Improved Exact Confidence Intervals for Discrete Distributions

Helge Blaker
University of Oslo

November 1998

Abstract

An explicit method is given for improving standard "exact" confidence intervals in discrete distributions. The improved intervals have smaller size but correct coverage probability. It is argued that one should consider confidence sets for all possible confidence levels as generated by a confidence curve. Improved confidence curves then lead to improved confidence sets. A general method for constructing confidence curves is given. For the special case of a one-parameter exponential family, the resulting confidence curve is Spjøtvoll's acceptability function. This confidence curve yields improved intervals for all discrete distributions; details are provided for the binomial, Poisson, negative binomial, and hypergeometric distribution. Nonparametric confidence intervals for a quantile are also considered.

Key words: Acceptability; Confidence curve, Binomial distribution; Poisson distribution; Negative binomial distribution; Hypergeometric distribution.

1 Introduction

Construction of confidence intervals in discrete distributions is an old problem where no definite solution seems to have been reached. The usual approach is based on inverting an equal-tailed test, giving standard "exact" intervals in distributions such as binomial, Poisson, negative binomial, or hypergeometric. Following Blyth and Still (1983), an exact confidence set has coverage probability larger than or equal to the nominal level for each possible parameter value. Standard exact confidence intervals tend to be very conservative, in particular for small and moderate sample sizes, and hence too wide. There are two different approaches for obtaining improved confidence intervals. The first is to use approximate solutions based on the normal distribution, see e.g. Vollset (1993) or Agresti and Coull (1998) for the binomial case. Such procedures typically yield shorter intervals, but the coverage probability is not above the nominal level for all parameter values and hence they are not correct confidence sets in the strict sense. The second approach is to look for less conservative exact intervals. For the binomial distribution, this is done by Sterne (1954), Crow (1956), and Blyth and Still (1983) and for the Poisson distribution by Crow and Gardner (1959). Casella and Robert (1989) address the question of improving confidence procedures with respect to length while maintaining correct confidence level through a numerical procedure they call a refinement process. The purpose of this article is to present an analytical approach which gives improved exact confidence intervals compared to the standard equal-tailed intervals for discrete distributions. Our approach builds on Spjøtvoll's (1983) acceptability function and is related to a procedure for improving binomial confidence intervals presented in Sterne (1954). It is no more difficult to use than the standard method and does not involve detailed analysis of acceptance regions for every possible parameter value, like the methods in Blyth and Still (1983). In fact, the improvement emanates from treating a confidence interval (or two-sided test) as a problem in its own right and not just as the intersection of two one-sided problems. Numerical calculations of intervals and coverage probabilities for the binomial, Poisson, negative binomial, and hypergeometric distribution show that our approach yields confidence intervals with considerably smaller size and coverage probability much closer to the nominal level. These distributions all have monotone likelihood ratio. Improved confidence sets can be constructed for all discrete distributions, but outside the class of distributions with monotone likelihood ratio, uniformly most powerful one-sided tests do not exist and hence the choice of statistic is more difficult. Also, it is no longer clear that we necessarily want a confidence interval rather than a more complicated set.

In the next section, we formulate the problem in terms of confidence curves and present a general method for constructing exact confidence sets. Section 3 introduces the acceptability function and shows why it leads to improved intervals compared to the standard method. The rest of the paper then studies the improved intervals for the binomial, Poisson, negative binomial, and hypergeometric distributions. We also look at nonparametric confidence intervals for a quantile. The appendix contains S-plus functions for computing the acceptability functions for the cases discussed so the reader can compute his own tables.

2 Preliminaries

Let X have a discrete distribution indexed by a real-valued parameter θ and let $p_\theta(x)$ be the density with respect to counting measure on the natural numbers. Assume a test of $H : \theta = \theta_0$ against $K : \theta = \theta_1$ is based on the statistic $T(X)$, such that H is rejected when T is large if $\theta_1 > \theta_0$ and when T is small if $\theta_1 < \theta_0$. Denote an observed value of T by t. In particular, this holds if the family of distributions has monotone likelihood ratio in $T(x)$, i.e. the ratio of densities $p_{\theta'}(x)/p_\theta(x)$ is a nondecreasing function of $T(x)$ for all $\theta < \theta'$. This includes the binomial, Poisson, negative binomial, and hypergeometric distributions, see Lehmann (1986) chapter 3. We are interested in constructing confidence sets at level $1-\alpha$ for θ, i.e. sets $C_\alpha(t)$ such that $\inf_\theta P_\theta(\theta \in C_\alpha(T)) \geq 1-\alpha$. If T is *stochastically increasing*, i.e. $P_\theta(T \leq t)$ is decreasing in θ for all t, it is natural to restrict attention to intervals. Families with monotone likelihood ratio are always stochastically increasing. Standard confidence intervals for θ are found by inverting the equal-tailed test of $H : \theta = \theta_0$, so $C_\alpha(t) = (\theta_L, \theta_U)$ where θ_L is the largest θ such that $P_\theta(T \geq t) \leq \alpha/2$ and θ_U is the smallest θ such that $P_\theta(T \leq t) \leq \alpha/2$, except for boundary cases. This interval will be exact, i.e. the coverage probability is at least $1-\alpha$ for every possible θ. However, there must be some room for improvement since we have replaced the condition

$$P_\theta(\theta_L(T) < \theta < \theta_U(T)) \geq 1-\alpha \tag{1}$$

by the stronger condition

$$P_\theta(\theta_L(T) \geq \theta) \leq \alpha/2 \text{ and } P_\theta(\theta_U(T) \leq \theta) \leq \alpha/2. \quad (2)$$

The p-value of the equal-tailed test of $H : \theta = \theta_0$ is $\beta(\theta; t) = \min\{\beta^0(\theta; t), 1\}$ where

$$\beta^0(\theta_0; t) = 2 \min\{P_{\theta_0}(T \geq t), P_{\theta_0}(T \leq t)\}.$$

Notice that $(\theta_L, \theta_U) = \{\theta : \beta(\theta; t) > \alpha\} = U_\alpha(t)$, say. The function $\beta(\theta; t)$ is the *confidence curve* corresponding to this confidence procedure, as advocated in Birnbaum (1961) and $\beta(\theta; t)$ as a function of θ for a fixed (observed) t gives all possible confidence intervals at all levels and ranks all possible θ-values according to how reasonable they are after $T = t$ is observed. It is a *preference function* in the sense of Spjøtvoll (1983) and its interpretation is akin to the likelihood function. In other words, θ_1 is better than θ_2 if $\beta(\theta_1; t) > \beta(\theta_2; t)$ and the most preferable value is the one maximizing $\beta(\theta; t)$. In this discrete setting, we will see that there is an interval of θ-values that are most preferable, but we can see that the median-unbiased estimator is one most preferable value. The improved confidence intervals are based on the idea of finding a better confidence curve than $\beta(\theta; t)$, i.e. a function $\alpha(\theta; t)$ such that $\alpha(\theta; t) \leq \beta(\theta; t)$ for all θ and t while $P_\theta[\alpha(\theta; T) > \alpha] \geq 1 - \alpha$. Confidence intervals based on $\alpha(\theta; t)$ are then shorter than the standard intervals but have the same minimum coverage probability. The next lemma gives the basic idea.

Lemma 1 *Let X have density $p_\theta(x)$ and let $\gamma(\theta, x)$ be any function of x and θ. Define $\lambda(\theta; x) = P_\theta[\gamma(\theta, X) \leq \gamma(\theta, x)]$. Then the set $S_\alpha(X) = \{\theta : \lambda(\theta; X) > \alpha\}$ is a $1 - \alpha$ confidence set for θ and the test which rejects if $\lambda(\theta_0; x) \leq \alpha$ is a level α test of the hypothesis $H : \theta = \theta_0$.*

Proof. Set $Z = \gamma(\theta, X)$ which is a random variable. Then $\lambda(\theta; x) = P_\theta(Z \leq z) = H_\theta(z)$, say, and consequently $P_\theta(\theta \notin S_\alpha(X)) = P_\theta(\lambda(\theta; X) \leq \alpha) = P_\theta(H_\theta(Z) \leq \alpha) \leq \alpha$. □

The level is exactly α for α and θ such that $\alpha = H_\theta(z)$ for some z (or equivalently $\alpha = \gamma(\theta, x)$ for some x). In particular, the level is α for all θ if $H_\theta(z)$ is continuous and strictly increasing. If $\gamma(\theta, x)$ is a constant in x, Lemma 1 is still true but vacuous.

One possible choice is $\gamma(\theta, x) = p_\theta(x)$. Then $\lambda(\theta; x)$ is the probability of obtaining a likelihood equal to or smaller than the one observed when θ is the true parameter. This function is used to obtain improved confidence limits for the binomial success probability by Sterne (1954). Notice that the sets from Lemma 1 will not in general be intervals, even if $\gamma(\theta, x) = p_\theta(x)$ and the density has monotone likelihood ratio. For instance, if $p_\theta(x)$ is increasing or decreasing for all x, we get one-sided intervals and if $p_\theta(x)$ is not unimodal, the resulting sets may be unions of disjoint intervals. For this reason, and to ease comparison with the standard approach, we prefer the choice $\gamma(\theta, x) = \min\{P_\theta(T \geq t), P_\theta(T \leq t)\}$ where $t = t(x)$.

3 The acceptability function

The following theorem provides valid $1 - \alpha$ confidence sets and unifies calculations in many standard problems.

Theorem 1 *Let the distribution of X be indexed by θ and let $T = T(X)$ be any statistic. Define the function $\gamma(\theta, t) = \min\{P_\theta(T \geq t), P_\theta(T \leq t)\}$ and let $\alpha(\theta; t) = P_\theta[\gamma(\theta, T) \leq \gamma(\theta, t)]$. Then*

1. *The set $S_\alpha(T) = \{\theta : \alpha(\theta; T) > \alpha\}$ is a $1 - \alpha$ confidence set for θ.*

2. *The following expressions hold:*

$$\alpha(\theta; t) = P_\theta(T \geq t) + P_\theta(T \leq t^*)$$

if $P_\theta(T \geq t) < P_\theta(T \leq t)$ and t^* is the largest u such that $P_\theta(T \leq u) \leq P_\theta(T \geq t)$,

$$\alpha(\theta; t) = 1$$

if $P_\theta(T \geq t) = P_\theta(T \leq t)$ and

$$\alpha(\theta; t) = P_\theta(T \leq t) + P_\theta(T \geq t^{**})$$

if $P_\theta(T \geq t) > P_\theta(T \leq t)$ and t^{**} is the smallest v such that $P_\theta(T \geq v) \leq P_\theta(T \leq t)$.

Proof.

