# Estimation of a probability with optimum guaranteed confidence in inverse binomial sampling

LUIS MENDO[*] and JOSÉ M. HERNANDO[**]

*E.T.S. Ingenieros de Telecomunicación, Polytechnic University of Madrid, 28040 Madrid, Spain.*
*E-mails: [*]lmendo@grc.ssr.upm.es; [**]hernando@grc.ssr.upm.es*

Sequential estimation of a probability $p$ by means of inverse binomial sampling is considered. For $\mu_1, \mu_2 > 1$ given, the accuracy of an estimator $\hat{p}$ is measured by the confidence level $P[p/\mu_2 \leq \hat{p} \leq p\mu_1]$. The confidence levels $c_0$ that can be guaranteed for $p$ unknown, that is, such that $P[p/\mu_2 \leq \hat{p} \leq p\mu_1] \geq c_0$ for all $p \in (0, 1)$, are investigated. It is shown that within the general class of randomized or non-randomized estimators based on inverse binomial sampling, there is a maximum $c_0$ that can be guaranteed for arbitrary $p$. A non-randomized estimator is given that achieves this maximum guaranteed confidence under mild conditions on $\mu_1, \mu_2$.

*Keywords:* confidence level; interval estimation; inverse binomial sampling; sequential estimation

## 1. Introduction

In a sequence of Bernoulli trials with probability of success $p$ at each trial, consider the estimation of $p$ by inverse binomial sampling. This sampling scheme, first discussed by Haldane (1945), consists in observing the sequence until a given number $r$ of successes is obtained. The resulting number of trials $N$ is a sufficient statistic for $p$ (Lehmann and Casella (1998), page 101). The uniformly minimum variance unbiased estimator of $p$ is (Mikulski and Smith (1976))

$$\hat{p} = \frac{r-1}{N-1} \tag{1.1}$$

and for $r \geq 3$, it has a normalized mean square error $\mathrm{E}[(\hat{p} - p)^2]/p^2$ (or $\mathrm{Var}[\hat{p}]/p^2$) lower than $1/(r-2)$, irrespective of $p$ (Mikulski and Smith (1976); Sathe (1977); Prasad and Sahai (1982)).

This paper analyzes inverse binomial sampling from a different point of view, related to interval estimation. Given $\mu_1, \mu_2 > 1$, the accuracy of an estimator $\hat{p}$ is measured by the probability $c$ that $\hat{p}$ lies in the interval $[p/\mu_2, p\mu_1]$. The motivation to use a relative interval $[p/\mu_2, p\mu_1]$, instead of an interval $[x_1, x_2]$ with $x_1, x_2$ fixed, is the fact that $P[p/\mu_2 \leq \hat{p} \leq p\mu_1]$, unlike $P[x_1 \leq \hat{p} \leq x_2]$, has a definite meaning independent of $p$. Moreover, using a relative interval allows another interpretation of $c$: since $p/\mu_2 \leq \hat{p} \leq p\mu_1$ if and only if $\hat{p}/\mu_1 \leq p \leq \hat{p}\mu_2$, $c$ gives the probability that the true value $p$ is covered by the random interval $[\hat{p}/\mu_1, \hat{p}\mu_2]$. In the sequel, $c$ will be referred to as the *confidence* (or confidence level) associated with $\mu_1, \mu_2$.

Recently, Mendo and Hernando (2006, 2008a) have shown that the estimator

$$\hat{p} = \frac{r-1}{N} \tag{1.2}$$

has the following properties for $r \geq 3$. The confidence level $c$ associated with $\mu_1$, $\mu_2$ has an asymptotic value $\bar{c}$ as $p \to 0$, namely $\bar{c} = \gamma(r, (r-1)\mu_2) - \gamma(r, (r-1)/\mu_1)$, where

$$\gamma(r, t) = \frac{1}{\Gamma(r)} \int_0^t s^{r-1} \exp(-s)\, ds$$

is the regularized incomplete gamma function. Furthermore, the confidence for any $p \in (0, 1)$ exceeds this asymptotic value provided that $\mu_1$ and $\mu_2$ satisfy certain lower bounds. Similar results have been established (Mendo and Hernando (2008b)) for the uniformly minimum variance unbiased estimator (1.1). Using a more general setting in which the Bernoulli random variables are replaced by arbitrary bounded random variables, Chen (2007) has obtained comparable (although somewhat less tight) results for (1.1), as well as for the maximum likelihood estimator

$$\hat{p} = \frac{r}{N}. \tag{1.3}$$

From the aforementioned results, the question naturally arises as to whether the attained confidence could be improved using estimators other than (1.1)–(1.3). This motivates the study of arbitrary estimators based on inverse binomial sampling. In this regard, it is noted that $1 - c$ can be expressed as the risk $\mathrm{E}[L(\hat{p})]$ corresponding to the loss function $L$ defined as

$$L(x) = \begin{cases} 0, & \text{if } x \in [p/\mu_2, p\mu_1], \\ 1, & \text{otherwise.} \end{cases} \tag{1.4}$$

Since $N$ is a sufficient statistic, for any estimator defined in terms of the observed Bernoulli random variables, there exists an estimator that depends on the observations through $N$ only and that has the same risk; this equivalent estimator is possibly a randomized one (Lehmann and Casella (1998), page 33). (The fact that $L$ is non-convex prevents application of a corollary of the Rao–Blackwell theorem (Lehmann and Casella (1998), page 48) to discard randomized estimators.) Thus, attention can be restricted to estimators that depend on the observed variables only through $N$; however, both randomized and non-randomized estimators have to be considered.

The purpose of the present paper is to investigate the limits on the confidence levels $c_0$ that can be guaranteed (in the sense that the actual confidence equals or exceeds $c_0$, irrespective of $p$) in inverse binomial sampling. More specifically, the objectives are:

- to determine the supremum of $\inf_{p \in (0,1)} c$ over all (randomized or non-randomized) estimators based on inverse binomial sampling;
- to find an estimator, if it exists, that can achieve this supremum.

Consequently, the main focus of this paper is on non-asymptotic results, valid for $p \in (0, 1)$. Nonetheless, asymptotic results for $p \to 0$ will also be derived since, apart from their own theoretical importance, they provide an upper bound on (and an indication of) what can be achieved for arbitrary $p$.

Section 2 shows that $\liminf_{p \to 0} c$ has a maximum over all estimators based on inverse binomial sampling and computes this maximum. This sets an upper bound on the confidence that can be guaranteed by an arbitrary estimator. Section 3 establishes that this upper bound is also a maximum, that is, estimators exist that can achieve this guaranteed confidence. Specifically, estimators are given that guarantee the maximum confidence for sufficiently small $p$ and that guarantee the maximum confidence for any $p$, under certain conditions on the relative interval being considered. Section 4 discusses the results, comparing them with those from other works. Section 5 contains proofs of all results in the paper.

