Confidence Bounds for a Parameter

V. BENTKUS
Institute of Mathematics and Informatics
Akademijos 4, 232600 Vilnius, Lithuania
bentkus@takas.lt

G. PAP
Institute of Mathematics and Informatics, University of Debrecen
P.O. Box 12, H-4010 Debrecen, Hungary
papgy@math.klte.hu

M. VAN ZUIJLEN
Department of Mathematics, University of Nijmegen
Postbus 9010, 6500 GL Nijmegen, The Netherlands
zuijlen@sci.kun.nl

Abstract—The subject of the paper—upper confidence bounds—originates from applications to auditing. Auditors are interested in upper confidence bounds for an unknown mean $\mu$, for all sample sizes $n$. The samples are drawn from populations such that often only a few observations are nonzero. The conditional distribution of an observation, given that it is nonzero, usually has a very irregular shape. In such situations parametric models seem to be somewhat unrealistic. In this paper, we consider confidence bounds and intervals for an unknown parameter in parametric and nonparametric models. We propose a reduction of the problem to inequalities for tail probabilities of relevant statistics. In the special case of an unknown mean and bounded observations, a similar approach has been used in [1] by applying Hoeffding [2] inequalities for sample means and variances. The bounds can be modified in order to involve a priori information (= professional judgment of an auditor), which leads to improvements of the bounds. Furthermore, the results hold for various sampling schemes and observations from measurable spaces provided that we possess the aforementioned inequalities for tail probabilities. © 2003 Elsevier Ltd. All rights reserved.

Keywords—Nonparametric model, Conservative, Lower and upper confidence bounds, Sample variance, Auditing, Stringer bound, Hoeffding inequalities, Finite population, Sampling without replacement, Martingale.

1. INTRODUCTION AND RESULTS

Let $X_1, \ldots, X_n$ be random observations with a common distribution function (or distribution) $F = L(X_1)$. Let $\theta = \theta(F)$, $\theta \in \mathbb{R}$, be an unknown parameter. Given a risk $0 < \alpha < 1$, we are interested in the construction of upper confidence bounds for $\theta$, that is, of statistics $b = b(X_1, \ldots, X_n)$ such that

$$\mathbb{P}\{\theta \leq b\} \geq 1 - \alpha. \quad (1.1)$$

Research partly supported by the Limperg Institute, which is the Interuniversity Research Institute for Accountancy in the Netherlands, and by the Netherlands Organization for Scientific Research (NWO), and by Max Planck Institute of Mathematics, Bonn, and by the Hungarian Scientific Research Fund, Grant No. OTKA-T032361/2000.

0895-7177/03/$-$ see front matter © 2003 Elsevier Ltd. All rights reserved. Typeset by AMSTeX
do: 10.1016/S0895-7177(03)00274-7
Often bounds $b$ are considered such that (1.1) holds approximately or asymptotically as $n \to \infty$. We shall consider only confidence bounds such that the inequality in (1.1) holds for all $n$ and all allowable distributions $F$ of the observations. We shall provide lower confidence bounds (for $\theta$) as well and, as a consequence, also confidence intervals.

Our motivation comes from auditing, where inequality (1.1) has to hold for all $n$ and all allowable distributions; "approximately" is insufficient. See a review paper [3] and references therein, as well as [4]. Finkelstein et al. [5] considered bounds of such kind in some parametric models.

We cannot use methods based on limit theorems since such methods usually provide only asymptotic upper confidence bounds, so that for finite $n$ the inequality in (1.1) holds only approximately. Moreover, we would like to cover cases of different limiting behavior, for example, of Poisson or normal type. Applications of bootstrap are questionable much since we would like to cover situations where only a few of $X_j$ are nonzero, so that the empirical distribution may have quite a pure structure like in the extremal cases when all $X_j$ are zero or only one of them is nonzero. Furthermore, usually bootstrap works well when normal approximations are possible.

We have found nothing better than to apply inequalities for tail probabilities which have to hold without special assumptions on the structure of distributions. The results depend on quality of such inequalities. For sample means and some other statistics, the best available inequalities of such flavor have been proved almost 40 years ago by Hoeffding (see [2] and a review therein). Since then, no improvements suitable for construction of confidence bounds have been obtained (except parametric models, where such inequalities are obvious). There is only a sole exception—a paper of Talagrand [6] where some missing factors are implemented in these inequalities. However, one cannot use this result since the factors contain certain inexplicit constants.

We shall assume that a nonempty set, say $\Theta \subseteq \mathbb{R}$, of possible values of $\theta$ is given. We write

$$\theta_l = \inf\{\theta \in \Theta\} \quad \text{and} \quad \theta_r = \sup\{\theta \in \Theta\}.$$ 

If the set $\Theta$ is not bounded, then $\theta_l$ or/and $\theta_r$ can be infinite. With the set $\Theta$ we associate the interval $[\theta_l, \theta_r]$, which as an endpoint can eventually have $\pm \infty$.