1. Follows from Lemma 1.

2. Fix θ and define $g(t) = P_\theta(T \geq t)$ and $h(t) = P_\theta(T \leq t)$. Then $g(t)$ is nonincreasing and $h(t)$ is nondecreasing. If $\gamma(\theta, t) = g(t)$,

$$\begin{aligned} \alpha(\theta; t) &= P_\theta[\min\{g(T), h(T)\} \leq g(t)] \\ &= P_\theta[g(T) \leq g(t)] + P_\theta[h(T) \leq g(t)] \\ &= P_\theta(T \geq t) + P_\theta(T \leq h^{-1}(g(t))] \end{aligned}$$

where $h^{-1}(t) = \sup\{u : h(u) \leq t\}$. The case $\gamma(\theta, t) = h(t)$ is similar. If $g(t) = h(t)$, $\gamma(\theta, t)$ is maximized and hence $P_\theta[\gamma(\theta, T) \leq \gamma(\theta, t)] = 1$. □

If T is the sufficient statistic in a one-parameter exponential family, the function $\alpha(\theta; t)$ is the *acceptability function* defined in Spjøtvoll (1983) and Theorem 1 is contained in that paper. Viewed as a preference function, the acceptability function has some optimality properties for this family. Here, we only view it as a convenient way of generating confidence sets which by Theorem 1 are the level sets of $\alpha(\theta; t)$, i.e. the set of θ-values for which $\alpha(\theta; t)$ is above α when $T = t$ is observed. The following corollary is an immediate consequence of Theorem 1 and shows why $\alpha(\theta; t)$ leads to improved confidence sets in discrete distributions. Recall $U_\alpha(t) = \{\theta : \beta(\theta; t) > \alpha\}$ is the standard $1 - \alpha$ confidence interval.

Corollary 1 *It always holds true that* $\alpha(\theta; t) \leq \beta(\theta; t)$ *and consequently* $S_\alpha(t) \subseteq U_\alpha(t)$ *while* $1 - \alpha \leq P_\theta(\theta \in S_\alpha(T)) \leq P_\theta(\theta \in U_\alpha(T))$.

This corollary shows that the confidence sets $S_\alpha(t)$ based on the acceptability function are subsets, sometimes proper subsets, of the standard intervals $U_\alpha(t)$ and are less conservative but have the correct confidence coefficient. The rest of this paper investigates the actual improvement for spesific discrete distributions. For continuous distributions, $\alpha(\theta; t) = \beta(\theta; t)$ and no improvement is possible. It should be mentioned that there is a minuscule possibility that $S_\alpha(t)$ is a union of several disjoint intervals rather than one interval. In that case, we use as our confidence interval the smallest interval containing $S_\alpha(t)$, i.e. (θ_L, θ_U) where θ_L is the smallest θ such that $\alpha(\theta; t) \geq \alpha$ and θ_U is the largest θ such that $\alpha(\theta; t) \geq \alpha$. Since

$\alpha(\theta; t) \leq \beta(\theta; t)$ and $\beta(\theta; t)$ always gives intervals for stochastically increasing distributions, it is clear that these modified intervals are still subsets of the standard intervals and have the required minimum coverage probability.

4 Binomial distribution

Let X denote a binomial variate for sample size n with success probability p. The confidence interval based on $\beta(\theta; x)$ is

$$U_\alpha(x) = \{p: \min(P_p(X \geq x), P_p(X \leq x)) > \frac{\alpha}{2}\} = (p_L, p_U),$$

say, where p_L satisfies $P_{p_L}(X \geq x) = \alpha/2$ and p_U satisfies $P_{p_U}(X \leq x) = \alpha/2$ if $x \notin \{0, n\}$. This confidence interval is known as the Clopper-Pearson "exact" confidence interval for p. Let f_{α,ν_1,ν_2} denote the upper α quantile in the Fisher distribution with ν_1 and ν_2 degrees of freedom. Using the fact that $P_p(X \geq x) = P[F_{2x,2(n-x+1)} < (n-x+1)p/(x(1-p))]$, Leemis and Trivedi (1996) p.67 show that the interval can be written

$$\left[1 + \frac{n-x+1}{x f_{2x,2(n-x+1),1-\alpha/2}}\right]^{-1} < p < \left[1 + \frac{n-x}{(x+1) f_{2(x+1),2(n-x),\alpha/2}}\right]^{-1},$$

This interval is usually treated as a gold standard when comparison between different confidence sets for p is made, but is often criticised for being too conservative. The confidence set $S_\alpha(x) = \{p: \alpha(p; x) > \alpha\}$ is superior to $U_\alpha(x)$ since it has the same nominal level but is shorter. This is illustrated in Figure 1 where we plot $\alpha(p; x)$ and $\beta(p; x)$ together with coverage probabilities as a function of p for $\alpha = .05$. Table 1 shows the improvement in mean coverage probability, i.e. $\int_0^1 C_n(p)dp$ where $C_n(p)$ is the coverage probability if p is the true value.

4.1 Comparison with approximate methods

A number of authors have been concerned with confidence intervals for p, e.g. Vollset (1993), Leemis and Trivedi (1996) and Agresti and Coull (1998). The consensus in these papers is that one should use some approximate method such as (continuity corrected) score intervals or adjusted Wald intervals since these intervals have almost the right level and are shorter than the Clopper-Pearson intervals. It is our contention that no approximate solution is necessary, since the interval $S_\alpha(x)$ has level $1-\alpha$ for all n and is considerably shorter than $U_\alpha(x)$. Table 1 shows average width of the Clopper-Pearson(C-P) and acceptability (ACC) intervals as well as average widths of the adjusted Wald (AW) and score (SC) intervals. Let $\hat{p} = X/n$ and let $\Phi(\cdot)$ be the cdf of the standard normal. Then the score interval is based on the approximate confidence curve $2\Phi(-|\hat{p}-p|/\sqrt{p(1-p)/n})$ and the Wald interval on $2\Phi(-|\hat{p}-p|/\sqrt{\hat{p}(1-\hat{p})/n})$. For adjusted Wald interval, replace n with $n+4$ and $\hat{p}$ with $(X+2)/(n+4)$, see Agresti and Coull (1998) who recommend these intervals rather than the Clopper-Pearson intervals. However,these procedures have confidence coefficient well below the nominal level. The acceptability intervals are about halfway in length between the Clopper-Pearson intervals and the score intervals and have approximately the same average length as the adjusted Wald intervals. In fact, the average width of the acceptability intervals is the shortest possible for any confidence interval with the correct coverage probability, at least for the cases studied in Table 1 and Figure 2.

4.2 Comparison with other exact procedures

Significant contributions to the construction of exact binomial confidence intervals are Sterne (1954), Crow (1956), and, more recently, Blyth and Still (1983). Sterne (1954) works with ideas close to ours. He considers the probability of obtaining a number of successes as probable as or less probable than the observed x, i.e. $\lambda(p;x) = P_p[\gamma(p,X) \leq \gamma(p,x)]$ where $\gamma(p,x) = P_p(X=x) = \binom{n}{x}p^x(1-p)^{n-x}$ in the notation of Lemma 1. He then argues that one should consider as a confidence set for p all values of p such that $\lambda(p;x) > \alpha$. Lemma 1 then guarantees that the

Table 1: Coverage Probabilities and Widths when $1-\alpha = .95$

	Mean Coverage				Minimum Coverage				Average Width			
n	C-P	ACC	AW	SC	C-P	ACC	AW	SC	C-P	ACC	AW	SC
5	.990	.980	.965	.955	.975	.950	.879	.832	.678	.626	.586	.558
10	.984	.973	.964	.954	.961	.950	.917	.835	.508	.475	.457	.435
30	.973	.963	.960	.953	.951	.950	.934	.837	.299	.282	.279	.271
50	.969	.960	.958	.952	.953	.950	.935	.838	.231	.220	.218	.213

corresponding confidence statement holds with the desired accuracy. The resulting confidence sets are numerically close, but not identical, to those obtained from acceptability. There are two disadvantages with the Sterne system, as pointed out by Crow (1956). Firstly, the resulting confidence sets may be the union of several disjoint intervals rather than one single interval (though this seems to happen only at very low levels, e.g. for $n = 3$, $1 - \alpha = .442$ as cited in Casella and Berger 1990 p.417). Secondly, the resulting intervals are not necessarily subsets of the corresponding Clopper-Pearson intervals, so there is no uniform improvement like there is for acceptability. The confidence sets from acceptability may also be unions of several disjoint intervals (though this does not happen at reasonable significance levels in practice), but this is not a big disadvantage since we then use the smallest single interval containing all values and the resulting confidence interval is still a subset of the Clopper-Pearson intervals and hence Corollary 1 still applies.

The intervals considered by Crow (1956) and Blyth and Still (1983) are based on the same technique, which is essentially to adjust Sterne's system to avoid unreasonable behavior. For each n and p_0, there is an acceptance interval $A_n(p_0) \leq X \leq B(p_0)$ of the hypothesis $H : p = p_0$ against $K : p \neq p_0$. These acceptance regions are then chosen to be as short as possible, i.e. include as few X-values as possible. Crow (1956) proved that every confidence interval given by a family of minimum-length acceptance intervals makes the sum of the $n + 1$ possible lenghts of the confidence interval as small as possible for the given level. In many cases, there are several equally short acceptance intervals. If one of them is the Clopper-Pearson acceptance interval, this should be chosen to guarantee that the corresponding intervals are included in the Clopper-Pearson intervals. Further, one must make sure that the acceptance intervals are nested, i.e. if p_0 increases, we must only include larger and not smaller X-values. This makes sure that the corresponding confidence sets are intervals. These requirements are not sufficient to guarantee uniqueness, so further rules must be introduced, see Blyth and Still (1983) p.110. For instance, when $n = 8$, $1 - \alpha = .95$, the possible shortest acceptance regions for $.3155 \leq p_0 \leq .3995$ are now $(0 \leq X \leq 5)$ or $(1 \leq X \leq 6)$. Blyth and Still now changes their acceptance region at the midpoint $p_0 = .3575$ while Crow always chooses the acceptance interval furthest to the right. The latter rule leads to some anomaly as the confidence intervals sometimes do not change regularly as X increases. For instance, if $n = 14$, $\alpha = .05$,

$X = 6, 7, 8$, Crow's intervals are $(.206, .688)$, $(.206, .794)$,and $(.312, .794)$ while the acceptability intervals are $(.201, .688)$, $(.231, .769)$, and $(.312, .799)$, respectively. To see that the intervals in Blyth and Still (1983) are not the same as $S_\alpha(x)$, notice that their acceptance interval for $n = 8$, $1 - \alpha = .95$ is $1 \leq X \leq 6$ for $p = .3575$ while the acceptability region is $0 \leq X \leq 5$ (it includes $X = 6$ for $p = .3585$, so the difference is not great). The acceptability intervals are very close to the Blyth-Still system, and compared to the Crow system, the acceptability intervals are longer when X is near 0 or n and shorter when X is close to $n/2$. The improved intervals in Crow (1956) and Blyth and Still (1983) all involve some arbitrary choice between acceptance regions and hence are not completely satisfactory. Also, it is a formidable task to compute the intervals to find e.g. the smallest α for which a given p is in the interval when X is observed. In practice, a fine grid of parameter values must be generated and then one needs to find shortest acceptance regions according to their rules. They do however have guaranteed minimum total lenght due to their acceptance region origin, and this property is not shared by the acceptability intervals in theory (but it is for all examples considered here, e.g in Table 1). A counterexample is $n = 40$, when $\alpha(.75, 27) = .1994$ and $\alpha(.75, 35) = .1507$, so the acceptance interval from acceptability is $[27, 35]$ but the shortest acceptance region is $[28, 35]$. The acceptability intervals have the advantage of being a more natural construction, directed at improving the standard interval, and with a precise control of the coverage probability.