## 2. Asymptotic analysis

It is assumed in the sequel that $r \geq 3$. Let $t^{(j)}$ denote $t(t-1)\cdots(t-j+1)$. The probability function of $N$, $f(n) = P[N = n]$, is

$$f(n) = \frac{(n-1)^{(r-1)}}{(r-1)!} p^r (1-p)^{n-r} \qquad \text{for } n \geq r. \tag{2.1}$$

As justified in Section 1, it is sufficient to consider (possibly randomized) estimators defined in terms of the sufficient statistic $N$. Let $\mathcal{F}$ denote the set of all functions from $\{r, r+1, r+2, \ldots\}$ to $\mathbb{R}$. A *non-randomized estimator* $\hat{p}$ can be described as $\hat{p} = g(N)$, with $g \in \mathcal{F}$. Thus, a non-randomized estimator is entirely specified by its function $g$. A *randomized estimator* is a random variable $\hat{p}$ whose distribution depends, in general, on the value taken by $N$. Let $\Pi_n$ denote the distribution function of $\hat{p}$ conditioned on $N = n$. The randomized estimator is completely specified by the functions $\Pi_n$, $n \geq r$. Thus, denoting by $\mathcal{F}_\mathcal{R}$ the class of all functions from $\{r, r+1, r+2, \ldots\}$ to the set of real functions of a real variable, a randomized estimator is defined by a function $G \in \mathcal{F}_\mathcal{R}$ that assigns $\Pi_n$ to each $n$.

Non-randomized estimators form a subset of the class of randomized estimators. Thus, any statement that applies to randomized estimators will also be valid, in particular, for non-randomized estimators. The reason that the specialized class of non-randomized estimators has been explicitly defined is that, on one hand, their simplicity makes them more attractive for applications and, on the other hand, it will be seen that, under certain conditions, no loss of optimality in guaranteed confidence is incurred by restricting to non-randomized estimators. Throughout the paper, when referring to an arbitrary estimator without specifying its type, the general class of randomized estimators (including the non-randomized ones) will be meant.

Given any estimator $G \in \mathcal{F}_\mathcal{R}$, the confidence associated with $\mu_1$, $\mu_2$ will be expressed in the sequel as a function $c(p)$ (that is, the dependence on $p$ will be explicitly indicated). Given $r$, $\mu_1$ and $\mu_2$, the latter function is determined by $G$. An estimator is said to *guarantee* a confidence level $c_0$ in the interval $(p_1, p_2)$ if $c(p) \geq c_0$ for all $p \in (p_1, p_2)$. If the interval is $(0, 1)$, then the estimator is said to *globally guarantee* this confidence level. An estimator *asymptotically guarantees* a certain $c_0$ if there exists $\varepsilon > 0$ such that the estimator guarantees $c_0$ in the interval $(0, \varepsilon)$.

The problem being addressed can be rephrased as that of optimizing the globally guaranteed confidence within the general class of estimators based on inverse binomial sampling, or finding a minimax estimator with respect to the risk defined by the loss function (1.4). The

following proposition, and its ensuing particularization, provides the motivation for studying $\liminf_{p \to 0} c(p)$ as part of this optimization problem.

**Proposition 1.** *If a given estimator, with confidence function $c(p)$, guarantees a confidence level $c_0$ in an interval $(p_1, p_2)$, then, necessarily, $\liminf_{p \to p_0} c(p) \geq c_0$ for any $p_0 \in [p_1, p_2]$.*

Particularizing Proposition 1 to $p_1 = p_0 = 0$, it is seen that $\liminf_{p \to 0} c(p)$ represents an upper bound on the confidence levels that can be globally or asymptotically guaranteed.

An important subclass of non-randomized estimators is formed by those defined by functions $g \in \mathcal{F}$ for which $c(p)$ has an asymptotic value, that is, for which $\lim_{p \to 0} P[p/\mu_2 \leq g(N) \leq p\mu_1]$ exists. The set of all such functions will be denoted $\mathcal{F}_{\mathrm{p}}$. As will be seen, another important subclass is that corresponding to the set of functions $g \in \mathcal{F}$ for which $\lim_{n \to \infty} ng(n)$ exists, is finite and non-zero. This set will be denoted $\mathcal{F}_{\mathrm{n}}$. The following result establishes that $\mathcal{F}_{\mathrm{n}} \subset \mathcal{F}_{\mathrm{p}}$.

**Proposition 2.** *For a non-randomized estimator defined by $g \in \mathcal{F}_{\mathrm{n}}$, $\lim_{p \to 0} c(p)$ exists and equals $\bar{c}$, given by*

$$\bar{c} = \gamma(r, \Omega\mu_2) - \gamma(r, \Omega/\mu_1), \qquad \Omega = \lim_{n \to \infty} ng(n). \tag{2.2}$$

The converse of Proposition 2 is not true; that is, $\mathcal{F}_{\mathrm{n}} \neq \mathcal{F}_{\mathrm{p}}$. A simple counterexample is given by

$$g(n) = \begin{cases} \omega'/n, & \text{if there exists } k \in \mathbb{N} \text{ such that } n = 2^k, \\ \omega/n, & \text{otherwise,} \end{cases}$$

with $\omega' \neq \omega$. It is easily seen that this function is in $\mathcal{F}_{\mathrm{p}}$, with $\lim_{p \to 0} c(p) = \gamma(r, \omega\mu_2) - \gamma(r, \omega/\mu_1)$; however, $\lim_{n \to \infty} ng(n)$ does not exist.

Given $r$, $\mu_1$ and $\mu_2$, the maximum of $\bar{c}$ over all $g \in \mathcal{F}_{\mathrm{n}}$ is attained when $\lim_{n \to \infty} ng(n)$ equals $\Omega^*$, given by

$$\Omega^* = r \frac{\log \mu_2 - \log(1/\mu_1)}{\mu_2 - 1/\mu_1}, \tag{2.3}$$

as is readily seen by differentiating $\bar{c}$ in (2.2). Let $c^*$ denote the resulting maximum. Defining

$$M = \mu_1\mu_2, \tag{2.4}$$

the terms $\Omega^*/\mu_1$ and $\Omega^*\mu_2$ can be expressed as

$$\Omega^*/\mu_1 = \frac{r \log M}{M - 1}, \qquad \Omega^*\mu_2 = \frac{r M \log M}{M - 1} \tag{2.5}$$

and thus

$$c^* = \gamma\left(r, \frac{r M \log M}{M - 1}\right) - \gamma\left(r, \frac{r \log M}{M - 1}\right). \tag{2.6}$$

The following theorem establishes that $c^*$ is not only the maximum of $\lim_{p \to 0} c(p)$ within the subclass of non-randomized estimators defined by $\mathcal{F}_n$, but also the maximum of $\liminf_{p \to 0} c(p)$ within the general class of randomized estimators defined by $\mathcal{F}_\mathcal{R}$.

**Theorem 1.** *The maximum of* $\liminf_{p \to 0} c(p)$ *over all estimators defined by functions* $G \in \mathcal{F}_\mathcal{R}$ *is* $c^*$, *given by* (2.6).

As can be seen from (2.6), $c^*$ depends on $\mu_1$ and $\mu_2$ only through $M$. An explanation of this result is as follows. Given $\mu_1$, $\mu_2$, let $\hat{p}$ be an arbitrary estimator and consider $a > 0$. If $\mu_1$ and $\mu_2$ are replaced by $\mu_1' = a\mu_1$ and $\mu_2' = \mu_2/a$, respectively, defining a modified estimator $\hat{p}' = a\hat{p}$, it is clear that $p/\mu_2' \le \hat{p}' \le p\mu_1'$ if and only if $p/\mu_2 \le \hat{p} \le p\mu_1$. This shows that any value of $\liminf_{p \to 0} c(p)$ that can be achieved for $\mu_1$, $\mu_2$ can also be achieved for $a\mu_1$, $\mu_2/a$ (using a different estimator), and conversely. Thus, $c^*$ is the same for $\mu_1$, $\mu_2$ and for $a\mu_1$, $\mu_2/a$.