Furthermore, we assume that a (nonempty) class $\mathcal{D}$ of allowable distributions $F \in \mathcal{D}$ is given, and that $\theta = \theta(F)$ depends on $F$. One may interpret $\mathcal{D}$ and $\Theta$ as a priori information or, in auditing, as a professional judgment of an auditor. We split the class $\mathcal{D}$ into a union of disjoint subclasses $\mathcal{D}_\theta$; that is,

$$\mathcal{D} = \bigcup_{\theta \in \Theta} \mathcal{D}_\theta, \quad \mathcal{D}_\theta = \{F \in \mathcal{D} : \theta(F) = \theta\}.$$ 

We allow the situation where some of the classes $\mathcal{D}_\theta$ are empty.

Naturally, to any parametric model there corresponds a class $\mathcal{D}$, and in this case each subclass $\mathcal{D}_\theta$ contains at most one distribution. In the nonparametric settings, usually the cardinality of the subclasses $\mathcal{D}_\theta$ is infinite.

Our construction depends on a statistic, say $T = T(X_1, \ldots, X_n)$, which serves as a starting point. Any statistic $T$ leads to an upper confidence bound of type (1.1) for $\theta$. However, the quality of the upper bound depends on the choice of $T$. For example, in the case of inference concerning an unknown mean $\mu$, most probably the best choice of $T$ is the sample mean $T = \bar{X} = (X_1 + \cdots + X_n)/n$, since $\bar{X}$ is a natural estimator of $\mu$. In the case of $\mu$, the choice of $\bar{X}$ is preferable as well from the mathematical point of view since the properties of $\bar{X}$ have been extensively studied. Of course, one could choose any other appropriate statistic. In general, it is not necessary to look for a $T$, which is an estimator of $\theta$.

We note, however, that the definition of $T$ as of a function of observations $X_1, \ldots, X_n$ is in fact superfluous. For the construction of confidence bounds we need only a set $\Theta \subseteq \mathbb{R}$ and a
family, say $T$, of random variables $T \in \mathcal{T}$ such that $\theta = \theta(T) = \theta(L(T))$ is a function of the distribution $L(T)$. We can represent

$$\mathcal{T} = \bigcup_{\theta \in \Theta} \mathcal{T}_\theta, \quad \mathcal{T}_\theta = \{T \in \mathcal{T} : \theta(T) = \theta\},$$

as a union of disjoint subclasses $\mathcal{T}_\theta$. In particular, in the special case where we observe $X_1, \ldots, X_n$ and have a statistic $T = T(X_1, \ldots, X_n)$, it suffices to assume that $X_j$ takes its values in a measurable space $(X_j, \mathcal{B}_j)$, and that $T = T(\cdot, \cdot, \ldots)$ is a measurable function of its arguments. The parameter $\theta$ has to be just a functional of the distribution of $T$.

We write

$$\mathcal{V}_\theta = \{L(T) : T \in \mathcal{T}_\theta\}.$$

We can summarize the setup as follows. We observe a realization of the random variable $T$. We have a parameter $\theta$, which is related to the distribution of $T$. We are interested in a construction of nonrandom functions $b$ (or "upper confidence bounds") such that $P\{\theta \leq b(T)\} \geq 1 - \alpha$. To optimize the construction, we would like to use a priori information in the form that the distribution of $T$ belongs to a given class of distributions or that we know that $\theta$ belongs to some given set $\Theta$. We propose a reduction of this problem to bounds of tail probabilities of $T$.

In cases where $T$ is a sum of real-valued random variables, this is usually called "inequalities for tail probabilities". In more sophisticated terminology related to nonlinear statistics the name "measure concentration phenomena" is used; see [7]. We have to remark here that for construction of upper confidence bounds, the inequalities for tail probabilities have to depend explicitly on $\theta$ and have to be as precise as possible.

Let us describe the construction of upper confidence bounds using inequalities for tail probabilities $P\{T \leq x\}$. Let $V(x, \theta)$ be a nonnegative real-valued function of variables $x$ and $\theta$ such that

$$\sup_{T \in \mathcal{T}_\theta} P\{T \leq x\} \leq V(x, \theta), \quad \text{for all } x \in \mathbb{R} \text{ and } \theta \in \Theta. \quad (1.2)$$

Without loss of generality, we assume that $V(x, \theta) \leq 1$ since otherwise we can replace $V$ in $(1.2)$ by the function $\min\{V(x, \theta); 1\}$.

In the case where $T$ is constructed from the observations $X_1, \ldots, X_n$, the function $V$ usually depends on $n$, which is not reflected in the notation. Inequality $(1.2)$ can be interpreted as an upper bound for tail probabilities of $T$.

Let fix the terminology related to monotone functions. We say that a function $f$ is increasing if $f(x) \leq f(y)$, for $x \leq y$. Similarly, we understand decreasing functions. In cases when $f(x) < f(y)$, for $x < y$, we say "strictly increasing".