4.3 When is the confidence coefficient exactly $1 - \alpha$?

As can be seen from Figure 1, Clopper-Pearson intervals have the undesirable feature that for p near 0 or 1, the actual coverage is about $1-\alpha/2$ rather than $1-\alpha$ (Vollset 1993 p.822). For $n \leq 5$ this is true for all p. No such problems arise with $S_\alpha(x)$. Consider the case $n = 1$ which is most transparent. Let $\alpha < 1/2$. Then $U_\alpha(x)$ is $[0, 1 - \alpha/2)$ if $x = 0$ and $(\alpha/2, 1]$ if $x = 1$, while $S_\alpha(x) = [0, 1 - \alpha]$ or $[\alpha, 1]$ respectively. It follows that $S_\alpha(X)$ has minimum coverage probability $1 - \alpha$ and $U_\alpha(X)$ has $1 - \alpha/2$.

It may be of some interest to see when the coverage probability is exactly $1 - \alpha$. From Lemma 1, if $\gamma(p, x) = \min\{P_p(X \geq x), P_p(X \leq x)\}$ and $Z = \gamma(p, X)$, the the level is exactly $1-\alpha$ for p such that

$$H_p(z)=\sum_{k=0}^{n}P_p(Z\le z|X=k)P_p(X=k)=\sum_{k=0}^{n}I\{\gamma(p,k)\le z\}P_p(X=k)=\alpha$$

for some z. Figure 3 shows $H_p(z)$ as a function of z for the eight values of p such that $H_p(z)=.05$ for some z when $n=5$. The coverage probability of the acceptability intervals is therefore exactly .95 for these values of p, which can also be seen in Figure 1.

Since the acceptability function is not continuous, confidence intervals are not available at all levels. There are intervals $\mathcal{I}$ such that $\{p\colon \alpha(p;x)>\alpha\}$ is the same for $\alpha\in\mathcal{I}=[\alpha_L,\alpha_U]$, i.e. the confidence intervals at level $1-\alpha_L$ and $1-\alpha_U$ are the same. For example, when $n=1$ and $\alpha>1/2$, $\{p\colon \alpha(p;X)>\alpha\}=[0,1/2]$ when $X=0$ and $[1/2,1]$ when $X=1$. This is due to the discreteness and shows that the entire confidence curve should be shown, not just the level sets.

We can in fact see directly which levels are available from $\alpha(\theta,t)$. Recall $\gamma(\theta,t)=\min\{P_\theta(T\ge t),P_\theta(T\le t)\}$ and $\alpha(\theta;t)=P_\theta[\gamma(\theta,T)\le\gamma(\theta;t)]$. The $1-\alpha$ confidence set from acceptability is $S_\alpha(t)=\{\theta\colon \alpha(\theta;t)>1-\alpha\}$. A confidence set has exact level $1-\alpha$ if $\inf_\theta P(\theta\in S_\alpha(T))=1-\alpha$. From Lemma 1, $S_\alpha(t)$ has exact level $1-\alpha$ if $H_\theta(z)=\alpha(\theta;t)$ attains the value α, i.e. if there are values θ_0, t_0 such that $\alpha(\theta_0;t_0)=\alpha$. If there are no values of θ, t such that $\alpha(\theta;t)=\alpha$, then the confidence sets $S_\alpha(t)$ must be equal for all $\alpha\in[\alpha_L,\alpha_U)$, say, and consequently S_α can not have confidence coefficient $1-\alpha$ for $\alpha\in(\alpha_L,\alpha_U)$. If there are only a finite number of possible values of t, we can generate all possible functions $\alpha(\theta;t)$ as a function of θ for fixed t. A plot of these then shows which levels $1-\alpha$ that can be achieved exactly.

Figures 4, 5 and 6 show the situation for $X\sim\text{Bin}(n,p)$, $n=1,2$, and 3. For $n=1$, $1-\alpha\ge 1/2$ is attainable, and for $n=2$, $1-\alpha\ge\sqrt{2}-1$ is attainable. Since $S_\alpha(t)$ does not shrink to a point as α approaches 1, it is clear that there will always be an interval of α-values near 1 where the acceptability intervals are not exact. We could hope that this was the only exception, but the case $n=3$ shows more complicated behavior. Here, the attainable levels are $1-\alpha\ge 0.5$ and $0.375\le 1-\alpha\le 0.444$. Notice that the values of α which are unattainable are exactly the values of y which the equation $y=\alpha(p;x)$ has no solution for any $x\in\{0,1,\ldots,n\}$.

Finally, notice from Figure 1 that the acceptability of p is 1 in an interval including the median-unbiased estimator $\hat{p}$, which means that all values of p in this interval will never be rejected at any level in a test when x is observed. This is quite reasonable and again reflects the discrete nature of the problem. If we observe $x = 3$ when $n = 5$ and test the hypothesis $H : p = .59$ against $K : p \neq .59$, one should never reject at any level because no other observed x-value supports the null hypothesis more than $x = 3$.

5 Poisson distribution

Let X be a Poisson variate with mean λ. The confidence interval based on $\beta(\lambda; x)$ is $U_\alpha(x)$ which can be written $(1/2)(\chi^2_{2X,1-\alpha/2}, \chi^2_{2(X+1),\alpha/2})$ since $P_\lambda(X \leq x) = 1 - G_{2(x+1)}(2\lambda)$ where $G_\nu(\cdot)$ is the cdf in the chi-square distribution with ν degrees of freedom. This interval was first given by Garwood (1936). The Garwood intervals suffer from the same problem as the Clopper-Pearson intervals, being too wide and yielding coverage probabilities strictly greater than $1-\alpha$, especially for small λ where the level is effectively $1-\alpha/2$. The intervals $S_\alpha(X) = \{\lambda\colon \alpha(\lambda; X) > \alpha\}$ based on the level sets of the acceptability function alleviate these problems and have coverage probabilities much closer to the nominal level, see Figure 8. We have also plotted the corresponding confidence curves when $X = 5$ is observed in Figure 7. Some authors, e.g. Agresti and Coull (1998) suggest using some approximate interval with close to the nominal level to get shorter intervals. Improvement on the Garwood interval while maintaining exact level $1 - \alpha$ is considered by Crow and Gardner (1959) and Casella and Robert (1989). In the latter paper, a numerical procedure for improving confidence sets is given and the confidence limits using this procedure are very close, but not identical, to the corresponding confidence sets from the acceptability function. The improvement in length using the procedure of Crow and Gardner (1959) is larger but those intervals have the disadvantage that the endpoints are not strictly increasing in x, which is rather counterintuitive. For instance, for $\alpha = .1$ and $x = 8, 9, 10, 11, 12, 13$, their confidence intervals are $(4.53, 13.55)$, $(4.53, 15.30)$, $(5.98, 15.99)$, $(5.98, 17.81)$, $(7.51, 18.40)$ and $(7.51, 20.05)$. The corresponding acceptability intervals are $(4.31, 14.23)$, $(4.72, 15.29)$, $(5.81, 16.74)$, $(6.23, 17.81)$, $(7.30, 19.23)$, and $(7.72, 20.26)$.

6 Negative binomial distribution

Let X be the number of noncases before r cases are observed with prevalence rate p so X follows the negative binomial distribution with parameters r and p,

$$P(X = x) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x, \quad x = 0, 1 \ldots$$

where r is known and p is unknown. This family of distributions has monontone likelihood ratio in $T(x) = -x$. The case $r = 1$ is the geometric distribution. Lui (1995) gives tables for the standard interval $U_\alpha(x)$ for $\alpha = .05$. In Figure 9, the confidence curves $\alpha(\theta; x)$ and $\beta(\theta; x)$ are plotted when $x = 5$ is observed and r is 1, 5 or 20. The interpretation of these curves is that they show all possible confidence intervals for θ when $X = x$ is observed. For example, $\alpha(.185; 5) = \alpha(.749; 5) = .05$ when $r = 5$ means the 95% interval when $X = 5$ is $(.185, .749)$ using acceptability. The standard interval is found to be $(.187, .788)$ since $\beta(.187; 5) = \beta(.788; 5) = .05$.

In Figure 9, we have also plotted the coverage probabilities for the standard interval and the acceptability interval when $\alpha = .05$. The standard interval has the undesirable feature that for p near 1, the level is about $1 - \alpha/2$ rather than $1 - \alpha$. The problem is worse when r is small. No such problems arise with the acceptability interval. For small α and p not too close to 0, the improved interval from the acceptability function is close to the interval obtained from the standard formula by using the α-quantile instead of the $\alpha/2$-quantile in the upper limit. This works well in particular for the geometric distribution.

7 Hypergeometric distribution

Assume a finite universe consisting of N units of which an unknown number A has a particular attribute. We sample n units from the universe without replacement and record the number of observed element with the attribute, x. It follows that X follows a hypergeometric distribution,

$$P(X = x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}.$$

We can work with A or $P = A/N$ as unknown parameter. Following Wright (1991), we regard A as the unknown parameter. It follows from Lehmann (1986) p.80 that this distribution has monotone likelihood ratio in $T(x) = x$. A standard reference on inference in this distribution is Wright (1991), which also contains extensive tables of standard confidence sets. Since A takes values in $\{0, 1, \ldots, N\}$, we are now considering confidence sets rather than intervals, but as shorthand, let $[A_1, A_2]$ mean the set $\{A_1, A_1 + 1, \ldots, A_2\}$. Once again, the standard confidence set is obtained by inverting the equal-tailed test (Buonaccorsi 1987) and can hence be written

$$U_\alpha(x) = \{A\colon \min(P_A(X \geq x), P_A(X \leq x)) > \alpha/2\},$$

while our improved set is $S_\alpha(x) = \{A\colon \alpha(A; x) > \alpha\}$. In Figure 10, we have plotted the confidence curves $\alpha(A; x)$ and $\beta(A; x)$ for $x = 10$ when $N = 100, n = 13$. Figure 11 shows coverage probabilities for the standard and improved intervals at level $1 - \alpha = .95$ for for $N = 20, n = 4$ and $N = 100, n = 13$ which corresponds to examples 3.8 and 3.9 in Wright (1991). The standard intervals are very conservative in these cases, having coverage probability at least .9855 and .9641, respectively. There is room for improvement, and our procedure, which also only uses the hypergeometric distribution, has level .9566 and .9510 in the two cases. Wright (1991), p.51, incorrectly states that the excess probability for the standard interval is as small as possible. In both cases, the 95% interval $S_{.05}(x)$ is a proper subset of the standard interval for all x. For $N = 20, n = 4$, the standard 95% sets for $x \in \{0, 1, 2, 3, 4\}$ are

$$[0, 11], [1, 15], [2, 18], [5, 19], [9, 20],$$

Wright (1991) p.51, while the improved sets are

$$[0, 10], [1, 14], [3, 17], [6, 19], [11, 20].$$

We can also obtain tests for $H : A = A_0$ against $K : A \neq A_0$ that are better than the standard equal-tailed test in the sense of having a type I-error closer to the nominal level, which in turn will give higher power. For instance, if $A_0 = 50$ when $N = 100, n = 13$, then the standard test at level .05 rejects when $X \geq 11$ or $X \leq 2$ with actual level .0147, while the test which rejects when $\alpha(A_0; x) \leq .05$ also includes $X = 10$ in the rejection region and has actual level .0430.