# 3. An optimum estimator for certain relative intervals

The asymptotic results in Section 2 impose a limit on the confidence levels that can be guaranteed, as established by the following corollary of Proposition 1 and Theorem 1.

**Corollary 1.** *No estimator can guarantee a confidence level greater than* $c^*$, *given by* (2.6), *in an interval* $(0, p_2)$.

According to Corollary 1, $c^*$ is an upper bound on the confidence that can be guaranteed either asymptotically or globally. It remains to be seen if there exists some estimator that can actually guarantee the confidence level $c^*$. If it exists, that estimator will be optimum from the point of view of guaranteed confidence. A related question is if one such optimum estimator can be found within the restricted class of non-randomized estimators. As will be shown, the answer to both questions turns out to be affirmative for all values of $\mu_1$ and $\mu_2$ in the case of asymptotic guarantee and for certain values of $\mu_1$ and $\mu_2$ in the case of global guarantee.

Consider a non-randomized estimator $\hat{p} = g(N)$ of the form

$$g(n) = \frac{\Omega}{n + d}, \tag{3.1}$$

where $\Omega$ and $d$ are parameters (with $\Omega = \lim_{n \to \infty} ng(n)$). This is a generalization of the estimators (1.1)–(1.3). Note that (3.1) has an asymptotic confidence $\bar{c}$ given by (2.2) and, for $\Omega = \Omega^*$, it achieves the maximum $\liminf_{p \to 0} c(p)$ that any estimator can have, according to Theorem 1.

Under a mild condition on $d$, the estimator given by (3.1) with $\Omega = \Omega^*$ can be shown to asymptotically guarantee the confidence $c^*$ for any $\mu_1$, $\mu_2 > 0$, as established by Theorem 2 below.

**Theorem 2.** *The non-randomized estimator*

$$\hat{p} = \frac{\Omega^*}{N + d} \tag{3.2}$$

*asymptotically guarantees the optimum confidence c\* given in* (2.6) *if*

$$d > \frac{1}{2}\left(-r + \frac{M+1}{M-1}\right),\tag{3.3}$$

*where M is defined by* (2.4).

An estimator of this form can also *globally* guarantee the confidence level $c^*$, provided that $\mu_1$, $\mu_2$ are not too small, as discussed in the following.

For $d \in \mathbb{Z}$, the estimator defined by (3.1) lends itself to an analysis similar to that carried out by Mendo and Hernando (2006, 2008a) for the particular case (1.2). This allows the derivation of sufficient conditions on $\mu_1$, $\mu_2$ which ensure that $c(p) \geq \bar{c}$ for all $p \in (0, 1)$. The least restrictive conditions are obtained for $d = 1$ and are given in the following proposition, which, particularized to $\Omega = \Omega^*$, will yield the desired result on globally guaranteeing the optimum confidence.

**Proposition 3.** *The confidence of the non-randomized estimator*

$$\hat{p} = \frac{\Omega}{N+1}\tag{3.4}$$

*exceeds its asymptotic value $\bar{c}$, given by* (2.2), *for all $p \in (0, 1)$ if*

$$\mu_1 \geq \frac{\Omega}{r - \sqrt{r}},\qquad \mu_2 \geq \frac{r + \sqrt{r} + 1}{\Omega}.\tag{3.5}$$

Particularizing to $\Omega = \Omega^*$, Proposition 3 establishes that there exists a non-randomized estimator that can globally guarantee the optimum confidence $c^*$ for certain values of $\mu_1$, $\mu_2$. It turns out that for $\Omega = \Omega^*$ and $r$ given, one of the two inequalities in (3.5) implies the other. Thus, for each $r$, only one of the inequalities needs to be considered in order to determine the allowed range for $\mu_1$, $\mu_2$. The result is stated in the following theorem.

**Theorem 3.** *The non-randomized estimator*

$$\hat{p} = \frac{\Omega^*}{N+1},\tag{3.6}$$

*with $\Omega^*$ as in* (2.3), *globally guarantees the optimum confidence c\* given by* (2.6) *if either of the following conditions is satisfied*:

$$\frac{M-1}{\log M} \geq \frac{r + \sqrt{r}}{r - 1}\qquad for\ r \in \{3, 4\};\tag{3.7}$$

$$\frac{M \log M}{M - 1} \geq \frac{r + \sqrt{r} + 1}{r}\qquad for\ r \geq 5,\tag{3.8}$$

*where M is defined by* (2.4). *These conditions can be jointly expressed as $M \geq h(r)$, where h is an increasing function.*

Given $r$, consider the region of $(\mu_1, \mu_2)$ values that satisfy the appropriate condition (3.7) or (3.8), with $\mu_1, \mu_2 > 1$. From Theorem 3, the boundary of this region is a continuous, decreasing, concave curve in $(1, \infty) \times (1, \infty)$, namely, the curve determined by the equation $\mu_1\mu_2 = h(r)$. The region in question is the union of this curve and the portion of the plane lying above and to the right. Furthermore, the region for $r' > r$ contains that for $r$.

According to the preceding results, the problem of optimum estimation of $p$, in the sense of globally guaranteeing the maximum possible confidence, is solved by the non-randomized estimator (3.6) for $\mu_1, \mu_2$ satisfying (3.7) or (3.8). Equivalently, this estimator is minimax with respect to the risk defined by the loss function given in (1.4) (maximin with respect to confidence).

The fact that $c^*$ depends on $\mu_1, \mu_2$ only through $M$ gives rise to another interpretation of the result in Theorem 3. For $r$ and $M$ given, consider the problem of finding, among all interval estimators of $p$ with a ratio $M$ between their end-points, that which maximizes the globally guaranteed confidence. The solution, if $M$ satisfies (3.7) or (3.8), is

$$\left[ \frac{r \log M}{(M-1)(N+1)}, \frac{r M \log M}{(M-1)(N+1)} \right] \tag{3.9}$$

and the resulting maximum is $c^*$, as expressed by (2.6). Equivalently, given $r$ and a prescribed confidence $c_0$, if a value for $M$ is computed such that (2.6) holds with $c^* = c_0$ and if it satisfies (3.7) or (3.8), then the interval estimator (3.9) minimizes the ratio between interval end-points subject to a globally guaranteed confidence level $c_0$. Observe that it is meaningful to prescribe, or minimize, the ratio of the interval end-points, rather than their difference, since a given value for the latter might be either unacceptably high or unnecessarily small, depending on the unknown $p$, whereas the ratio has a definite meaning, regardless of $p$.

According to Proposition 3, conditions (3.5) are sufficient; however, they may not be necessary. The same applies to (3.7) and (3.8). Determining the most general conditions which assure optimality of (3.6) is a difficult problem.[1] Nevertheless, as will be illustrated, the sufficient conditions (3.7) or (3.8) cover most cases of interest.

## 4. Discussion

Figure 1(a) depicts the relationship between $c^*$, $M$ and $r$, for $M$ satisfying (3.7) or (3.8). The guaranteed confidence $c^*$ is represented, for convenience, as a function of $\sqrt{M} - 1$, each dashed curve corresponding to a different $r$. The figure also represents, with solid line, the minimum $c^*$ that fulfills inequalities (3.7) or (3.8); this corresponds to the lowest $r$ for which the applicable inequality holds. Figure 1(b) shows this minimum $r$ as a function of $\sqrt{M} - 1$, with $c^*$ as a parameter. From Corollary 1, this figure also has a more general interpretation as the minimum $r$ that is required in order to guarantee (either globally or asymptotically) a desired confidence level using any estimator based on inverse binomial sampling.

