt)^{y-1}}{(y-1)!} e^{-nt} dt \quad . \end{aligned} \tag{2.4}$$

Making the transformation $\omega=2nt$ allows us to write

$$P_\theta(Y\geq y) = P(\chi^2_{2y} \leq 2n\theta) \quad , \tag{2.5}$$

which is a well-known relationship between the Poisson and chi-squared distributions.

The above argument also shows that the fiducial argument is not necessary, and the method works even when

$$\frac{\partial}{\partial t} \sum_{k=x}^{\infty} P_t(X=k)\Big|_{t=\theta} \tag{2.6}$$

or its absolute value is not a density in θ. In fact, it is advantageous at times only to differentiate and integrate a portion of $P_\theta(X=x)$.

The fiducial approach also runs into slight difficulty since two different fiducial distributions arise when working with discrete distributions. Stevens (1950) calls these a pair of fiducial distributions. Again, these technicalities do not create problems since the derived intervals are still $1-\alpha$ confidence intervals.

Using (2.2) the following $1-\alpha$ confidence intervals can be derived (see McCulloch and Casella, 1983, for details). For the parameter θ in a binomial distribution, i.e., $P_\theta(X=x) = \binom{n}{x} \theta^x(1-\theta)^{n-x}$ the interval is

$$\left[\frac{1}{1+\frac{n-X+1}{X} F^{2(n-X+1)}_{2X,\alpha/2}}, \frac{\frac{X+1}{n-X} F^{2(X+1)}_{2(n-X),\alpha/2}}{1+\frac{X+1}{n-X} F^{2(X+1)}_{2(n-X),\alpha/2}}\right], \tag{2.7}$$

where, in general, $F^{\nu_1}_{\nu_2,\gamma}$ satisfies $P\left(F^{\nu_1}_{\nu_2} > F^{\nu_1}_{\nu_2,\gamma}\right) = \gamma$. For the parameter θ in a Poisson distribution, i.e., $P_\theta(X=x) = \frac{\theta^x e^{-\theta}}{x!}$, the interval is

$$\left(\frac{1}{2n} \chi^2_{2Y,1-\alpha/2}, \frac{1}{2n} \chi^2_{2(Y+1),\alpha/2}\right), \tag{2.8}$$

where $Y = \sum X_i$ and $\chi^2_{\nu,\alpha}$ satisfies $P(\chi^2_\nu > \chi^2_{\nu,\alpha}) = \alpha$. For the parameter θ in a negative binomial distribution (r assumed known), i.e., $P_\theta(X=x) = \binom{r+x-1}{x} \theta^r(1-\theta)^x$ the interval is

$$\left[\frac{1}{1+\frac{X+1}{r} F^{2(X+1)}_{2r,\alpha/2}}, \frac{\frac{r}{X} F^{2r}_{2X,\alpha/2}}{1+\frac{r}{X} F^{2r}_{2X,\alpha/2}}\right]. \tag{2.9}$$

3. SHARPNESS OF THE POISSON AND NEGATIVE BINOMIAL INTERVALS

In this section we address the question of whether the coverage probabilities of the Poisson and negative binomial intervals ever achieve the infimum, $1-\alpha$. We show, somewhat surprisingly, that even though these intervals were constructed in a 'conservative' manner, they achieve the stated infimum, i.e., they are sharp.

For notational simplicity, and without loss of generality, we assume in this section that $n=1$.

Theorem 3.1: Let $X \sim \text{Poisson}(\theta)$. For fixed α, the confidence intervals

$$\left(\tfrac{1}{2}\chi^2_{2X,1-(\alpha/2)}\ ,\ \tfrac{1}{2}\chi^2_{2(X+1),\alpha/2}\right) \tag{3.1}$$

are sharp. That is,

$$\inf_{\theta} P_\theta\left(\tfrac{1}{2}\chi^2_{2X,1-(\alpha/2)} \le \theta \le \tfrac{1}{2}\chi^2_{2(X+1),\alpha/2}\right) = 1-\alpha \quad . \tag{3.2}$$

Proof: We will show that, as $\theta \to \infty$, the limit of the coverage probabilities is $1-\alpha$. We use the notation χ^2_{2X} to denote a chi-squared random variable with $2X$ degrees of freedom, where $X \sim \text{Poisson}(\theta)$. [More precisely, the distribution of χ^2_{2X} given $X=x$ is χ^2_{2x}, while X is marginally Poisson (θ).]