8 Confidence intervals for quantiles

Let $X_1, \ldots, X_n$ be i.i.d. with continuous distribution $F(x)$ and let $F(x_\varepsilon) = \varepsilon$, i.e. x_ε is the ε-quantile for this distribution. Nonparametric inference about x_ε can be based on the statistic $S_n(\theta) = \sum_{i=1}^n I\{X_i > \theta\}$ since $S_n(x_\varepsilon)$ has a binomial distribution with parameters n and $1-\varepsilon$. The p-value of the equal-tailed test of $H: x_\varepsilon = \theta$ against $K: x_\varepsilon \neq \theta$, is $\beta(\theta; X) = \min\{\beta^0(\theta; X), 1\}$ where

$$\beta^0(\theta; X) = 2\min\{P(Z \geq S_n(\theta)), P(Z \leq S_n(\theta))\}$$

and $Z \sim \text{Bin}(n, 1-\varepsilon)$. The set $U_\alpha(X) = \{\theta: \beta(\theta; X) \geq \alpha\}$ is a $1-\alpha$ nonparametric confidence interval for x_ε (Lehmann 1975 p.185). Theorem 1 is applicable and the acceptability of θ can be found from the formula in Theorem 1, replacing t with $S_n(\theta)$.

The corresponding confidence interval $S_\alpha(X) = \{\theta: \alpha(\theta; X) > \alpha\}$ is consequently a subset of the standard interval and has confidence level at least $1-\alpha$. Figure 12 shows $\alpha(\theta; X)$ and $\beta(\theta; X)$ when $\theta = x_{.25}$ and $n = 25$. This is plotted for $X_i = i$ for $i = 1, 2, \ldots, 25$, which is of course artificial but convenient since $X_{(i)} = i$. The standard 95% interval is $(X_{(2)}, X_{(12)})$ while the improved interval is $(X_{(2)}, X_{(11)})$, and the coverage probabilities are .982 and .963, respectively. Since, if $Z \sim \text{Bin}(25, .25)$, $P(Z \leq 1) = .007$, $P(Z \leq 2) = .032$, $P(Z \geq 12) = .011$ and $P(Z \geq 11) = .030$, it is obvious that we can obtain a shorter interval in this case by requiring (1) rather than (2), and in this simple situation it is clear how acceptability works. But it is interesting that these conclusions are obtained through Theorem 1. For $\varepsilon = 1/2$, $Z \sim \text{Bin}(n, 1/2)$ so $P(Z \geq z) = P(Z \leq n - z)$ and consequently $\alpha(\theta; X) = \beta(\theta; X)$ so this method does not give improved confidence intervals for the median.

9 Summary and discussion

We have introduced a new method for constructing exact confidence sets that leads to shorter and less conservative exact confidence sets for discrete distributions. For stochastically increasing distributions, attention can be restricted to intervals and the method gives intervals that are easy to compute and do not involve anything except the distribution at hand. Numerical results show that the improved intervals are much less conservative than standard exact intervals. Since our improved intervals do not have coverage probability equal to the nominal level for all parameter values, and indeed this can not be achieved by any non-randomized procedure, it may be possible to construct even more accurate exact confidence intervals. For instance, procedures like the construction in Crow and Gardner (1959) may be used, but the resulting intervals have some non-intuitive flavor since some arbitrary choice between acceptance regions has to be made. They will not uniformly dominate acceptability, but may give shorter intervals in an average sense, at the expense of even more conservative behavior in some parts of the parameter space. We find the intervals based on the acceptability function to have a more intuitive appeal. In conclusion, the improved intervals are always subsets of the standard intervals and are guaranteed to have the nominal confidence level, so there is nothing to lose but there may be a substantial gain from adapting the new procedure.

Appendix: S-Plus functions

This section contains S-Plus code (Venables and Ripley 1994) for computing the acceptability function $\alpha(\theta; x)$ as a function of x and θ in the binomial, Poisson, negative binomial, and hypergeometric distributions. Recall that in S-plus, the quantile function $Q(u)$ for discrete distributions gives the smallest integer m such that $F(m) \geq u$. The checks for the negative binomial and hypergeometric distributions are necessary because $Q(u)$ sometimes breaks down for u very close to 0 or 1.

```
acceptbin_function(x, n, p){
#Computes the acceptability of p in the
#binomial distribution when x is the
#observed number of successes on n attempts
       p1_1 - pbinom(x - 1, n, p)
       p2_pbinom(x, n, p)
       a1_p1 + pbinom(qbinom(p1, n, p) - 1, n, p)
       a2_p2 + 1 - pbinom(qbinom(1 - p2, n, p),n,p)
       return(min(a1,a2))
}
```

```
acceptpoisson_function(x, lambda){
#Computes the acceptability of lambda in
#the poisson distribution when x is observed
       p1_1 - ppois(x-1, lambda)
       p2_ppois(x, lambda)
       a1_1
       if (x != 0){
          a1_p1 + ppois(qpois(p1, lambda) -1,lambda)
       }
       a2_p2 + 1 - ppois(qpois(1 - p2, lambda), lambda)
       return(min(a1,a2))
}

acceptnegbin_function(x, r, p){
#Computes the acceptability in the negative binomial distribution when x
#is the observed number of failures before r successes is observed
       if (x !=0){
          p1_1 - pnbinom(x - 1, r, p)
          if (p1 > 0.99999) a1_1
          if (p1 < 0.00001) a1_0
          if (p1 < 0.99999 && p1 > 0.00001) {
              a1_p1 + pnbinom(qnbinom(p1, r, p) -1, r, p)
          }
       }
       p2_pnbinom(x, r, p)
       if (p2 > 0.99999) a2_1
       if (p2 < 0.00001) a2_0
       if (p2 < 0.99999 && p2 > 0.00001) {
           a2_p2 + 1 - pnbinom(qnbinom(1 - p2, r, p), r, p)
       }
       if (x == 0) return(a2)
       return(min(a1, a2))
}
```

```
accepthyper_function(x, A, N, n){
#Computes the acceptability of A (taking values 0, 1, ...,N) when x is
#an observation from the hypergeometric distribution with N units, A of
#which are of a particular type, and x of this type are obtained on n attempts.
      if (A < x) return(0)
      if (A > N - (n - x)) return(0)
      p1_1 - phyper(x - 1, A, N-A, n)
      if (p1 > 0.99999) a1_1
      if (p1 < 0.00001) a1_0
      if (p1 < 0.99999 && p1 > 0.00001){
          a1_p1 + phyper(qhyper(p1, A, N-A, n) - 1, A, N-A, n)
      }
      p2_phyper(x, A, N-A, n)
      if (p2 > 0.99999) a2_1
      if (p2 < 0.00001) a2_0
      if (p2 < 0.99999 && p2 > 0.00001){
          a2_p2 + 1 - phyper(qhyper(1 - p2, A, N-A, n), A, N-A, n)
      }
      return(min(a1,a2))
}
```

Acknowledgements

This research was supported by grant 121566/410 from the Research Council of Norway.

References

Agresti, A., and Coull, B.A. (1998). Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, **52**, 119-126.

Birnbaum, A. (1961), Confidence Curves: An Omnibus Technique for Estimation and Testing Statistical Hypotheses, *Journal of the Americal Statistical Association,* **56**, 246-249.

Blyth, C.R., and Still, H.A. (1983). Binomial Confidence Intervals, *Journal of the Americal Statistical Association,* **78**, 108-116.

Buonaccorsi, J.P. (1987). A Note on Confidence Intervals for Proportions in Finite Populations, *The American Statistician*, **41**, 215-218.

Casella, G. and Robert, C. (1989). Refining Poisson Confidence Intervals, *Canadian Journal of Statistics,* **17**, 45-57.

Casella, G. and Berger, R. (1990). *Statistical Inference.* Wadsworth, Pacific Grove, California.

Clopper, C.J. and Pearson, E.S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika,* **26**, 404-413.

Crow, E.L. (1956). Confidence Intervals for a Proportion, *Biometrika,* **43**, 423-435.

Crow, E.L. and Gardner, R.S. (1959). Confidence Intervals for the Expectation of a Poisson Variable, *Biometrika,* **46**, 441-453.

Garwood, F. (1936). Fiducual Limits for the Poisson Distribution, *Biometrika,* **28**, 437-442.

Leemis, L.M. and Trivedi, K.S. (1996). A Comparison of Approximate Interval Estimators for the Bernoulli Parameter, *The American Statistician*, **50**, 63-68.

Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

Lehmann, E.L. (1986). *Testing Statistical Hypothesis.* Wiley, New York.

Lui, K.-J.(1995). Confidence Limits for the Population Prevalence Rate Based on the Negative Binomial Distribution. *Statistics in Medicine,* **14**, 1471-1477.

Spjøtvoll, E. (1983), Preference Functions, in *A Festschrift for Erich L. Lehmann,* eds. P.J.Bickel, K.Doksum, and J.L.Hodges Jr, 409-432, Wadsworth, Belmont, Califonia.

Sterne, T.H. (1954). Some Remarks on Confidence or Fiducial Limits, *Biometrika,* **41**, 275-278.

Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus.* Springer, New York.

Vollset, S.E. (1993) Confidence Intervals for a Binomial Proportion, *Statistics in Medicine,* **12**, 809-824.

Wright, T. (1991). *Exact Confidence Bounds when Sampling from Small Finite Universes.* Springer, New York.

Helge Blaker, Department of Mathematics, University of Oslo, P.O.Box 1053 Blindern, N-0316 Oslo, Norway.