---

[1] Although $c(p)$ (or a lower bound thereof) can be expressed in terms of the Gauss hypergeometric function $_2F_1(a, b; c; t)$ (see the proof of Theorem 2), standard algorithms for evaluation of hypergeometric sums (Petkovšek et al. (1996)) are not directly applicable because of the existing dependence between $b$ and $t$.
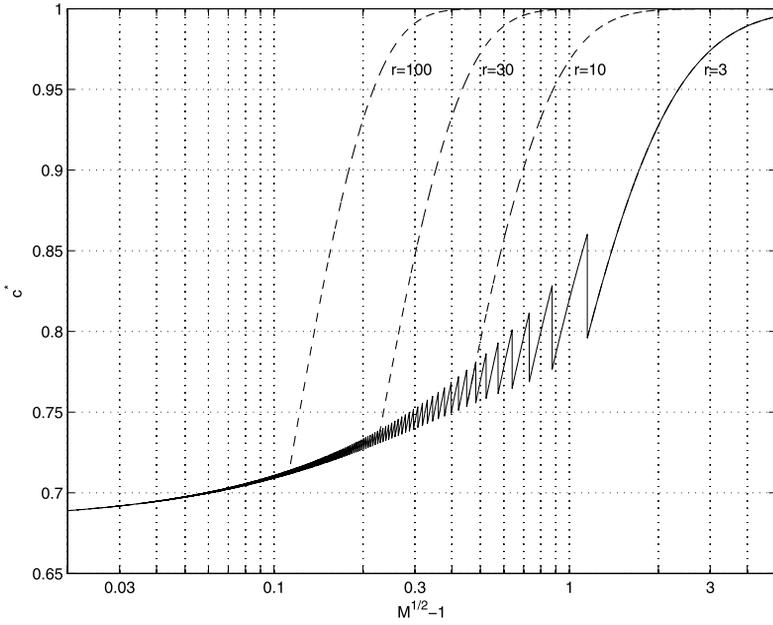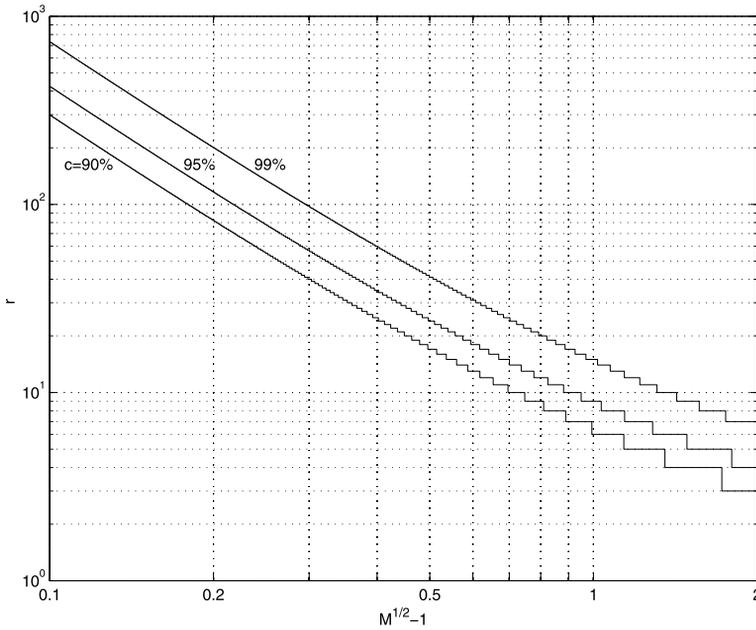
**Figure 1.** Behavior of optimum guaranteed confidence as a function of $\sqrt{M} - 1$: (a) $c^*$ (dashed) and minimum $c^*$ (solid); (b) minimum $r$ that guarantees a confidence level.

Given $\mu_1$, $\mu_2$ and $c$, if the point $(\sqrt{M} - 1, c)$ with $M = \mu_1\mu_2$ lies in the region above the solid curve in Figure 1(a), then there exist values of $r$ for which the estimator (3.6) globally guarantees the confidence level $c$ for the relative interval defined by $\mu_1$ and $\mu_2$; the minimum such $r$ is displayed in Figure 1(b). As mentioned earlier, the region referred to in Figure 1(a) (or, equivalently, conditions (3.7) and (3.8)) covers most cases of interest. For example, any confidence greater than 85% can be globally guaranteed for any relative interval with $\sqrt{M} - 1 \leq$ 1.108, that is, such that $\mu_1\mu_2 \leq 4.443$.

An important subclass of relative intervals is that for which $\mu_1 = \mu_2 = 1 + m$, $m > 0$. An interval of this form corresponds to the requirement that $\hat{p}$ and $p$ do not deviate from each other by a factor greater than $1 + m$; the parameter $m$ is thus interpreted as a relative error margin. The guaranteed confidence and required $r$ in this case can be read directly from Figures 1(a) and 1(b), as $\sqrt{M} - 1 = m$. This particular case is analyzed by Mendo and Hernando (2006, 2008a) for the estimator (1.2). Comparing Figure 1 of Mendo and Hernando (2008a) with Figure 1(a), the individual (dashed) curves in the latter are seen to be above those in the former (this is most noticeable for small $r$), in accordance with the fact that the estimator (3.6) is optimum. This yields a reduction in the error margin $m$ for a given $r$ and a desired guaranteed confidence. For example, taking $r = 10$, the estimator (1.2) guarantees a confidence level of 95% for $m = 0.9074$, whereas (3.6) guarantees the same confidence for $m = 0.8808$. Furthermore, the latter value is the smallest $m$ for which a confidence level of 95% can be guaranteed by any estimator with the $r$ in question.

As another manifestation of the optimum character of the estimator (3.6), the curves in Figure 1(b) for $\sqrt{M} - 1 = m$ are farther to the left than those in Figure 5(a) of Mendo and Hernando (2006). Since, from (2.1), $E[N] = r/p$, for certain combinations of $m$ and $c$, this provides a reduction in average observation time to achieve a globally guaranteed confidence $c$ for an error margin $m$. (The fact that the reduction is obtained only for certain combinations of $m$ and $c$ is a consequence of the discrete character of $r$.) Thus, for $m = 50\%$, the estimator (1.2) requires $r = 18$ in order to globally guarantee a 90% confidence level, whereas $r = 17$ suffices for the estimator (3.6); furthermore, this is the lowest required $r$ that can be achieved by any estimator.

Comparing the attainable region shown in Figure 1(a), that is, the region above the solid curve, with the corresponding region in Figure 1 of Mendo and Hernando (2008a), they are seen to have similar shape and size, except that for small $m$, the boundary curve is slightly higher in Figure 1(a). Thus, the applicability of the estimator (3.6) is similar to that of (1.2) (while achieving better performance).

Another interesting particularization is $\mu_1 = 1 + m$, $\mu_2 = 1/(1 - m)$, $0 < m < 1$, which corresponds to requiring that the absolute error $|\hat{p} - p|$ does not exceed a fraction $m$ of the true value $p$. The results for this case can be compared with those of Mendo and Hernando (2008a) and Chen (2007). For $m = 40\%$ with a globally guaranteed confidence level of 90%, the estimator (1.2) requires $r = 17$ (Mendo and Hernando (2008a), Proposition 1). The results of Chen (2007), Theorem 2, give a sufficient value of $r = 46$ for the estimator (1.3). On the other hand, from Theorem 3, the estimator (3.6) only requires $r = 16$; moreover, it is assured that any other estimator requires at least this $r$ to guarantee (either globally or asymptotically) the same confidence for the $m$ under consideration.