For $x \in \mathbb{R}$ and $\theta \in \Theta$, consider the following conditions on the function $V$:

- the function $x \mapsto V(x, \theta)$ is increasing; \hspace{1cm} (1.3)
- the function $\theta \mapsto V(x, \theta)$ is decreasing. \hspace{1cm} (1.4)

Assumption (1.3) is quite natural since the function $x \mapsto P\{T \leq x\}$ is a distribution function. It is natural to expect that the function in (1.4) is monotone. Usually (1.4) holds in cases where $\theta$ is a "location" parameter, like the mean, but is fulfilled also for some other type of parameters. If, alternatively, the function in (1.4) is increasing, then (1.2) can be used to construct lower confidence bounds; see the text below.

For $x \in \mathbb{R}$, introduce the function

$$b(x) = \sup\{\theta : \theta \in \Theta \text{ and } V(x, \theta) > \alpha\}. \quad (1.5)$$

We define $\sup\{\theta : \theta \in \Theta\} = \theta_l$. Hence, the function $b$ can assume the values $\pm \infty$. 

THEOREM 1.1. Let a set $\Theta$, a family $\mathcal{T}$, and a function $V$ satisfy (1.2)-(1.4). Let $b$ be the function defined by (1.5). Then, the statistic $b(T)$ is an upper confidence bound for $\theta$ with risk $\alpha$; that is,

$$P\{\theta \leq b(T)\} \geq 1 - \alpha, \quad \text{for all } \theta \in \Theta \text{ and } T \in \mathcal{T}. \quad (1.6)$$

We provide the proofs of all the theorems in Section 4.

Theorem 1.1 reduces the construction of upper confidence bounds for $\theta$ to a construction of upper bounds for tail probabilities of related statistics. That is, for the class $\mathcal{F}$ of random variables (or for the corresponding class of their distributions), one has to seek for explicitly given functions $V(\cdot, \cdot)$ such that inequality (1.2) holds and is as tight as possible. In parametric models, that is, in cases where each class $\mathcal{D}_\theta$ contains at most one distribution, if we take $V(x, \theta) = IP\{T \leq x\}$, for $T \in \mathcal{T}_\theta$, the upper confidence bound $b$ is exact. In semiparametric and nonparametric models this bound will be rather conservative even in cases when we know the function

$$V^*(x, \theta) = \sup_{T \in \mathcal{T}_\theta} P\{T \leq x\}.$$ 

To improve the bounds one has to introduce additional parameters and confidence bounds for them. This leads to improvements of the upper bound (1.2) for tail probabilities. In the case of mean approach based on the variance analysis was proposed in [1]. The results of the present paper make just a necessary initial step in this direction.

In tax examinations and some other areas (see [3] for a description), lower confidence bounds for $\theta$ are of interest. This is a statistic $d$ such that $P\{d \leq \theta\} \geq 1 - \alpha$. Similar to the case of upper confidence bounds, we shall construct nonrandom functions, say $l$, such that $P\{l(T) \leq \theta\} \geq 1 - \alpha$. Using lower and upper confidence bounds for $\theta$, say $l(T)$ and $b(T)$, with risk $\alpha$ and $\beta$, respectively, we can construct a two-sided confidence interval for $\theta$ with risk $\alpha + \beta$; that is, we have

$$P\{l(T) \leq \theta \leq b(T)\} \geq 1 - (\alpha + \beta).$$

Let us start with lower confidence bounds based on inequalities of type (1.2). A bit later we provide upper and lower confidence bounds which are defined using inequalities for tail probabilities $P\{T \geq x\}$.

For $x \in \mathbb{R}$ and $\theta \in \Theta$, consider the following condition on the function $V$ from (1.2):

the function $\theta \mapsto V(x, \theta)$ is increasing; \hfill (1.7)

cf. condition (1.4) where it is assumed that the same function is decreasing. Applying condition (1.7) instead of (1.4) and using definition (1.8) instead of (1.5), we arrive at lower confidence bounds.

Introduce the function

$$l(x) = \inf \{\theta : \theta \in \Theta \text{ and } V(x, \theta) > \alpha\}, \quad x \in \mathbb{R}. \quad (1.8)$$

We define $\inf \{\theta : \theta \in \Theta\} = \theta_\ast$. Hence, the function $l$ can assume values $\pm \infty$.

THEOREM 1.2. Let a set $\Theta$, a family $\mathcal{T}$, and a function $V$ satisfy (1.2), (1.3), and (1.7). Let $l$ be the function defined by (1.8). Then, the statistic $l(T)$ is a lower confidence bound for $\theta$ with risk $\alpha$; that is,

$$P\{l(T) < \theta\} > 1 - \alpha, \quad \text{for all } \theta \in \Theta \text{ and } T \in \mathcal{T}. \quad (1.9)$$

In Theorems 1.1 and 1.2 the construction of upper and lower confidence bounds was based on upper bounds for the distribution function $x \mapsto P\{T \leq x\}$. The survival function $x \mapsto P\{T \geq x\}$ can be used as well. Let us provide details. Let $V(x, \theta)$ be a nonnegative real-valued function of variables $x \in \mathbb{R}$ and $\theta \in \Theta$ such that

$$\sup_{T \in \mathcal{T}_\theta} P\{T \geq x\} \leq V(x, \theta), \quad \text{for all } x \in \mathbb{R} \text{ and } \theta \in \Theta. \quad (1.9)$$
For $x \in \mathbb{R}$ and $\theta \in \Theta$, consider the following condition on the function $V$:

\[
\text{the function } x \mapsto V(x, \theta) \text{ is decreasing.} \tag{1.10}
\]

Assumption (1.10) is quite natural since the function $x \mapsto \mathbb{P}\{T \geq x\}$ is a survival function.