It is straightforward to check that $E[\chi^2_{2X}] = 2\theta$ and $\text{Var}(\chi^2_{2X}) = 8\theta$, and

$$\frac{\chi^2_{2X} - 2\theta}{(8\theta)^{\frac{1}{2}}} \to Z \text{ in distribution as } \theta \to \infty \tag{3.3}$$

where Z is a standard normal random variable. The convergence in (3.3) can be established by using the moment generating function of χ^2_{2X}, $\exp\{-\theta + [\theta/(1-2t)]\}$.

Next

$$P_\theta\left(\tfrac{1}{2}\chi^2_{2X,1-(\alpha/2)} \le \theta \le \tfrac{1}{2}\chi^2_{2(X+1),\alpha/2}\right)$$

$$= P_\theta\left(\chi^2_{2X,1-(\alpha/2)} \le \left(\frac{2\theta}{\chi^2_{2X}}\right) \chi^2_{2X} \le \chi^2_{2(X+1),\alpha/2}\right)$$

$$= P_\theta\left(\frac{\chi^2_{2X,1-(\alpha/2)}-2\theta}{(8\theta)^{\frac{1}{2}}} \le \frac{\left(\frac{2\theta}{\chi^2_{2X}}\right)\chi^2_{2X}-2\theta}{(8\theta)^{\frac{1}{2}}} \le \frac{\chi^2_{2(X+1),\alpha/2}-2\theta}{(8\theta)^{\frac{1}{2}}}\right). \tag{3.4}$$

An application of (3.3) shows that

$$\frac{\chi^2_{2X,1-(\alpha/2)}-2\theta}{(8\theta)^{\frac{1}{2}}} \to z_{1-(\alpha/2)} \quad , \tag{3.5}$$

and (3.3), together with the fact that $\chi^2_{2X}/2\theta \to 1$ in probability as $\theta \to \infty$, shows that the middle term in (3.4) converges to a standard normal variate. Similarly, it can be shown that the rightmost term in (3.4) converges to $z_{\alpha/2}$. Hence,

$$\lim_{\theta\to\infty} P_\theta\left(\tfrac{1}{2}\chi^2_{2X,1-(\alpha/2)} \le \theta \le \tfrac{1}{2}\chi^2_{2(X+1),\alpha/2}\right) = P\left(z_{1-(\alpha/2)} \le Z \le z_{\alpha/2}\right)$$
$$= 1-\alpha \quad , \tag{3.6}$$

establishing the theorem. ||

A similar result holds for the negative binomial intervals. The proof is similar to that of Theorem 3.1, and hence will only be sketched.

<u>Theorem 3.2</u>: Let $X \sim NB(r,\theta)$. For fixed α, the confidence intervals given in (2.9) are sharp, i.e.,

$$\inf_\theta P_\theta\left(\frac{1}{1+\frac{(X+1)}{r}F^{2(X+1)}_{2r,\alpha/2}} \le \theta \le \frac{\frac{r}{X}F^{2r}_{2X,\alpha/2}}{1+\frac{r}{X}F^{2r}_{2X,\alpha/2}}\right) = 1-\alpha \quad . \tag{3.7}$$

Proof: We use the facts that, as $\theta \to 0$,

i) $2\theta X \to \chi^2_{2r}$ in distribution

ii) $\dfrac{F^{2(X+1)}_{2r}}{2r} \to \dfrac{1}{\chi^2_{2r}}$ in distribution

iii) $\dfrac{F^{2(X+1)}_{2r,\alpha/2}}{2r} \to \dfrac{1}{\chi^2_{2r,1-(\alpha/2)}}$ in probability .