## 5. Proofs

The following notation is introduced for convenience:

$$\phi(t) = \frac{t^{r-1} \exp(-t)}{(r-1)!}, \tag{5.1}$$

$$n_1 = \lceil \Omega/(p\mu_1) - d \rceil, \qquad n_2 = \lfloor \Omega\mu_2/p - d \rfloor. \tag{5.2}$$

**Proof of Proposition 1.** Let $c_i$ denote $\liminf_{p \to p_0} c(p)$ and assume that $c_i < c_0$. For $\varepsilon = (c_0 - c_i)/2$, the definition of limit inferior implies that there exists $p_\varepsilon \in (p_1, p_2)$ such that $c(p_\varepsilon) < c_i + \varepsilon < c_0$ and thus the estimator does not guarantee the confidence level $c_0$ in $(p_1, p_2)$. This establishes the result. □

**Lemma 1.** *For all $t > 0$, the function $\phi$ defined in (5.1) satisfies*

$$0 < \phi(t) \leq \frac{(r-1)^{r-1} \exp(-r+1)}{(r-1)!} < 1. \tag{5.3}$$

**Proof.** The first inequality in (5.3) is obvious. Let $Q_r = (r-1)^{r-1} \exp(-r+1)/(r-1)!$. Maximizing $\phi(t)$ with respect to $t$, it is seen that $\phi(t) \leq Q_r$. Since $Q_{r+1}/Q_r = (1 + 1/(r-1))^{r-1} \exp(-1) < 1$ and $Q_3 = 2\exp(-2) < 1$, it follows that $Q_r < 1$ for all $r \geq 3$. □

**Proof of Proposition 2.** Consider $g \in \mathcal{F}_n$ and let $\Omega = \lim_{n \to \infty} ng(n)$. Given $\varepsilon > 0$, there exists $n_\varepsilon$ such that $|ng(n) - \Omega| < \varepsilon$ for all $n \geq n_\varepsilon$, that is, $(\Omega - \varepsilon)/n < g(n) < (\Omega + \varepsilon)/n$. Using this, the confidence $c(p) = P[p/\mu_2 \leq g(N) \leq p\mu_1]$ can be bounded for $p \leq (\Omega - \varepsilon)/(\mu_1 n_\varepsilon)$ as

$$c(p) \geq P[\lceil (\Omega + \varepsilon)/(p\mu_1) \rceil \leq N \leq \lfloor (\Omega - \varepsilon)\mu_2/p \rfloor], \tag{5.4}$$

$$c(p) \leq P[\lceil (\Omega - \varepsilon)/(p\mu_1) \rceil \leq N \leq \lfloor (\Omega + \varepsilon)\mu_2/p \rfloor]. \tag{5.5}$$

Let $b(n, p; i)$ denote the binomial probability function with parameters $n$ and $p$ evaluated at $i$. From the relationship between binomial and negative binomial distributions, (5.4) is written as

$$c(p) \geq \sum_{i=0}^{r-1} b\big(\lceil (\Omega + \varepsilon)/(p\mu_1) \rceil - 1, p; i\big) - \sum_{i=0}^{r-1} b\big(\lfloor (\Omega - \varepsilon)\mu_2/p \rfloor, p; i\big). \tag{5.6}$$

According to the Poisson theorem (Papoulis and Pillai (2002), page 113), the right-hand side of (5.6) converges to $\gamma(r, (\Omega - \varepsilon)\mu_2) - \gamma(r, (\Omega + \varepsilon)/\mu_1)$ as $p \to 0$. This implies that

$$\liminf_{p \to 0} c(p) \geq \gamma\big(r, (\Omega - \varepsilon)\mu_2\big) - \gamma\big(r, (\Omega + \varepsilon)/\mu_1\big). \tag{5.7}$$

Since $\partial\gamma(r, t)/\partial t = \phi(t)$, Lemma 1 establishes that $0 < \partial\gamma(r, t)/\partial t < 1$. Using this, (5.7) yields $\liminf_{p \to 0} c(p) \geq \bar{c} - (\mu_2 + 1/\mu_1)\varepsilon$. Taking into account that this holds for all $\varepsilon > 0$, it follows that $\liminf_{p \to 0} c(p) \geq \bar{c}$. Applying similar arguments to (5.5) gives $\limsup_{p \to 0} c(p) \leq \bar{c}$. Thus, $\lim_{p \to 0} c(p)$ exists and equals $\bar{c}$. □

**Lemma 2 (Abramowitz and Stegun (1970), inequality (4.1.33)).** $t/(1 + t) \leq \log(1 + t) \leq t$ *for* $t > -1$, *with equality if and only if* $t = 0$.

**Lemma 3.** *Let* $(p_k)$ *be a positive sequence which converges to* 0. *For any* $v_1$, $v_2$ *with* $v_2 > v_1 > 0$, *the sequence of functions* $(\phi_k)$ *defined as*

$$\phi_k(v) = \frac{(1 - p_k)^{v/p_k - r}}{(r - 1)!} \prod_{i=1}^{r-1} (v - ip_k) \tag{5.8}$$

*converges uniformly to* $\phi(v)$, *given by* (5.1), *in the interval* $[v_1, v_2]$.

**Proof.** Since the sequence $(p_k)$ is positive and converges to 0, there exists $k_1$ such that $p_k < \min\{v_1/(r - 1), 1\}$ for $k \geq k_1$. Thus, $\phi_k(v) > 0$ for $v \in [v_1, v_2]$, $k \geq k_1$ and, therefore,

$$|\phi_k(v) - \phi(v)| = \min\{\phi_k(v), \phi(v)\} \left( \max\left\{ \frac{\phi_k(v)}{\phi(v)}, \frac{\phi(v)}{\phi_k(v)} \right\} - 1 \right). \tag{5.9}$$

Lemma 1 implies that $\min\{\phi_k(v), \phi(v)\} \leq 1$. On the other hand,

$$\max\left\{ \frac{\phi_k(v)}{\phi(v)}, \frac{\phi(v)}{\phi_k(v)} \right\} = \exp\left| \log \frac{\phi_k(v)}{\phi(v)} \right|.$$

Substituting into (5.9), we have

$$|\phi_k(v) - \phi(v)| \leq \exp\left| \log \frac{\phi_k(v)}{\phi(v)} \right| - 1. \tag{5.10}$$

From (5.1) and (5.8), it follows that, assuming $v \in [v_1, v_2]$ and $k \geq k_1$,

$$\left| \log \frac{\phi_k(v)}{\phi(v)} \right| \leq \left| \sum_{i=1}^{r-1} \log\left( 1 - \frac{ip_k}{v} \right) \right| + \left| \left( \frac{v}{p_k} - r \right) \log(1 - p_k) + v \right|. \tag{5.11}$$

Using Lemma 2, the first term in the right-hand side of (5.11) is bounded as follows:

$$\left| \sum_{i=1}^{r-1} \log\left( 1 - \frac{ip_k}{v} \right) \right| = - \sum_{i=1}^{r-1} \log\left( 1 - \frac{ip_k}{v} \right) < \sum_{i=1}^{r-1} \frac{ip_k}{v - ip_k}. \tag{5.12}$$