**Theorem 1.3.** Let a set $\Theta$, a family $\mathcal{T}$, and a function $V$ satisfy (1.9) and (1.10). Then, we have the following.

(i) Let $V$ satisfy (1.7) and let $l$ be the function defined by (1.8). Then, the statistic $l(T)$ is a lower confidence bound for $\theta$ with risk $\alpha$.

(ii) Let $V$ satisfy (1.4) and let $b$ be the function defined by (1.5). Then, the statistic $b(T)$ is an upper confidence bound for $\theta$ with risk $\alpha$.

The choice of different statistics $T$ can lead to different confidence bounds for $\theta$. Similarly, different functions $V$ from (1.2) and (1.9) can provide different confidence bounds as well. A comparison of the bounds is not a simple question. A clear answer to the question is possible if a bound is dominated by another one. For example, if $b_1$ and $b_2$ are two different upper confidence bounds, then the bound $b_1$ is obviously better than $b_2$ if $b_1 \leq b_2$ with probability one, for all $\theta$.

**2. SOME EXAMPLES OF PARAMETRIC MODELS.**

In the case of parametric models the class $\mathcal{T}_0$ contains at most one statistic $T_0$. A natural choice of the function $V$ from (1.2) would be $V(x, \theta) = \mathbb{P}\{T_0 \leq x\}$. Then, condition (1.3) is fulfilled since the function $x \mapsto \mathbb{P}\{T_0 \leq x\}$ is a distribution function.

We consider some examples of upper confidence bounds. In a similar way, one can introduce lower confidence bounds.

In this section, $[x]$ stands for the integer part of $x \in \mathbb{R}$ and

\[
[x] = \begin{cases} 
  [x], & \text{if } x \text{ is not an integer,} \\
  [x] - 1, & \text{if } x \text{ is an integer.}
\end{cases}
\]

**Example 2.1. The Binomial Model.** Let $0 \leq p \leq 1$. Let $T$ be the number of successes in $n$ independent Bernoulli trials with success probability $p$. The statistic $T$ has the binomial distribution $\text{Bin}(n, p)$. Integrating several times by parts, we have

\[
\mathbb{P}\{T > x\} = \sum_{k=j+1}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = (j + 1) \left( \frac{n}{j+1} \right) \int_0^p t^j (1 - t)^{n-j-1} dt, \tag{2.1}
\]

for $x \geq 0$, where $j = [x]$. The integral representation (2.1) shows that the function

\[
p \mapsto \mathbb{P}\{T \leq x\} = 1 - \mathbb{P}\{T > x\}
\]

is a continuous strictly decreasing function of $p$.

Assume that $n$ is known and that we would like to construct an upper confidence bound for unknown $p$ using the observed $T$. Let $\Theta = [0, 1]$ be the parameter set. The function $V(x, p) = \mathbb{P}\{T \leq x\}$ is continuous and strictly decreasing function of $p \in [0, 1]$. Hence, condition (1.4) is fulfilled. Using definition (1.5) and representation (2.1), the function $b(x)$ (we suppress $n$ and $\alpha$ in this notation though $b(x)$ depends on $n$ and $\alpha$) is a unique solution of the equation

\[
(j + 1) \left( \frac{n}{j+1} \right) \int_0^p t^j (1 - t)^{n-j-1} dt = 1 - \alpha, \quad \text{where } j = [x], \tag{2.2}
\]

with respect to $p$ as an unknown variable (if $0 \leq x < n$; if $x = n$ then clearly $b(n) = 1$). Hence, by Theorem 1.1, the statistic $b(T)$ is an upper confidence bound for $p$; that is,

\[
\mathbb{P}\{p \leq b(T)\} \geq 1 - \alpha. \tag{2.3}
\]
It is clear that 
\[ 0 < b(0) < b(1) < \cdots < b(n - 2) < b(n - 1) < b(n) = 1. \]

Moreover, we have \( b(0) = 1 - \alpha^{1/n} \) and \( b(n - 1) = (1 - \alpha)^{1/n}. \)

Assume now that we know in addition that \( a \leq p \leq d \) with some \( 0 \leq a \leq d \leq 1. \) Then, the parameter set \( \Theta = [a, d] \) and in this case the bound provided by (2.3) can be improved. Now, the function
\[
 b_{a,d}(x) = \begin{cases} 
 a, & \text{if } b(x) \leq a, \\
 b(x), & \text{if } a \leq b(x) \leq d, \\
 d, & \text{if } b(x) \geq d,
\end{cases}
\]

provides the upper confidence bound \( b_{a,d}(T) \) for \( p \) with risk \( \alpha. \)

Now, suppose that \( p \) is known, and we seek an upper confidence bound for unknown \( n \) using the observed \( T_n = T \) (in this case it is convenient to indicate the dependence of \( T \) on the parameter \( n \)).