From (3.7), we can write the coverage probability as

$$P_\theta\left[\theta + (2\theta X)\left(\frac{F^{2X}_{2r,1-(\alpha/2)}}{2r}\right) \le 1 < \theta + (2\theta X)\left(\frac{F^{2(X+1)}_{2r,\alpha/2}}{2r}\right) + \frac{\theta}{r}\, F^{2(X+1)}_{2r,\alpha/2}\right] \quad . \quad (3.8)$$

As $\theta \to 0$, using i) – iii), we see that (3.8) converges to

$$P\left[\chi^2_{2r}\left(\frac{1}{\chi^2_{2r,\alpha/2}}\right) \le 1 \le \chi^2_{2r}\left(\frac{1}{\chi^2_{2r,1-(\alpha/2)}}\right)\right] \quad (3.9)$$

$$= P\left(\chi^2_{2r,1-(\alpha/2)} \le \chi^2_{2r} \le \chi^2_{2r,\alpha/2}\right)$$

$$= 1-\alpha \quad . \quad \|$$

Here we see the theoretical advantage of having the explicit formulas for the discrete intervals. The proofs of sharpness rely heavily on their use. We also note that the same proofs will work even if the tail probabilities are not equally split, i.e., putting γ in the upper tail and $1-\alpha+\gamma$ in the lower tail ($0 \le \gamma \le \alpha$) will still result in a sharp $1-\alpha$ procedure.

4. OPTIMALITY CONSIDERATIONS

In the previous section it was shown that the Poisson and negative binomial intervals were sharp at level $1-\alpha$ for all possible combinations of tail probabilities that add to α. It is natural to next investigate the size of the intervals as a function of the "α-split."

Having the explicit formulas for the confidence intervals makes such an investigation somewhat easier. It is simple to calculate the size of an interval for a given α-split and compare it with a second α-split. Unfortunately, calculations of expected size are still quite unwieldy. Even if these were undertaken, it would not be clear as to what optimality criteria should be investigated. Criteria usually used in the continuous case (unbiasedness, probability of false coverage) are really not applicable in discrete problems.

To get some idea of how small the intervals can be made, the following "procedure" was considered: For each x, select the α-split dependent on x in order to minimize the size of the confidence interval. This strategy does not give a $1-\alpha$ confidence procedure, but a procedure with confidence less than $1-\alpha$. Still, it serves as a measure of absolute size of the confidence intervals, and comparison of a fixed α-split procedure to this "optimal procedure" will give an indication of the performance of the fixed α-split procedure.

4.1. Poisson Intervals

A natural measure of size of the Poisson intervals is the ratio of the upper to the lower endpoint, i.e., for fixed γ, $0 \le \gamma \le \alpha$, define

$$S_\gamma(x) = \chi^2_{2(x+1),\gamma}/\chi^2_{2x,1-\alpha+\gamma} \quad . \tag{4.1}$$

For each x, we can calculate an optimal value of γ by minimizing $S_\gamma(x)$. This is a straightforward minimization, and the minimizing value, say $\gamma^*(x)$, is easily found numerically.

In Figure 1, for $\alpha=.1$ and $\alpha=.05$, we plot the ratio $S_{\alpha/2}(x)/S_{\gamma^*(x)}(x)$ for $x=0,\cdots,30$. (For $x=0$ we define $\chi^2_{2x}=0$ and $S_{\alpha/2}(\theta)/S_{\gamma^*}(0) = \chi^2_{2(x+1),\alpha/2}/\chi^2_{2(x+1),\gamma}$). The results, surprisingly, show that for most values of x, $S_{\alpha/2}(x)$ is nearly as good as $S_{\gamma^*}(x)$. (Recall that $S_{\gamma^*}(x)$ does not correspond to a $1-\alpha$ confidence procedure.) As $x \to \infty$, normal theory takes over so we know that $\alpha/2$ is optimal for large x. It is surprising, however, that for x as small as 2 the equal-tail interval is nearly optimal.

4.2. Negative Binomial Intervals

A similar analysis was carried out for the negative binomial intervals, but here we use length as our criterion. For each r, define,

$$S_\gamma(x) = \frac{r}{r + xF^{2x}_{2r,1-\alpha+\gamma}} - \frac{r}{r + (x+1)F^{2(x+1)}_{2r,\gamma}} \quad . \tag{4.2}$$

We similarly define $\gamma^*(x)$ as the value that minimizes (4.2) for each x, and again investigate the ratio $S_{\alpha/2}(x)/S_{\gamma^*}(x)$. For $\alpha=.05$ and .1 these are shown in Figures 2 and 3 for $r=2,5,10,20$ and $x=0,\cdots,35$.