As for the second term, since $v \in [v_1, v_2]$, there exists $k_2 \geq k_1$ such that $p_k < \min\{v_1/r, 1\}$ for $k \geq k_2$. Thus, $(v/p_k - r) \log(1 - p_k) < 0$, whereas $v > 0$. From Lemma 2, $-p_k/(1 - p_k) < \log(1 - p_k) < -p_k$. It follows that

$$\left| \left( \frac{v}{p_k} - r \right) \log(1 - p_k) + v \right| < \max\left\{ \left| -\left( \frac{v}{p_k} - r \right) p_k + v \right|, \left| -\left( \frac{v}{p_k} - r \right) \frac{p_k}{1 - p_k} + v \right| \right\}$$

$$= \max\left\{ r, \frac{|r - v|}{1 - p_k} \right\} p_k < \max\{r, v_2\} \frac{p_k}{1 - p_k}. \tag{5.13}$$

Combining (5.10)–(5.13) yields the following bound, valid for all $v \in [v_1, v_2]$, $k \geq k_2$:

$$|\phi_k(v) - \phi(v)| < \exp\left[\left(\sum_{i=1}^{r-1} \frac{i}{v - ip_k} + \frac{\max\{r, v_2\}}{1 - p_k}\right)p_k\right] - 1. \tag{5.14}$$

The right-hand side of (5.14) tends to 0 as $k \to \infty$. Therefore, $\sup_{v \in [v_1, v_2]} |\phi_k(v) - \phi(v)|$ also tends to 0, which establishes the result. $\qquad\square$

**Proof of Theorem 1.** For $\alpha \in (0, 1]$, let the sequence $(p_k)$, $k \in \mathbb{N}$ be defined as $p_k = \alpha M^{-k}$, with $M$ given by (2.4), and let the sequence of intervals $(I_k)$ be defined as $I_k = (p_k/\mu_2, p_k\mu_1]$. Consider also the sequence $(f_k)$, where $f_k$ is the probability function of $N$ with parameters $r$ and $p_k$. From (2.1) and (5.8),

$$f_k(n) = p_k\phi_k(np_k). \tag{5.15}$$

To facilitate the development, it is convenient to first analyze non-randomized estimators, and then generalize to randomized estimators. Given a non-randomized estimator specified by $g \in \mathcal{F}$, let the sequence of sets $(S_k)$ be defined such that $n \in S_k$ if and only if $g(n) \in I_k$. Since the intervals $I_k$ are disjoint for different $k$, the sets $S_k$ are also disjoint. Let the function $\sigma$ be defined such that $\sigma(n) = k$ if and only if $n \in S_k$, with $\sigma(n) = 0$ (or an arbitrary negative value) if $n \notin S_k$ for all $k \in \mathbb{N}$. Thus, $\sigma$ gives the index $k$ of the interval $I_k$ that $g$ associates to each $n$, if any. The function $\sigma$ is determined by $g$; and, for a given $n_0$, $\sigma(n_0)$ can be modified without affecting the rest of values $\sigma(n)$, $n \neq n_0$ by adequately choosing $g(n_0)$.

For the considered non-randomized estimator, let $c_k$ denote the probability that $\hat{p}$ lies in $I_k$ when $p = p_k$, i.e. $c_k = \sum_{n \in S_k} f_k(n)$. Further, let $C(h, H) = \sum_{k=h}^{h+H-1} c_k$ for $h, H \in \mathbb{N}$ arbitrary. Defining

$$s(n) = \begin{cases} f_{\sigma(n)}(n), & \text{if } n \in S_h \cup \cdots \cup S_{h+H-1}, \\ 0, & \text{otherwise,} \end{cases}$$

the sum $C(h, H)$ can be expressed as $\sum_{n=r}^{\infty} s(n)$. It follows that

$$C(h, H) \leq \sum_{n=r}^{\infty} \max_{k \in \{h, \dots, h+H-1\}} f_k(n), \tag{5.16}$$

which holds with equality if the estimator satisfies

$$\sigma(n) = \arg\max_{k \in \{h, \dots, h+H-1\}} f_k(n) \qquad \text{for all } n \in \mathbb{N}, n \geq r, \tag{5.17}$$

where, if the maximum is reached at more than one index $k$, the arg max function is arbitrarily defined to give the lowest such index.

As for randomized estimators, consider an arbitrary function $G \in \mathcal{F}_\mathcal{R}$, that for each $n$ specifies $\Pi_n$, the distribution function of $\hat{p}$ conditioned on $N = n$. Let $I^c = (-\infty, p_{h+H-1}/\mu_2] \cup (p_h\mu_1, \infty)$. Conditioned on $N = n$, let $\pi_{n,k}$ and $\pi_n^c$ respectively denote the probabilities that $\hat{p}$ is in $I_k$ and in $I^c$. Obviously,

$$\pi_{n,h} + \cdots + \pi_{n,h+H-1} + \pi_n^c = 1. \tag{5.18}$$

For $N = n$, the numbers $\pi_{n,h}, \ldots, \pi_{n,h+H-1}, \pi_n^c$ indicate how the conditional probability associated to all possible values of $\hat{p}$ (that is, 1 in total) is divided among $I_h, \ldots, I_{h+H-1}, I^c$. For a given $n_0$, any combination of values $\pi_{n_0,h}, \ldots, \pi_{n_0,h+H-1}, \pi_{n_0}^c$ allowed by (5.18) can be realized, without affecting other values $\pi_{n,k}, \pi_n^c$ for $n \neq n_0$, by adequately choosing the distribution function $\Pi_{n_0}$. Defining $c_k$ and $C(h, H)$ as in the non-randomized case, the former is expressed as

$$c_k = \sum_{n=r}^{\infty} \pi_{n,k} f_k(n). \tag{5.19}$$

Since all terms in (5.19) are positive, the series converges absolutely, and thus $C(h, H)$ can be written as

$$C(h, H) = \sum_{k=h}^{h+H-1} c_k = \sum_{n=r}^{\infty} \sum_{k=h}^{h+H-1} \pi_{n,k} f_k(n).$$

It is evident from this expression that $C(h, H)$ is maximized if, for each $n$, the values $\pi_{n,k}, \pi_n^c$ are chosen as $\pi_{n,l} = 1$ for $l = \arg\max_k f_k(n)$, $\pi_{n,k} = 0$ for $k \neq l$, $\pi_n^c = 0$; and the resulting maximum coincides with the right-hand side of (5.16). Thus the inequality (5.16), initially derived for non-randomized estimators, also holds for the general class of randomized estimators. This implies that for any randomized estimator there exists a non-randomized estimator that attains the same or greater $C(h, H)$.

Let the function $\bar{g} \in \mathcal{F}$ be defined as $\bar{g}(n) = \Omega^*/n$, and consider the non-randomized estimator specified by this function. The sets $S_k$ and function $\sigma$ associated to this estimator will be denoted as $\bar{S}_k$ and $\bar{\sigma}$ respectively. It is seen that

$$\bar{S}_k = \left[\Omega^*/(\mu_1 p_k), \Omega^* \mu_2/p_k\right) \cap \{r, r+1, r+2, \ldots\}. \tag{5.20}$$

Let the sequence of functions $(\bar{f}_k)$ be defined as $\bar{f}_k(n) = p_k \phi(n p_k)$. It is readily seen that the equation $\bar{f}_k(t) = \bar{f}_{k+1}(t)$ has only one solution, given as $t = \Omega^* \mu_2/p_k$. Similarly, $\bar{f}_{k-1}(t) = \bar{f}_k(t)$ has the solution $t = \Omega^*/(\mu_1 p_k)$. Taking into account that the functions $\bar{f}_k$ are unimodal, this implies that $\bar{f}_k(n) \geq \bar{f}_{k-1}(n)$ and $\bar{f}_k(n) \geq \bar{f}_{k+1}(n)$ for $n \in \bar{S}_k$. An analogous argument shows that $\bar{f}_k(n) > \bar{f}_{k-i}(n)$ and $\bar{f}_k(n) > \bar{f}_{k+i}(n)$ for $n \in \bar{S}_k$ and $i \geq 2$. Therefore, the function $\bar{\sigma}$ associated to $\bar{g}$ satisfies a modified version of (5.17) in which $f_k$ is replaced by $\bar{f}_k$, that is, $\phi_k$ in (5.15) is replaced by its limit $\phi$. In the following, using the fact that the difference between $\phi_k$ and $\phi$ is small for large $k$, the estimator defined by $\bar{g}$ will be used to derive from (5.16) a more explicit upper bound on $C(h, H)$. Since the sum $C(h, H)$ attained by any estimator is equalled or exceeded by some non-randomized estimator, it will be sufficient to restrict to the class of non-randomized estimators.