The parameter set \( \Theta = \{1, 2, 3, \ldots\} \). The function \( V(x, n) = P\{T_n \leq x\} \) is a decreasing function of the parameter \( n \in \{1, 2, 3, \ldots\} \). Indeed, introduce a sequence \( \varepsilon_1, \varepsilon_2, \ldots \) of i.i.d. Bernoulli random variables with the success probability \( p \). Then, \( P\{T_n \leq x\} = P\{\varepsilon_1 + \cdots + \varepsilon_n \leq x\}. \)

Using that the variable \( \varepsilon_{n+1} \) is nonnegative, we have
\[
 V(x, n + 1) = P\{T_{n+1} < x\} = P\{T_n + \varepsilon_{n+1} < x\} = P\{T_n < x\} = V(x, n).
\]

Hence, condition (1.4) is fulfilled. Definition (1.5) combined with (2.1) shows that \( b(x) \) is a maximal \( n \in \Theta \) (if there is any; otherwise \( b(x) = 1 \)) which satisfies the inequality
\[
 (j + 1) \binom{n}{j + 1} \int_0^1 t^j (1 - t)^{n-j-1} dt > \alpha, \quad \text{where } j = \lfloor x \rfloor.
\]

By Theorem 1.1, the statistic \( b(T_n) \) is an upper confidence bound for \( n \) with risk \( \alpha \), that is, \( P\{n \leq b(T_n)\} \geq 1 - \alpha \). It is clear that
\[
 b(0) = \max \left\{ 1; \left[ \frac{\ln \alpha}{\ln(1 - p)} \right] \right\}, \quad (2.4)
\]

Now suppose that we allow the value \( n = 0 \) as well. Then, instead of the parameter \( \Theta = \{1, 2, 3, \ldots\} \) we will have \( \Theta = \{0, 1, 2, \ldots\} \). This will change the confidence bound—now,
\[
 b(0) = \left[ \frac{\ln \alpha}{\ln(1 - p)} \right],
\]

cf. (2.4) (to be definite we define \( T = 0 \) if \( n = 0 \)).

The choice of the statistic \( T \) is irrelevant to estimation of parameters. However, the quality of upper bounds depends on the choice of \( T \). Consider, for example, the statistic \( T = 0 \), if in a sequence of \( n \) Bernoulli trials there are no successes, and \( T = 1 \) otherwise. In the case of unknown \( p \), this statistic leads to the rather rough upper confidence bound for \( p : b(T) = 1 - \alpha^{1/n}, \) if \( T = 0 \), and \( b(T) = 1, \) if \( T = 1 \).

The exposed above construction of confidence bounds for \( p \) in the binomial model is well known; see, for example, [8].

**Example 2.2. The Poisson Model.** Let \( 0 < \lambda < \infty \). Let \( X_1, \ldots, X_n \) be independent Poisson observations such that
\[
 P\{X_j = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{for } k \in \mathbb{N}_0 \text{ and } 1 \leq j \leq n.
\]
The statistic $T = X_1 + \cdots + X_n$ has the Poisson distribution $\text{Po}(n\lambda)$. Integrating by parts, we have
\[
P\{T > x\} = \sum_{k=j+1}^{\infty} \frac{(n\lambda)^k}{k!} e^{-n\lambda} = \int_0^{n\lambda} \frac{u^j}{j!} e^{-u} \, du, \quad j = [x],
\]
for $0 \leq x < \infty$.

Assume that $n$ is known, $\lambda$ is unknown, and that we observe $T$. Now, $\Theta = (0, \infty)$. It follows from (2.5) that the function $V(x, \lambda) = P\{T \leq x\}$ is a continuous strictly decreasing function of $\lambda$. By Theorem 1.1, a unique solution, say $b(x)$, of the equation
\[
\int_{n\lambda}^{\infty} \frac{u^j}{j!} e^{-u} \, du = \alpha, \quad j = [x],
\]
with respect to $\lambda$ as an unknown variable, provides an upper confidence bound for $\lambda$, that is, $P\{b(T) \geq \lambda\} \geq 1 - \alpha$. Notice that $b(0) = (1/n) \ln(1/\alpha)$.

Assume that $n$ is unknown, $\lambda$ is known, that we observe $T$, and that $\Theta = \{1, 2, \ldots\}$. It follows from (2.5) that the function $V(x, n) = P\{T \leq x\}$ is a continuous strictly decreasing function of $n$. By Theorem 1.1, a unique maximal solution, say $b(x)$, of the inequality
\[
\int_{n\lambda}^{\infty} \frac{u^j}{j!} e^{-u} \, du > \alpha, \quad j = [x],
\]
with respect to $n$ as an unknown variable, provides an upper confidence bound for $n$ with risk $\alpha$; that is, $P\{b(T) \geq n\} \geq 1 - \alpha$ (if the inequality has no solutions in $n$ then $b(x) = 1$). Notice that $b(0) = \max\{1; [(1/\lambda) \ln(1/\alpha)]_x\}$.