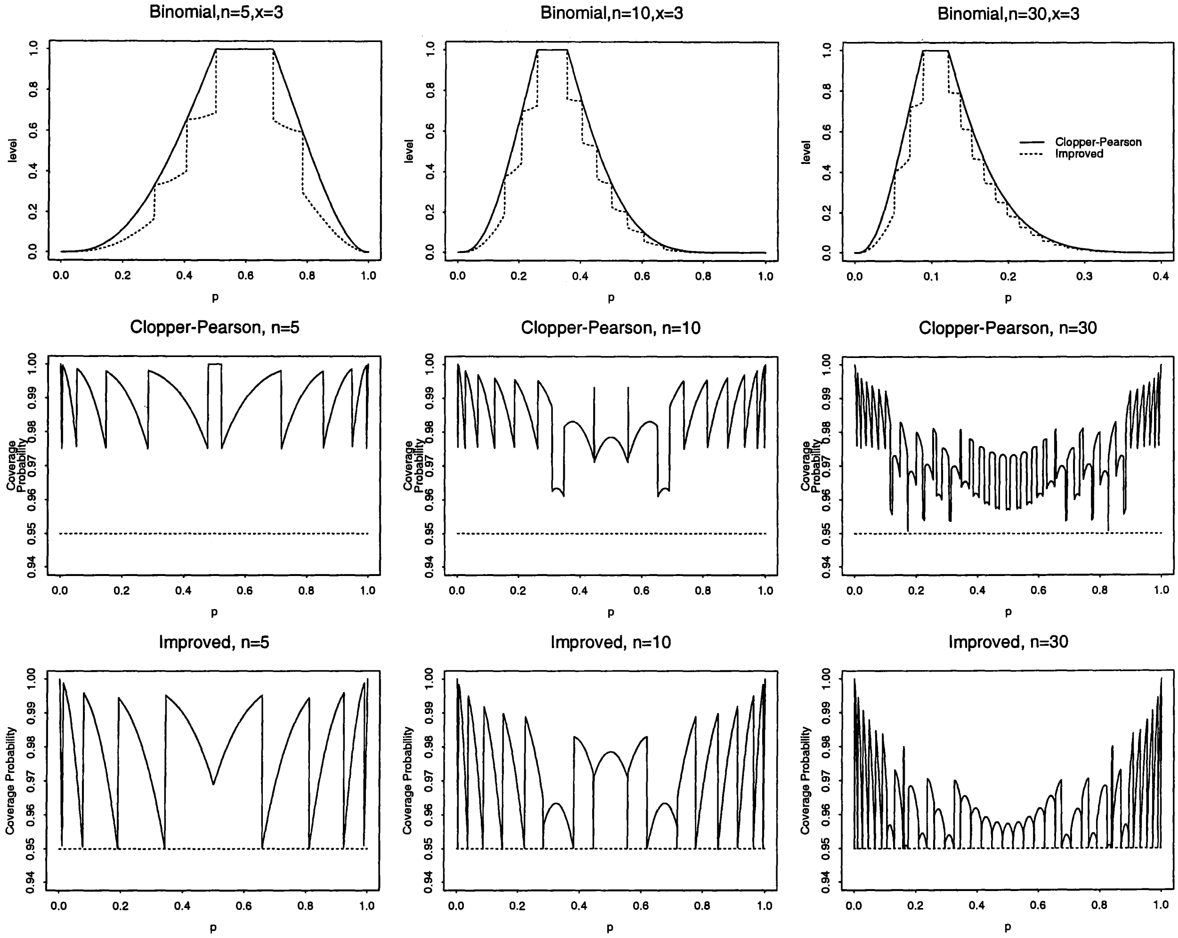

Figure 1: Top: Exact and approximate confidence curves; dotted line is exact $\alpha(p;t)$, unbroken line is approximate, $\beta(p;t)$. Middle: Coverage probability for 95% Clopper-Pearson confidence interval. Bottom: Coverage probability for improved 95% confidence interval.

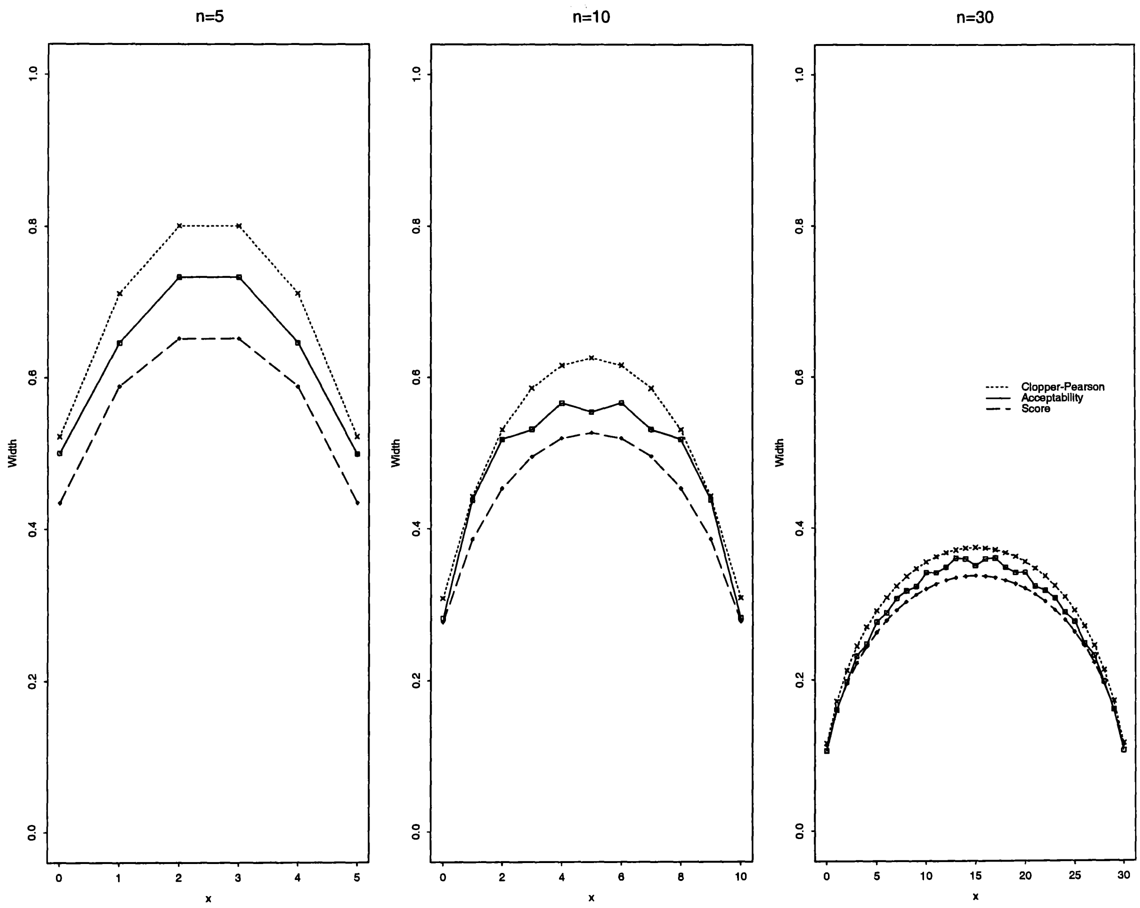

Figure 2: Length of different intervals for $n = 5, 10$ and 30.

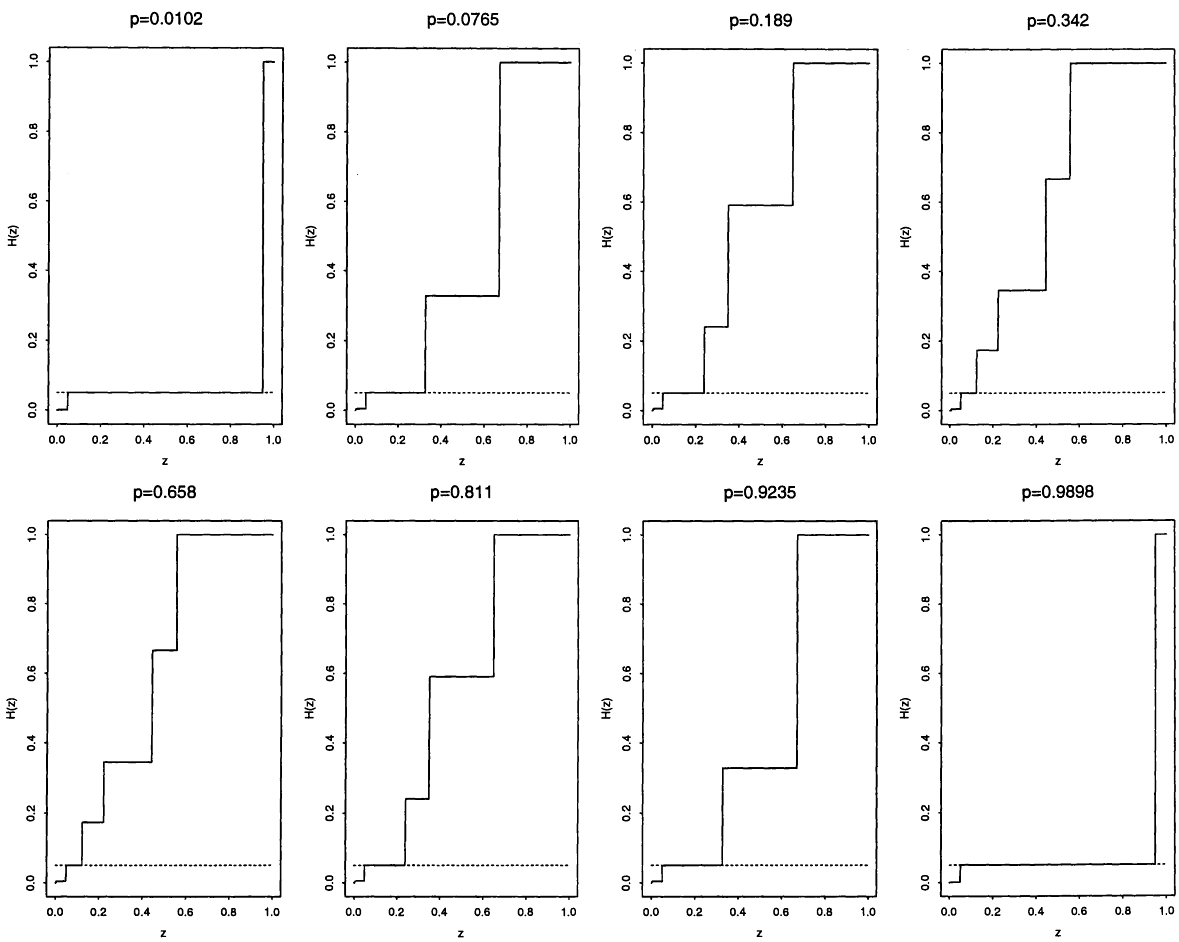

Figure 3: The function $H_p(z)$ in the binomial distribution when $n = 5$, $p = .0102$, $.0765, .189, .342, .658, .811, .9235, .9898$ which are the eight values for which $H_p(z) = .05$ for some z so the corresponding confidence interval has exact coverage probability .95 for these values of p.