Consider an arbitrary non-randomized estimator defined by $g \in \mathcal{F}$ with its corresponding function $\sigma$. Differentiating (2.1) with respect to $p$, it is seen that $\partial f(n)/\partial p$ is positive for $p < r/n$ and negative for $p > r/n$. According to (5.20),

$$\Omega^*/(\mu_1 n) \leq p_k < \Omega^* \mu_2/n \qquad \text{for } n \in \bar{S}_k. \tag{5.21}$$

Using Lemma 2, it stems from (2.5) that

$$\Omega^*/\mu_1 < r < \Omega^*\mu_2. \tag{5.22}$$

The first inequality in (5.21) and the second in (5.22) imply that $p_{k-1} = Mp_k \geq \Omega^*\mu_2/n > r/n$ for $n \in \bar{S}_k$. Thus $\partial f(n)/\partial p < 0$ for $p \geq p_{k-1}$. This implies that $f_{k+i}(n) < f_{k-1}(n)$ for $n \in \bar{S}_k$ and $i \leq -2$. It follows that if $\sigma(n_0) < \bar{\sigma}(n_0) - 1$ for a given $n_0$, the sum $C(h, H)$ could be made larger by modifying the value $g(n_0)$ so as to attain $\sigma(n_0) = \bar{\sigma}(n_0) - 1$; unless $\sigma(n_0) = h + H - 1$, in which case modifying $g(n_0)$ cannot make $C(h, H)$ larger. Analogously, the second inequality in (5.21) and the first in (5.22) yield $f_{k+i}(n) < f_{k+1}(n)$ for $n \in \bar{S}_k$ and $i \geq 2$. Therefore, if $\sigma(n_0) > \bar{\sigma}(n_0) + 1$ for some $n_0$, $C(h, H)$ could be made larger by modifying $g(n_0)$, unless $\sigma(n_0) = h$.

According to the above, in order to obtain an upper bound on $C(h, H)$, it suffices to consider non-randomized estimators such that $\sigma(n) \in \{\bar{\sigma}(n) - 1, \bar{\sigma}(n), \bar{\sigma}(n) + 1\}$ for $n \in \bar{S}_h \cup \cdots \cup \bar{S}_{h+H-1}, \sigma(n) = h$ for $n < \lceil \Omega^*/(\mu_1 p_h) \rceil$ and $\sigma(n) = h + H - 1$ for $n \geq \lceil \Omega^*\mu_2/p_{h+H-1} \rceil$. Thus, from (5.16),

$$C(h, H) \leq \sum_{k=h}^{h+H-1} \sum_{n \in \bar{S}_k} \max_{i \in \{-1,0,1\}} f_{k+i}(n) + \sum_{n=r}^{\lceil \Omega^*/(\mu_1 p_h) \rceil - 1} f_h(n)$$

$$+ \sum_{n=\lceil \Omega^*\mu_2/p_{h+H-1} \rceil}^{\infty} f_{h+H-1}(n) \tag{5.23}$$

$$\leq \sum_{k=h}^{h+H-1} \sum_{n \in \bar{S}_k} \max_{i \in \{-1,0,1\}} f_{k+i}(n) + 2.$$

Let $\nu_1 = \Omega^*/\mu_1$, $\nu_2 = \Omega^*\mu_2$, and consider $\varepsilon > 0$. From Lemma 3, there exists $k_1$ such that $|\phi_k(\nu) - \phi(\nu)| < \varepsilon$ for $\nu \in [\nu_1, \nu_2]$, $k \geq k_1$. In addition, (5.20) implies that $np_k \in [\nu_1, \nu_2]$ for $n \in \bar{S}_k$. From these facts and (5.15) it stems that $|f_k(n)/p_k - \phi(np_k)| < \varepsilon$ for $n \in \bar{S}_k$, $k \geq k_1$. Thus, for $k \geq k_1 + 1$

$$\max_{i \in \{-1,0,1\}} f_{k+i}(n) < \max_{i \in \{-1,0,1\}} \left[ p_{k+i}\big(\phi(np_{k+i}) + \varepsilon\big) \right]$$

$$\leq \max_{i \in \{-1,0,1\}} (p_{k+i}\phi(np_{k+i})) + \varepsilon p_{k-1}. \tag{5.24}$$

As previously shown, $\bar{f}_k(n) = p_k\phi(np_k)$ equals or exceeds $\bar{f}_{k-1}(n)$ and $\bar{f}_{k+1}(n)$ for $n \in \bar{S}_k$. Therefore, from (5.24)

$$\max_{i \in \{-1,0,1\}} f_{k+i}(n) < p_k\phi(np_k) + \varepsilon p_{k-1}. \tag{5.25}$$

Substituting (5.25) into (5.23),

$$C(h, H) \leq \sum_{k=h}^{h+H-1} \sum_{n \in \bar{S}_k} \big(p_k\phi(np_k) + \varepsilon p_{k-1}\big) + 2. \tag{5.26}$$

The number of elements in the set $\bar{S}_k$ is less than $\Omega^*(\mu_2 - 1/\mu_1)/p_k + 1$, according to (5.20). From (2.5), this upper bound equals $r \log M / p_k + 1$. Using this in (5.26), and taking into account that $p_{k-1} < 1$,

$$C(h, H) < \sum_{k=h}^{h+H-1} \sum_{n \in \bar{S}_k} p_k \phi(np_k) + \varepsilon H (rM \log M + 1) + 2. \tag{5.27}$$

The term $\sum_{n \in \bar{S}_k} p_k \phi(np_k)$ in (5.27) converges to $\int_{\Omega^*/\mu_1}^{\Omega^*\mu_2} \phi(v)\,dv = c^*$ as $k \to \infty$. Thus, for the considered $\varepsilon$, there exists $k_2$ such that for all $k \geq k_2$

$$\left| \sum_{n \in \bar{S}_k} p_k \phi(np_k) - c^* \right| < \varepsilon. \tag{5.28}$$

Consequently, defining $k_0 = \max\{k_1 + 1, k_2\}$, inequalities (5.27) and (5.28) imply that for all $h \geq k_0$, for all $H$, and for $\alpha \in (0, 1]$,

$$\begin{aligned} C(h, H) &< H(c^* + \varepsilon) + \varepsilon H (rM \log M + 1) + 2 \\ &= H(c^* + \varepsilon P) + 2, \end{aligned} \tag{5.29}$$

where $P = rM \log M + 2 > 0$.