Examination of these plots shows that, although the equal split performs well, the performance is not nearly as good as in the Poisson case. In particular, as $x \to \infty$, the equal split is not optimal as the ratio $S_{\alpha/2}(x)/S_{\gamma^*(x)}(x)$ approaches a limit $\neq 1$. To evaluate this limit, consider

$$\lim_{x\to\infty} \frac{S_{\alpha/2}(x)}{S_\gamma(x)} = \lim_{x\to\infty} \frac{\dfrac{1}{1 + \frac{x}{r} F^{2x}_{2r,1-(\alpha/2)}} - \dfrac{1}{1 + \frac{(x+1)}{r} F^{2(x+1)}_{2r,\alpha/2}}}{\dfrac{1}{1 + \frac{x}{r} F^{2x}_{2r,1-\alpha+\gamma}} - \dfrac{1}{1 + \frac{(x+1)}{r} F^{2(x+1)}_{2r,\gamma}}} \quad . \tag{4.3}$$

Dividing top and bottom by x, and recalling that $\frac{1}{r} F^{2x}_{2r,\beta} \to (\chi^2_{2r,1-\beta})^{-1}$ as $x \to \infty$, we have

$$\lim_{x\to\infty} \frac{S_{\alpha/2}(x)}{S_\gamma(x)} = \frac{\chi^2_{2r,\alpha/2} - \chi^2_{2r,1-(\alpha/2)}}{\chi^2_{2r,\alpha-\gamma} - \chi^2_{2r,1-\gamma}} \quad . \tag{4.4}$$

We can now find $\gamma^*(\infty)$, the value which minimizes $\chi^2_{2r,\alpha-\gamma} - \chi^2_{2r,1-\gamma}$ subject to the probability constraint. The values $S_{\alpha/2}(\infty)/S_{\gamma^*(\infty)}(\infty)$ are the asymptotes in Figures 2 and 3.

We next investigated whether $\gamma^*(\infty)$ might provide a better means for splitting the probability between the two tails. (Note that $\gamma^*(\infty)$ is independent of x, so it does provide a $1-\alpha$ confidence procedure.) A plot of $S_{\gamma^*(\infty)}(x)/S_{\gamma^*(x)}(x)$ produces pictures very much like that in Figure 1 (the Poisson case) and shows that $\gamma^*(\infty)$ is indeed a reasonably good choice for the probability split. Comparing $\gamma^*(\infty)$ to $\alpha/2$ (Figure 4) shows that for all but the smallest values of x, $\gamma^*(\infty)$ gives shorter intervals than $\alpha/2$, with the improvement diminishing as r increases.

Overall, it appears that $\gamma^*(\infty)$ is the better choice for negative binomial confidence intervals. Of course, by using $\gamma^*(\infty)$ rather than $\alpha/2$, it is not longer possible to use standard F-tables to construct the confidence intervals. However, with the use of a computer or even a programmable calculator, the intervals based on $\gamma^*(\infty)$ are not difficult to construct. Table 1 gives values of $\gamma^*(\infty)$ for $\alpha=.05$ and .1, $r=1,\cdots,50$.