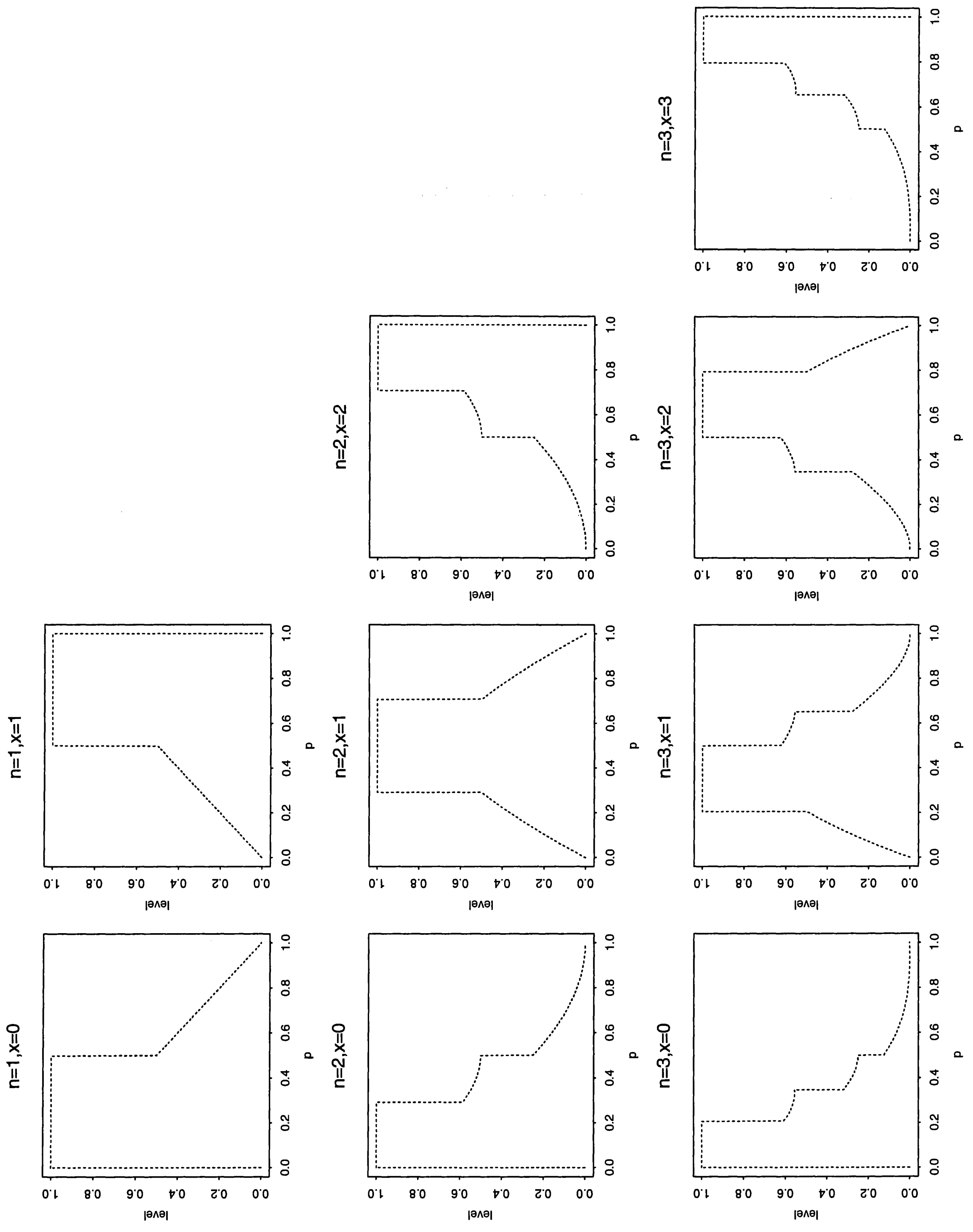

Figure 4: All possible confidence curves $\alpha(p;x)$ in the binomial distribution when $n = 1, 2$ or 3.

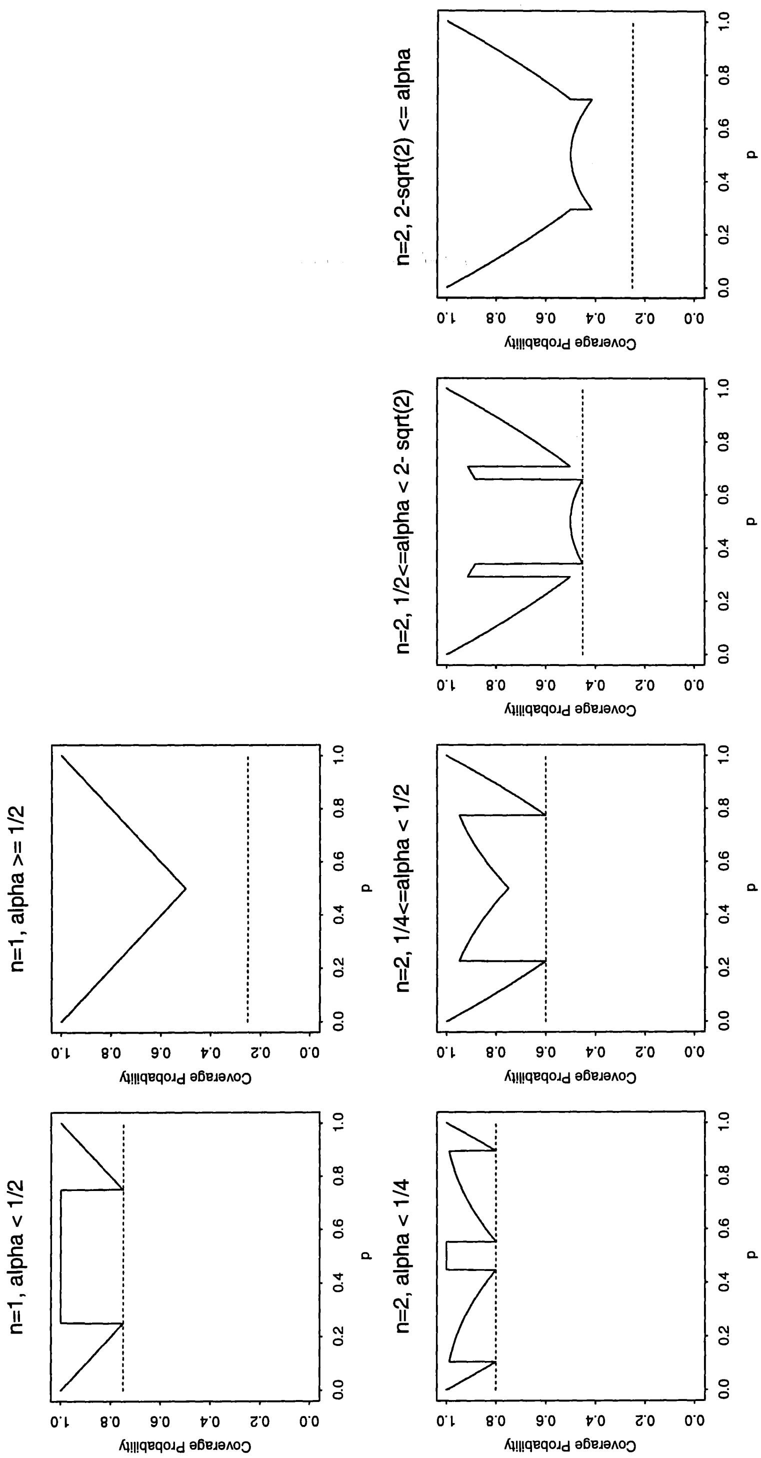

Figure 5: Coverage probabilities for $1 - \alpha$ acceptability intervals when $n = 1$ or 2. The intervals do not have correct confidence coefficient when $1 - \alpha < 1/2$ for $n = 1$ and when $1 - \alpha < \sqrt{2} - 1$ when $n = 2$. The broken line is $y = 1 - \alpha$.

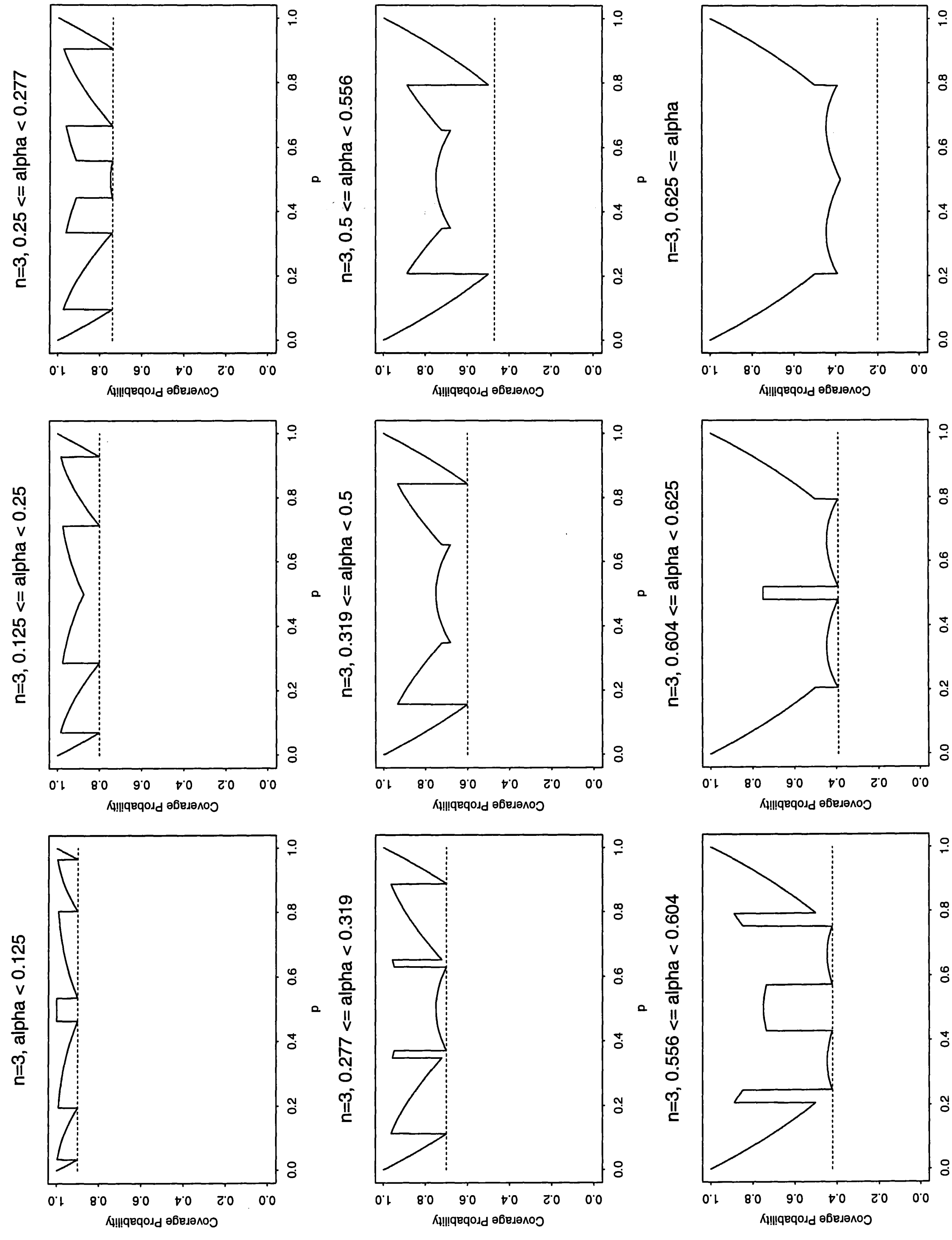

Figure 6: Coverage probabilities for $1-\alpha$ acceptability intervals when $n = 3$. The intervals do not have correct confidence coefficient when $.444 < 1-\alpha < 0.5$ or $1-\alpha < .375$. The broken line is $y = 1-\alpha$.