**EXAMPLE 2.3. THE SCALE IN A STABLE MODEL.** Let $Z$ be a stable nonnegative random variable with the exponent $0 < \gamma < 1$ and positive density $p : (0, \infty) \rightarrow (0, \infty)$. Let $0 < s < \infty$ be a scale parameter. Assume that we observe $X_1 = sZ_1, \ldots, X_n = sZ_n$, where $Z_j$ are independent and have the distribution of $Z$. The statistic
\[
T = n^{-1/\gamma}(X_1 + \cdots + X_n)
\]
has the distribution of the random variable $sZ$. Hence, we have
\[
P\{T \leq x\} = \int_0^{x/s} p(u) \, du, \quad (2.6)
\]
for $0 < x < \infty$. Notice that the distribution of $T$ is independent of $n$.

Assume that $n$ is known, $s$ is unknown, and that we observe $T$. Now, $\Theta = (0, \infty)$. It follows from (2.6) that the function $V(x, s) = P\{T \leq x\}$ is a continuous strictly decreasing function of $s \in (0, \infty)$. Let $I(x/s)$ stand for the integral in (2.6). By Theorem 1.1, a unique solution, say $b(x)$, of the equation $I(x/s) = \alpha$ with respect to $s$ as an unknown variable provides an upper confidence bound for $s$, that is, $P\{b(T) \geq s\} \geq 1 - \alpha$.

Assume that $n$ is unknown, $s$ is known, and that we observe $T$. Now, $\Theta = \{1, 2, \ldots\}$. The function $V(x, n) = P\{T \leq x\}$ is independent of the variable $n$, and definition (1.5) yields $b(x) = \infty$, if $x$ satisfies $I(x/s) > \alpha$, and $b(x) = 1$, if $x$ satisfies $I(x/s) \leq \alpha$. By Theorem 1.1,
\[
b(T) = \begin{cases} 
\infty, & \text{if } I\left(\frac{T}{s}\right) > \alpha, \\
1, & \text{if } I\left(\frac{T}{s}\right) \leq \alpha,
\end{cases}
\]
is an upper confidence bound for $n$ with risk $\alpha$. 
Example 2.4. The Hypergeometric Model. In this model, we have a population \( x_1, \ldots, x_N \), such that \( x_j = 0 \), for \( 1 \leq j \leq N - M \), and \( x_j = 1 \), for \( N - M < j < N \).

We draw \( X_1 \) with equal probability and without replacement. We continue the procedure obtaining \( X_2, \ldots, X_n \). The statistic \( T = X_1 + \cdots + X_n \) has the hypergeometric distribution, that is,

\[
q_j \overset{\text{def}}{=} \mathbb{P}(T = j) = \binom{M}{j} \binom{N-M}{n-j} \binom{n}{j}, \quad j = 0, \ldots, n,
\]

for \( n \leq M \), \( n \leq N - M \), and \( \mathbb{P}(T > x) = q_{j+1} + \cdots + q_n \) with \( j = \lfloor x \rfloor \).

The probability \( \mathbb{P}(T \leq x) \) is a decreasing function of \( n \), for fixed \( N \) and \( M \). For fixed \( n \) and \( N \), the probability is a decreasing function of \( M \) as well. For fixed \( n \) and \( M \), the probability is an increasing function of \( N \). Hence, we can apply Theorems 1-3 to construct confidence bounds for any of parameters \( n, N, M \). One can enrich the model as in the Bernoulli case using the monotonicity properties and a priori information in the form of restrictions on two of three parameters \( n, N, M \).

3. Some Examples of Nonparametric Models

Confidence Bounds for the Mean

Let \( D \) be the class of all possible distributions supported by the interval \([0, 1] \). Let \( X_1, \ldots, X_n \) be i.i.d. observations with a distribution \( F \) from the class \( D \). Write \( \mu \) and \( \sigma^2 \) for the mean and variance of \( X_1 \), respectively. Let

\[
T = \frac{X_1 + \cdots + X_n}{n}
\]

be the sample mean. The simplest Hoeffding inequality says that (see Theorem 1 in [2])

\[
\mathbb{P}(T \leq x) \leq H^n(1 - x; 1 - \mu), \quad \mathbb{P}(T \geq x) \leq H^n(x; \mu), \quad (3.1)
\]

where, for \( 0 \leq \nu \leq 1 \),

\[
H(a; \nu) = \left( \frac{1 - \nu}{1 - a} \right)^{1 - a} \left( \frac{\nu}{a} \right)^a, \quad \text{for } \nu < a \leq 1, \quad (3.2)
\]

and \( H(a; \nu) = 1 \), for \( a \leq \nu \), \( H(a; \nu) = 0 \), for \( a > 1 \). The function \( H \) allows the following very rough upper bound (see [2], Theorem 1):

\[
H(\nu + t; \nu) \leq \exp \left\{ -2t^2 \right\}, \quad \text{for } t \geq 0. \quad (3.3)
\]