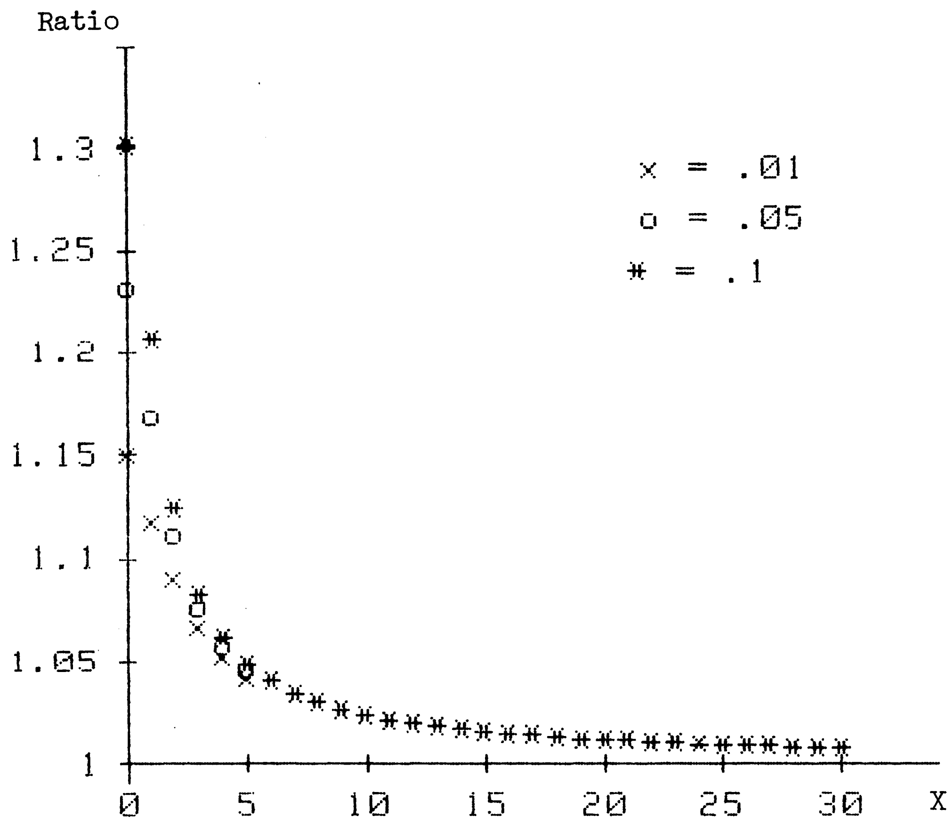

Figure 1. Ratio of the Size of the Poisson Interval Using Equal Tail Probabilities to the Size of the "Optimal" Intervals.

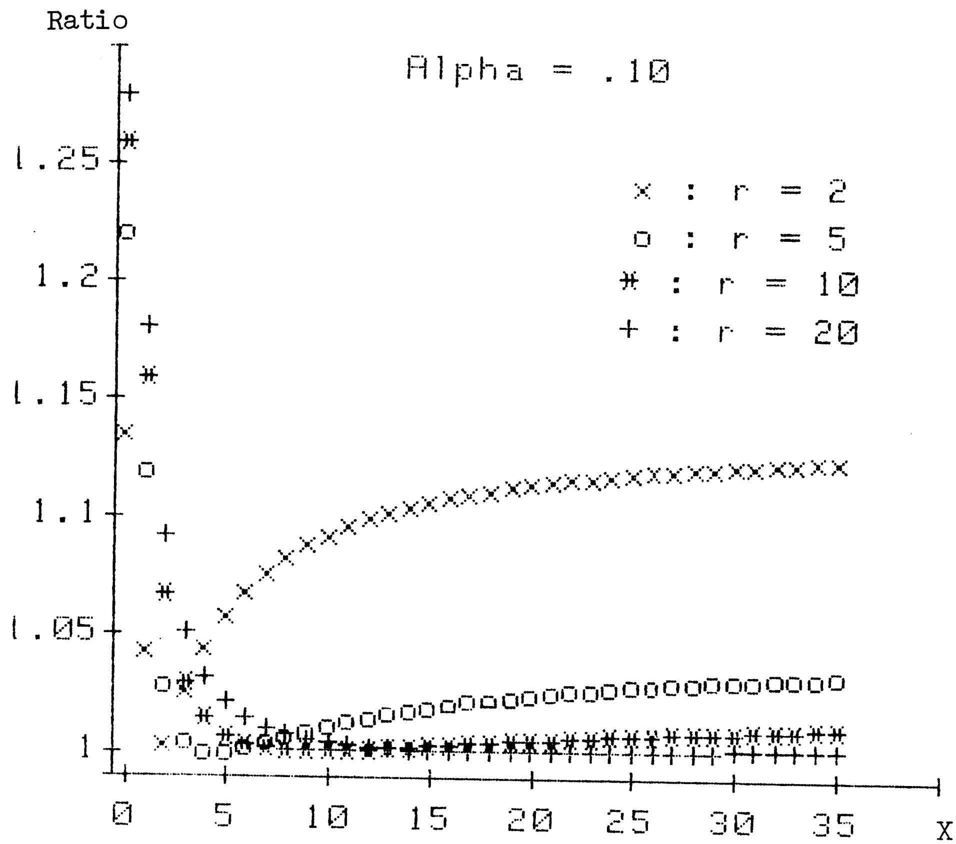

Figure 2. Ratio of the Length of the Negative Binomial Intervals Based on Equal Tail Probabilities to the Length of the "Optimal" Intervals.