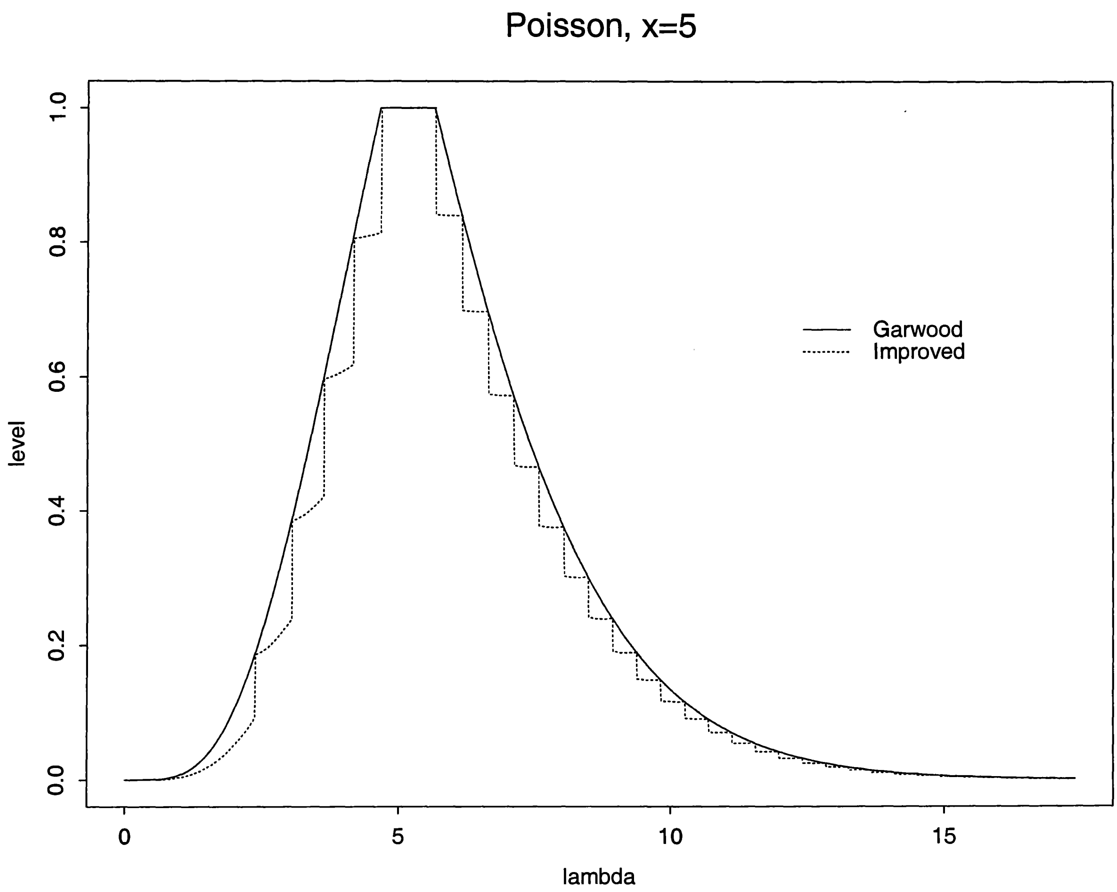

Figure 7: Confidence curves from acceptability and standard method in Poisson distribution when $X = 5$ is observed.

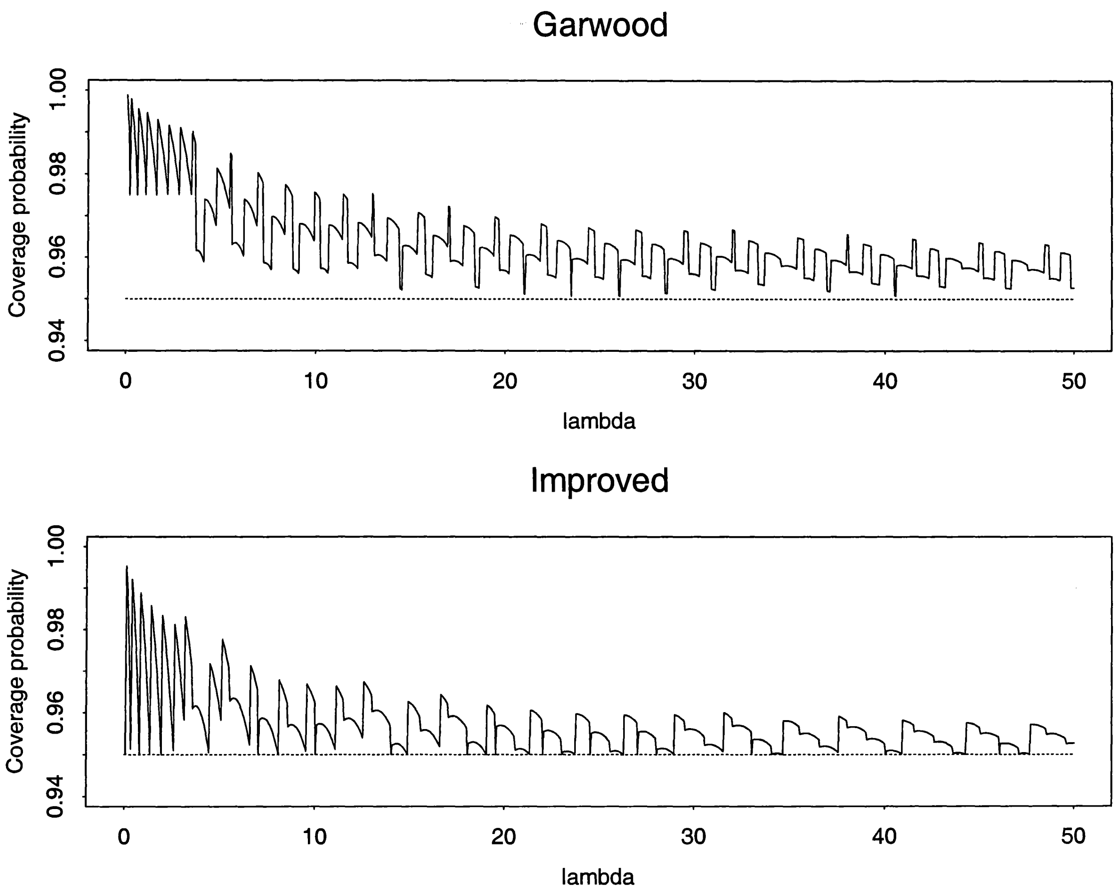

Figure 8: Top: Coverage probability for 95% Garwood intervals. Bottom: Coverage probability for 95% acceptability intervals.

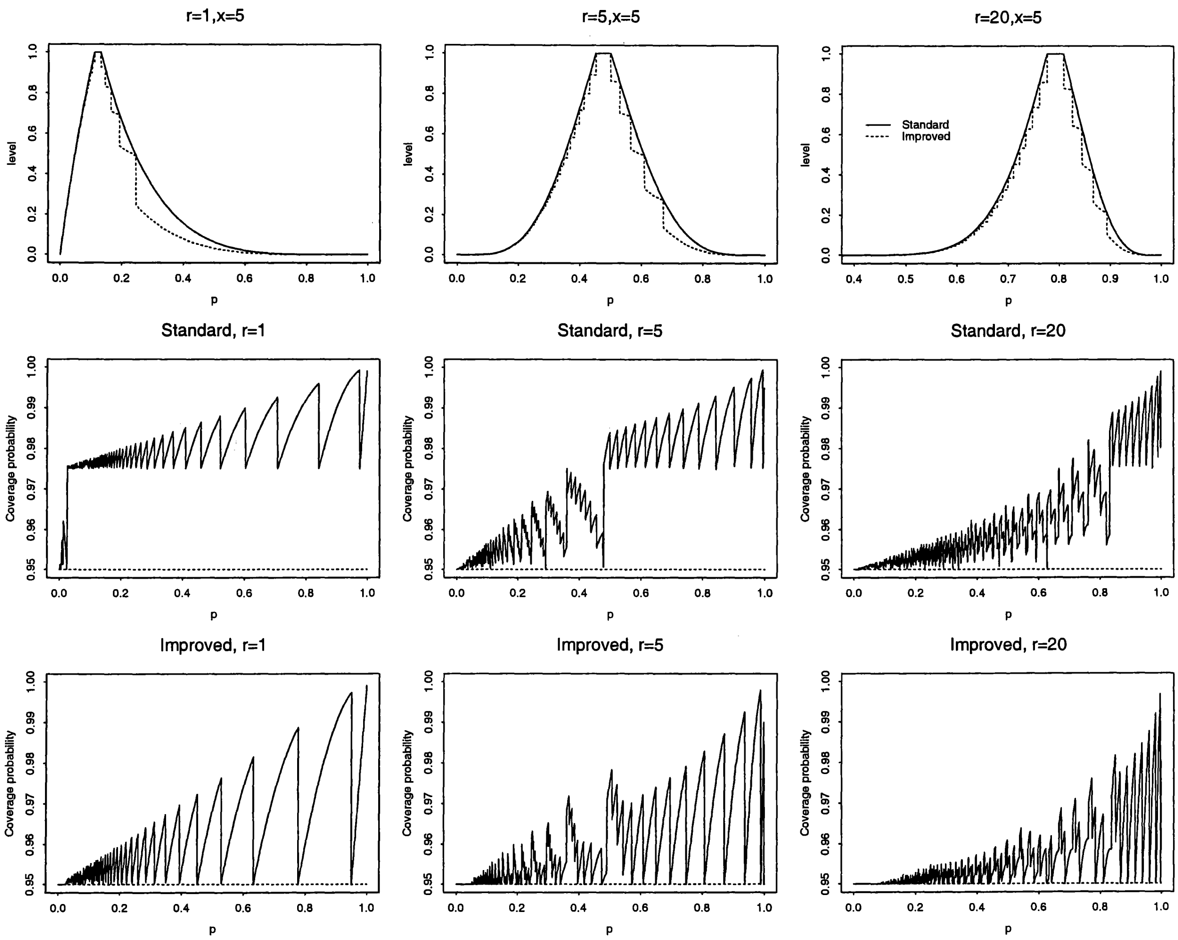

Figure 9: Top:Confidence curves for the negative binomial distribution based on standard approach and acceptability. Middle:Coverage probability for standard 95% confidence interval. Bottom: Coverage probability for 95% confidence interval based on acceptability.