Since \( H(a; \nu) \) is decreasing in \( a \) and increasing in \( \nu \), the upper bounds (3.1) can be used to construct confidence bounds for \( \mu \). By (i) of Theorem 1.3, the statistic \( l(T) \) defined using the function

\[
l(x) = \min \left\{ \mu : 0 \leq \mu \leq 1 \text{ and } H(x; \mu) \geq \alpha^{1/n} \right\} \quad (3.4)
\]

is a lower confidence bound for \( \mu \) with risk \( \alpha \). By Theorem 1.1, the statistic \( b(T) \) defined using the function

\[
b(x) = \max \left\{ \mu : 0 \leq \mu \leq 1 \text{ and } H(1 - x; 1 - \mu) \geq \alpha^{1/n} \right\} \quad (3.5)
\]

is an upper confidence bound for \( \mu \) with risk \( \alpha \). However, the bounds coming from (3.4) and (3.5) being simple in the form are rather rough; see [1, Proposition 1.1]. A more sophisticated application of Hoeffding’s inequalities which involves variances allows us to obtain much better confidence bounds; see [1].
Confidence Bounds for the Variance

The construction is based on applications of Hoeffding's inequalities. These inequalities hold for a wide class of statistics, which allows us to extend the results to other examples. We follow the approach used in [1], leaving aside generalizations of the most sophisticated bounds.

Let \( D \) be the class of distributions \( F \) supported by the interval \([0, 1]\) (in what comes below, one can replace the class \( D \) by the wider class of all possible shifts of distributions from \( D \)). Let \( X_1, \ldots, X_n \) be independent observations taken from a distribution \( F \in D \), and let \( \sigma^2 \) be the variance of \( X_1 \). As a basic statistic we shall use the sample variance

\[
\hat{\sigma}^2 = \frac{1}{n(n-1)} \sum_{i<k} (X_i - X_k)^2. \tag{3.6}
\]

To construct confidence bounds for \( \sigma^2 \), we shall use the following inequalities for tail probabilities (cf. Lemma 3.2 in [1]).

**Lemma 3.1.** The sample variance satisfies

\[
P\{\hat{\sigma}^2 \leq x\} \leq H^{[n/2]}(1 - 2x; 1 - 2\sigma^2), \quad \text{for all } x \leq \sigma^2, \tag{3.7}
\]

and

\[
P\{\hat{\sigma}^2 \geq x\} \leq H^{[n/2]}(2x; 2\sigma^2), \quad \text{for all } x \geq \sigma^2. \tag{3.8}
\]

**Proof.** Consider the following U-statistic:

\[
U = \frac{1}{n(n-1)} \sum_{1 \leq i < k \leq n} g(X_i, X_k), \tag{3.9}
\]

with a function \( g \) such that \( 0 \leq g \leq 1 \), where the sum in (3.9) is taken over all possible pairs \( i, k \) of distinct \( i \) and \( k \) (cf. (3.6) where the sum is taken over \( 1 \leq j < k \leq n \)). Then, for \( t \geq 0 \), we have (see comments following (5.7) in [2])

\[
P\{U > E(U) + t\} \leq H^{[n/2]}(E(U) + t; E(U)). \tag{3.10}
\]

Let us prove (3.7). It is clear that

\[
P\{\hat{\sigma}^2 \leq x\} = P\{1 - 2\hat{\sigma}^2 \geq 1 - 2x\}.
\]

The statistic \( 1 - 2\hat{\sigma}^2 \) is a statistic of the form (3.9) with \( g(x, y) = 1 - (x - y)^2 \). Hence, we may apply (3.10), and (3.7) follows.

We omit the proof of (3.8) since it is similar to the proof of (3.7). ■

By (i) of Theorem 1.3, the statistic \( l(T) \) defined using the function

\[
l(x) = \min \left\{ \sigma^2 : 0 \leq \sigma^2 \leq \frac{1}{4} \text{ and } H(2x; 2\sigma^2) \geq \alpha^{1/[n/2]} \right\} \tag{3.11}
\]

is a lower confidence bound for \( \sigma^2 \) with risk \( \alpha \). By Theorem 1.1, the statistic \( b(T) \) defined using the function

\[
b(x) = \max \left\{ \sigma^2 : 0 \leq \sigma^2 \leq \frac{1}{4} \text{ and } H(1 - 2x; 1 - 2\sigma^2) \geq \alpha^{1/[n/2]} \right\} \tag{3.12}
\]

is an upper confidence bound for \( \sigma^2 \) with risk \( \alpha \).

The bounds (3.11) and (3.12) are relatively rough. Following the approach of [1], it is possible to construct much better bounds which can include a priori information. However, it will require analysis of statistics related to the fourth moment of observations.
4. PROOFS

In this section, we prove the theorems of the introduction.