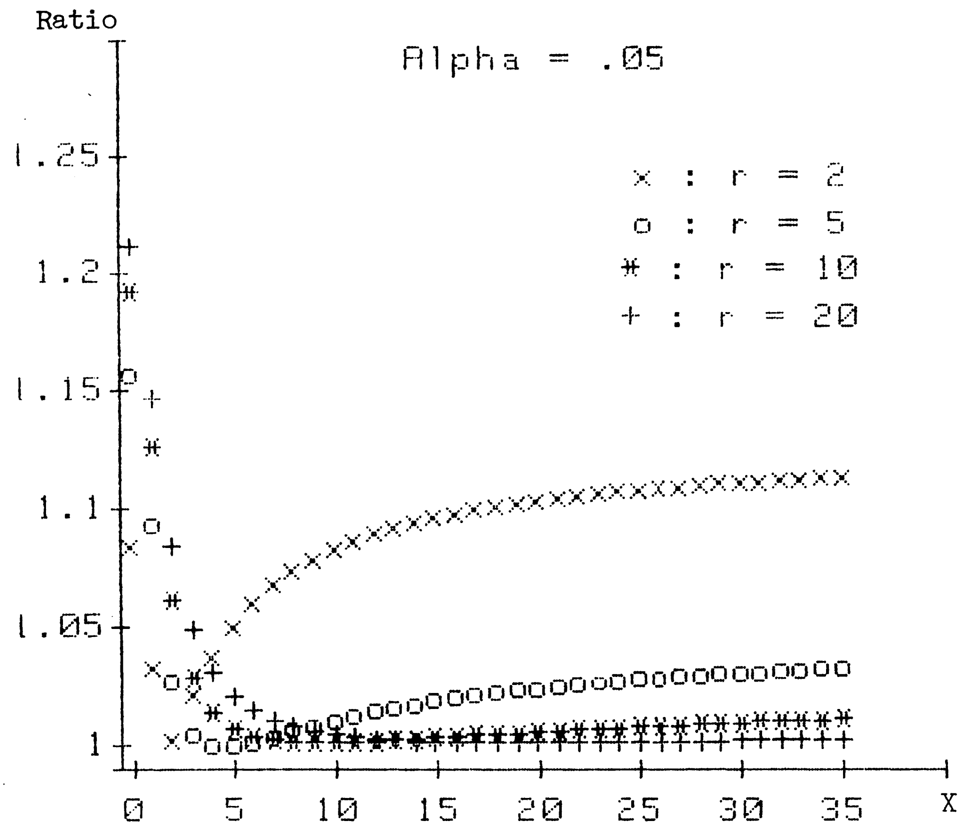

Figure 3. Ratio of the Length of the Negative Binomial Intervals Based on Equal Tail Probabilities to the Length of the "Optimal" Intervals.