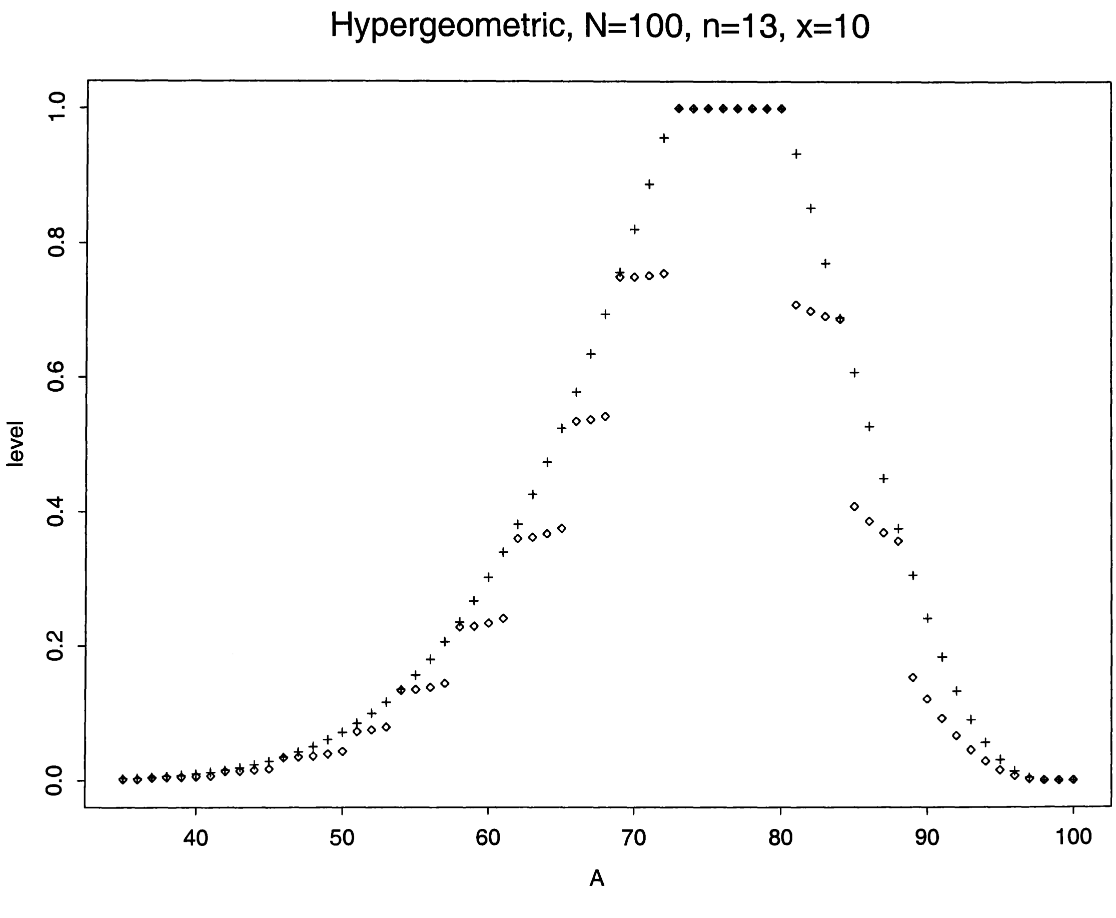

Figure 10: Confidence curves for A in the hypergeometric distribution, crosses are $\beta(A;x)$, squares are $\alpha(A;x)$.

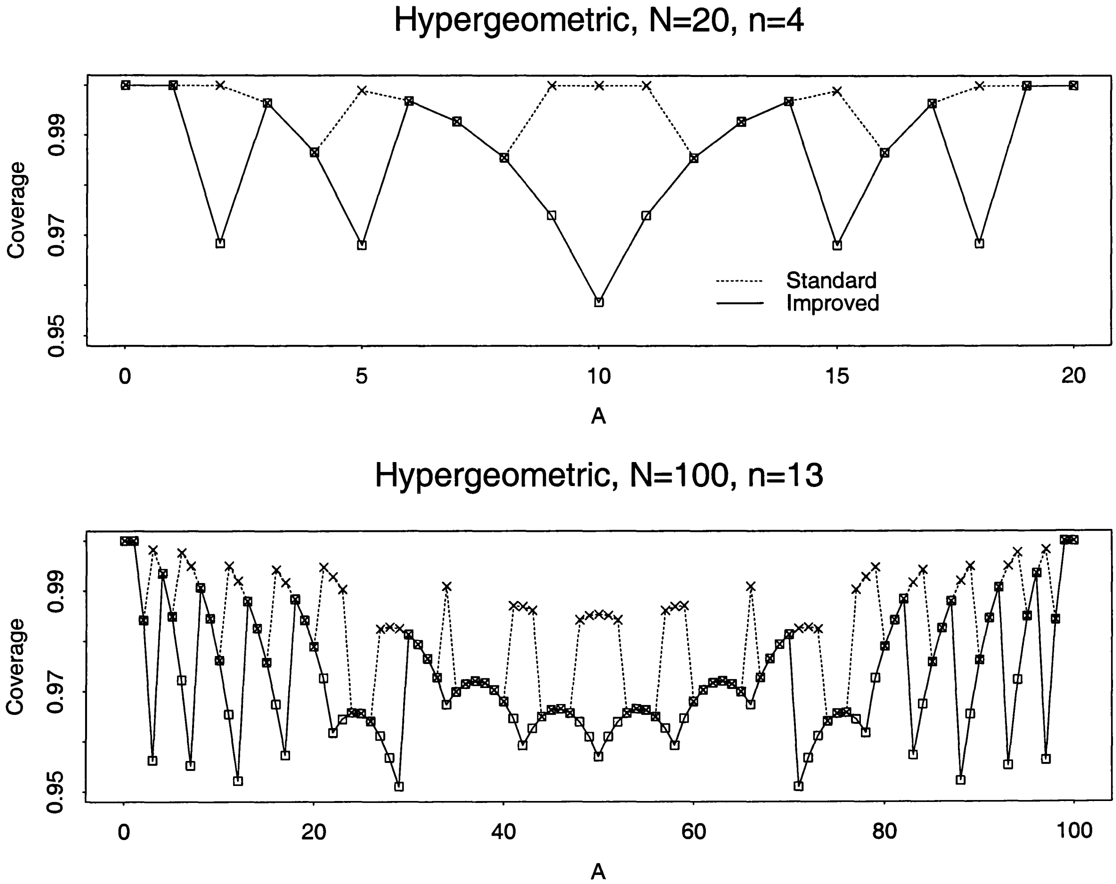

Figure 11: Coverage probabilities for standard and improved intervals. Standard interval is marked by cross, improved interval by box.

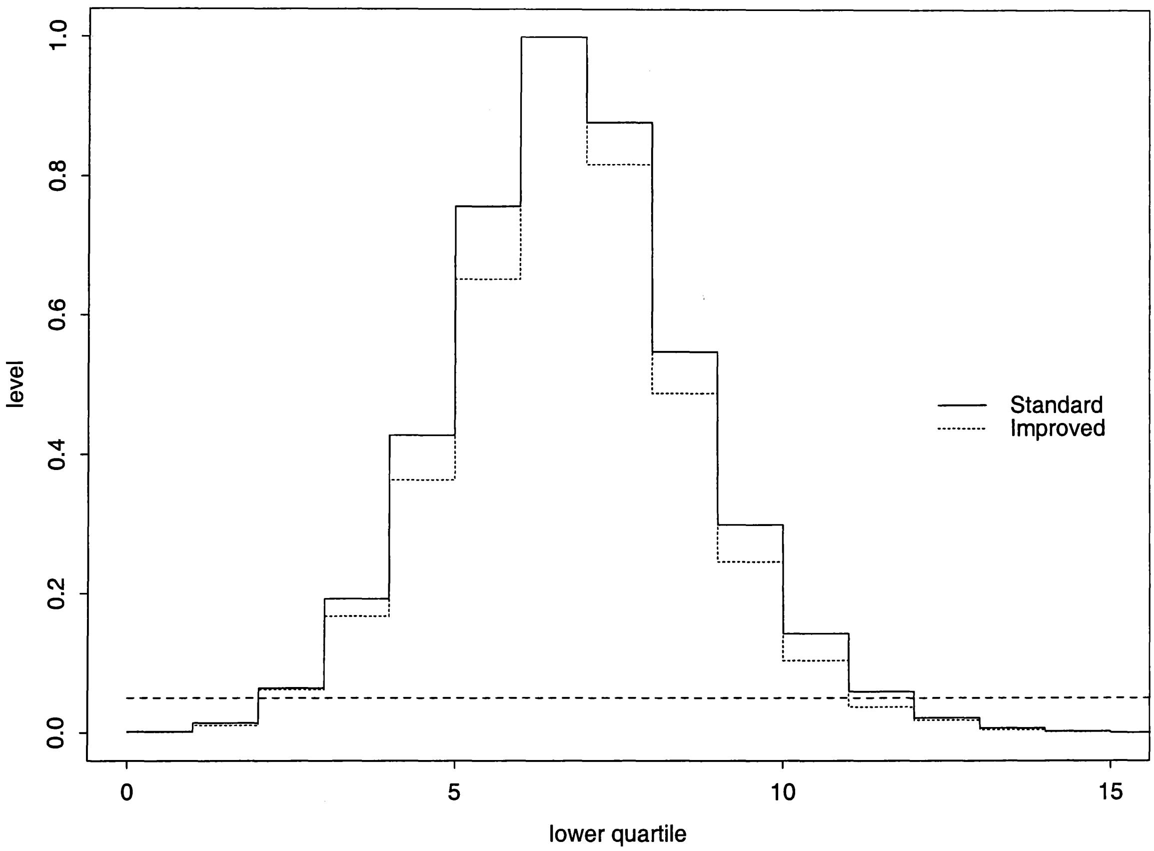

Figure 12: Nonparametric confidence curves for the lower quartile if $X_1, \ldots, X_{25} = 1, \ldots, 25$ is observed. The horizontal line is $\alpha = .05$.