PROOF OF THEOREM 1.1. Fix \( \theta \in \Theta \). Let \( T \) be a statistic from the class \( T_\theta \). Instead of (1.6) it suffices to show that

\[
\mathbb{P}\{b(T) < \theta\} \leq \alpha.
\]

(4.1)

Assumption (1.3) and definition (1.5) of \( b \) show that the function \( x \mapsto b(x) \) is increasing. Hence, the set

\[
A = \{ x \in \mathbb{R} : b(x) < \theta \}
\]

is empty or an interval. Therefore, there exists \( x_\theta \in \mathbb{R} \) such that only one of the following four cases is possible:

(i) \( A = \emptyset \);
(ii) \( A = (-\infty, x_\theta] \);
(iii) \( A = (-\infty, x_\theta) \);
(iv) \( A = \mathbb{R} \).

(4.2)

(i) \( A = \emptyset \) yields \( b(x) \geq \theta \), for all \( x \in \mathbb{R} \), and (4.1) is obviously fulfilled, since now \( \mathbb{P}\{b(T) < \theta\} = 0 \).

(ii) We have \( b(x_\theta) < \theta \). We can rewrite definition (1.5) as

\[
b(x_\theta) = \sup \{ \rho : \rho \in \Theta \text{ and } V(x_\theta, \rho) > \alpha \}.
\]

(4.3)

Let us note that \( V(x_\theta, \theta) \leq \alpha \). Indeed, otherwise we have \( V(x_\theta, \theta) > \alpha \), and (4.3) yields \( b(x_\theta) \geq \theta \), which contradicts the inequality \( b(x_\theta) < \theta \). By condition (1.2) we have \( \mathbb{P}\{T \leq x_\theta\} \leq V(x_\theta, \theta) \), which combined with \( V(x_\theta, \theta) \leq \alpha \) implies \( \mathbb{P}\{T \leq x_\theta\} \leq \alpha \). This inequality yields (4.1) since by the definition of the set \( A \) we have

\[
\{ T \leq x_\theta \} = \{ T \in A \} = \{ b(T) < \theta \}.
\]

(iii) Let \( \varepsilon > 0 \) be an arbitrary positive number. Then, \( b(x_\theta - \varepsilon) < \theta \). Arguing as in Case (ii), we derive \( \mathbb{P}\{T \leq x_\theta - \varepsilon\} \leq \alpha \). Passing to the limit as \( \varepsilon \downarrow 0 \), we have \( \mathbb{P}\{T < x_\theta\} \leq \alpha \). In the case under consideration the definition of the set \( A \) shows that \( \{ T < x_\theta \} = \{ b(T) < \theta \} \). Hence, the inequality \( \mathbb{P}\{T < x_\theta\} \leq \alpha \) is equivalent to (4.1).

(iv) Let \( N > 0 \) be an arbitrary large positive number. Then, \( b(N) < \theta \). Arguing as in Case (ii), we derive \( \mathbb{P}\{T \leq N\} \leq \alpha \). Passing to the limit as \( N \uparrow \infty \), we have \( \mathbb{P}\{T < \infty\} \leq \alpha \), which shows that Case (iv) is impossible since \( \alpha < 1 \) and the statistic \( T \) assumes only finite values.

PROOF OF THEOREM 1.2. To prove the theorem we use a continuous decreasing one-to-one transformation \( g : \mathbb{R} \rightarrow \mathbb{R} \). Using this transformation we can introduce the new parameter \( \rho = g(\theta) \in g(\Theta) \). We can consider the family \( T \) of statistics as a family such that to each \( T \in T \) corresponds \( \rho = g(\theta) \). Then, inequality (1.2) reads as

\[
\sup_{T \in T_\theta} \mathbb{P}\{T \leq x\} \leq V(x, g^{-1}(\rho)) = V(x, g^{-1}(\rho)).
\]

(4.4)

The function \( \rho \mapsto V(x, g^{-1}(\rho)) \) is a decreasing function by assumption (1.7) and since the function \( g^{-1} \) is strictly decreasing. Hence, (4.4) allows us to apply Theorem 1.1, and

\[
b(x) = \sup \{ \rho : \rho \in g(\Theta) \text{ and } V(x, g^{-1}(\rho)) > \alpha \}.
\]

(4.5)

is an upper confidence bound for \( \rho \), that is, \( \mathbb{P}\{\rho \leq b(T)\} \geq 1 - \alpha \). Using \( \rho = g(\theta) \) and that the inverse function \( g^{-1} \) is continuous and strictly decreasing, we can rewrite (4.5) as

\[
b(x) = \sup \{ g(\theta) : g(\theta) \in g(\Theta) \text{ and } V(x, \theta) > \alpha \} = g(\inf \{ \theta : \theta \in \Theta \text{ and } V(x, \theta) > \alpha \})
\]

(4.6)
Using $P\{\rho \leq b(T)\} \geq 1 - \alpha$, the equality $\rho = g(\theta)$, and (4.6), we have

$$1 - \alpha \leq P\{g(\theta) \leq b(T)\} = P\{\theta \geq l(T)\},$$

proving the theorem.

**Proof of Theorem 1.3.** This proof is similar to that of Theorems 1.1 and 1.2.

**References**