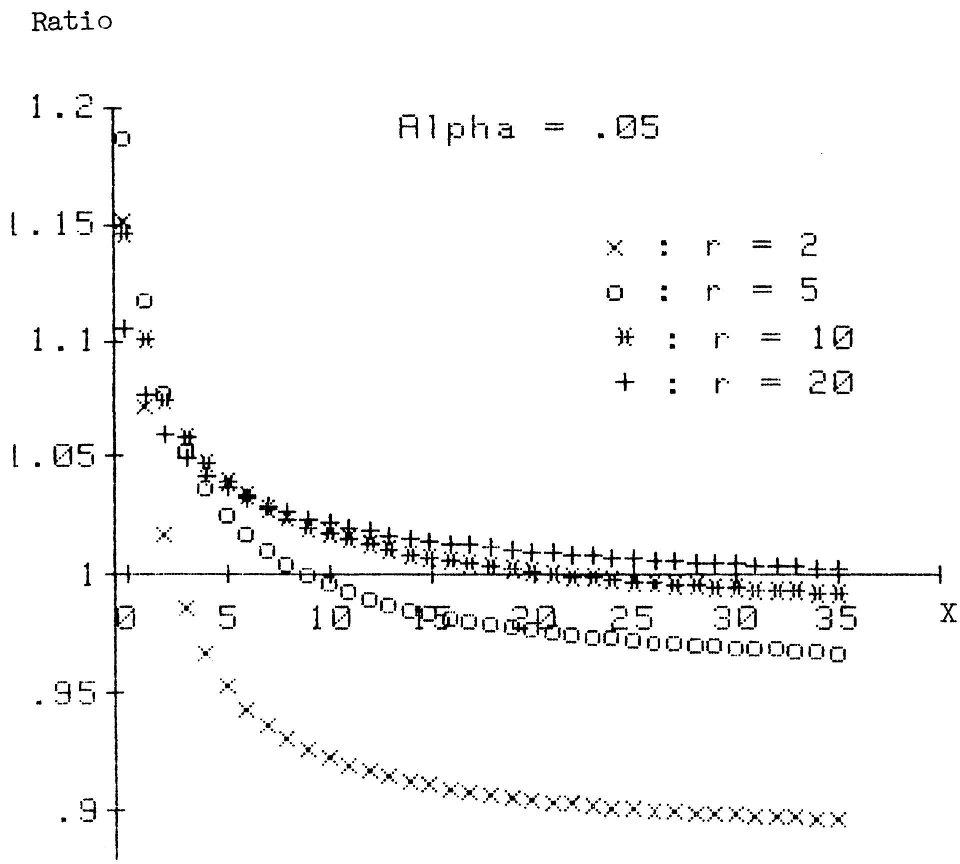

Figure 4. Ratio of the Length of the Negative Binomial Intervals Based on $\gamma^*(\infty)$ to Those with Equal Tail Probabilities.

Table 1. Tail Probabilities to be Used to Get F Cutoff Points

Based on Infinite Optimal Split

r	Alpha	Upper Tail	Alpha	Upper Tail
1	.050	.000	.100	.000
2	.050	.001	.100	.003
3	.050	.004	.100	.010
4	.050	.006	.100	.015
5	.050	.008	.100	.019
6	.050	.009	.100	.022
7	.050	.010	.100	.024
8	.050	.011	.100	.025
9	.050	.012	.100	.027
10	.050	.013	.100	.028
11	.050	.013	.100	.029
12	.050	.014	.100	.030
13	.050	.014	.100	.031
14	.050	.015	.100	.032
15	.050	.015	.100	.032
16	.050	.015	.100	.033
17	.050	.016	.100	.033
18	.050	.016	.100	.034
19	.050	.016	.100	.034
20	.050	.016	.100	.035
21	.050	.017	.100	.035
22	.050	.017	.100	.035
23	.050	.017	.100	.036
24	.050	.017	.100	.036
25	.050	.017	.100	.036
26	.050	.017	.100	.036
27	.050	.018	.100	.037
28	.050	.018	.100	.037
29	.050	.018	.100	.037
30	.050	.018	.100	.037
31	.050	.018	.100	.038
32	.050	.018	.100	.038
33	.050	.018	.100	.038
34	.050.	.018	.100	.038
35	.050	.018	.100	.038
36	.050	.019	.100	.039
37	.050	.019	.100	.039
38	.050	.019	.100	.039
39	.050	.019	.100	.039
40	.050	.019	.100	.039
41	.050	.019	.100	.039
42	.050	.019	.100	.039
43	.050	.019	.100	.040
44	.050	.019	.100	.040
45	.050	.019	.100	.040
46	.050	.019	.100	.040
47	.050	.019	.100	.040
48	.050	.019	.100	.040
49	.050	.019	.100	.040
50	.050	.019	.100	.040

REFERENCES

Clopper, C.J., and Pearson, E.S. (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404-414.

Garwood, F. (1936), "Fiducial Limits for the Poisson Distribution," *Biometrika*, 28, 437-441.

Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. John Wiley and Sons, New York.

McCulloch, C.E., and Casella, G. (1983), "Explicit Formulas for Confidence Interval Estimation in Discrete Distributions, BU-820-M in the Biometrics Unit Series, Cornell University.

Mood, A.M., Graybill, F.A., and Boes, D.C. (1974), *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

Stevens, W.L. (1950), "Fiducial Limits of the Parameter of a Discontinuous Distribution," *Biometrika*, 37, 117-129.