Since the minimum of a set cannot be larger than the average of the set, from (5.29) it follows that there is some $k_3 \in \{h, \dots, h + H - 1\}$ such that $c_{k_3} \leq C(h, H)/H < c^* + \varepsilon P + 2/H$.

Using the foregoing results, the bound $\liminf_{p \to 0} c(p) \leq c^*$ for an arbitrary estimator can be established by contradiction. Assume that there is some estimator, defined by $G \in \mathcal{F}_{\mathcal{R}}$, such that $\liminf_{p \to 0} c(p) = c^* + d$ with $d > 0$. This means that for any $\varepsilon' > 0$ there exists $p_0$ such that $c(p) \geq c^* + d - \varepsilon'$ for all $p \leq p_0$. Thus taking $\varepsilon' = d/3$, there is $p_0$ such that

$$c(p) \geq c^* + 2d/3 \qquad \text{for all } p \leq p_0. \tag{5.30}$$

On the other hand, taking $\varepsilon = d/(6P)$, $H = \lceil 12/d \rceil$, and with $\alpha \in (0, 1]$ arbitrary, the result in the preceding paragraph assures that for any $h$ not smaller than a certain $k_0$ (which depends on the considered $\varepsilon$) there exists $k_3 \in \{h, \dots, h + H - 1\}$ with

$$c_{k_3} < c^* + \varepsilon P + 2/H \leq c^* + d/3. \tag{5.31}$$

Let $h$ be selected such that $h \geq \max\{k_0, -\log p_0/\log M\}$. For each $k = h, \dots, h + H - 1$, let $X_k$ denote the (possibly empty) set of all points $x \in \mathbb{R}$ such that, for the considered estimator and for $p = p_k$, the probability that $\hat{p}$ equals $x$ is at least $d/3$. The number of points in $X_k$ cannot exceed $\lfloor 3/d \rfloor$, for otherwise the sum of their probabilities would be greater than 1. The set $X_k$ is determined by $f_k$ and $G$ (or by $f_k$ and $g$, in the case of a non-randomized estimator defined as $\hat{p} = g(N)$).

Let $\alpha \in (0, 1]$ be chosen such that

$$p_k/\mu_2 \notin X_k \qquad \text{for all } k = h, \ldots, h + H - 1. \tag{5.32}$$

Such $\alpha$ necessarily exists because (5.32) excludes only a finite number of possible values from $(0, 1]$. This choice of $\alpha$ assures that for $p = p_k$, $k = h, \ldots, h + H - 1$, the probability that $\hat{p}$ equals $p_k/\mu_2$ is smaller than $d/3$. Thus $c(p_k) < c_k + d/3$ for $k = h, \ldots, h + H - 1$, which, together with (5.31), gives

$$c(p_{k_3}) < c^* + 2d/3. \tag{5.33}$$

On the other hand, since $k_3 \geq h$, from the choice of $h$ it follows that

$$p_{k_3} = \alpha M^{-k_3} \leq \alpha M^{-h} \leq \alpha M^{\log p_0/\log M} = \alpha p_0 \leq p_0.$$

This implies, according to (5.30), that $c(p_{k_3}) \geq c^* + 2d/3$, in contradiction with (5.33). Therefore $\liminf_{p\to 0} c(p) \leq c^*$.

It has been shown that, for any estimator, $\liminf_{p\to 0} c(p)$ cannot exceed $c^*$. In addition, any non-randomized estimator defined by a function $g \in \mathcal{F}_n$ with $\lim_{n\to\infty} ng(n) = \Omega^*$ achieves $\liminf_{p\to 0} c(p) = \lim_{p\to 0} c(p) = c^*$. Therefore, $c^*$ is the maximum of $\liminf_{p\to 0} c(p)$ over all randomized or non-randomized estimators. $\qquad\square$

**Proof of Theorem 2.** Let $I_p(z, w)$ denote the regularized incomplete beta function:

$$I_p(z, w) = \frac{1}{B(z, w)} \int_0^p t^{z-1}(1 - t)^{w-1} \, dt,$$

$$B(z, w) = \int_0^1 t^{z-1}(1 - t)^{w-1} \, dt$$

$$= \frac{\Gamma(z)\Gamma(w)}{\Gamma(z + w)}.$$

From Abramowitz and Stegun (1970), equation (6.6.4), $P[N \leq n] = I_p(r, n - r + 1)$ for $n \in \mathbb{N}$, $n \geq r$. The confidence for the estimator (3.2) can thus be written as

$$c(p) = P[n_1 \leq N \leq n_2] = I_p(r, n_2 - r + 1) - I_p(r, n_1 - r), \tag{5.34}$$

where $n_1$ and $n_2$ are given by (5.2) with $\Omega = \Omega^*$. It is easy to show that $I_p(z, w)$ is an increasing function of $w$ for $w \in \mathbb{R}$. Since $n_1 < \Omega^*/(p\mu_1) - d + 1$ and $n_2 > \Omega^*\mu_2/p - d - 1$, it follows from (5.34) that

$$c(p) > \tilde{c}_2(p) - \tilde{c}_1(p) \tag{5.35}$$

with $\tilde{c}_1(p) = I_p(r, \Omega^*/(p\mu_1) - r - d + 1)$ and $\tilde{c}_2(p) = I_p(r, \Omega^*\mu_2/p - r - d)$.

Expressing $I_p(z, w)$ as (Abramowitz and Stegun (1970), equation (26.5.23))

$$I_p(z, w) = \frac{{}_2F_1(z, 1 - w; z + 1; p)p^z}{B(z, w)z},$$

$${}_2F_1(a, b; c; t) = \sum_{j=0}^{\infty} \frac{(a + j - 1)^{(j)}(b + j - 1)^{(j)}t^j}{(c + j - 1)^{(j)}j!},$$

the term $\tilde{c}_2(p)$ can be written as

$$\tilde{c}_2(p) = \frac{1}{B(r, \Omega^*\mu_2/p - r - d)} \frac{p^r}{r} {}_2F_1(r, r + d + 1 - \Omega^*\mu_2/p; r + 1; p) \quad (5.36)$$

$$= \frac{(\Omega^*\mu_2/p - d - 1)^{(r)}}{(r - 1)!} \sum_{j=0}^{\infty} \frac{(-1)^j (\Omega^*\mu_2/p - r - d - 1)^{(j)}}{(r + j)j!} p^{r+j}$$

$$= \frac{1}{(r - 1)!} \sum_{j=0}^{\infty} \frac{(-1)^j (\Omega^*\mu_2 - (d + 1)p) \cdots (\Omega^*\mu_2 - (d + r + j)p)}{(r + j)j!}. \quad (5.37)$$

Similarly,

$$\tilde{c}_1(p) = \frac{1}{(r - 1)!} \sum_{j=0}^{\infty} \frac{(-1)^j (\Omega^*/\mu_1 - dp) \cdots (\Omega^*/\mu_1 - (d + r + j - 1)p)}{(r + j)j!}. \quad (5.38)$$

According to (5.37) and (5.38), both $\tilde{c}_1(p)$ and $\tilde{c}_2(p)$ can be expressed as power series in $p$: $\tilde{c}_1(p) = \sum_{i=0}^{\infty} u_i p^i$, $\tilde{c}_2(p) = \sum_{i=0}^{\infty} v_i p^i$. Thus, the right-hand side of (5.35) is also a power series, $\sum_{i=0}^{\infty} w_i p^i$, with $w_i = v_i - u_i$. The zero-order coefficient is

$$w_0 = \frac{1}{(r - 1)!} \left( \sum_{j=0}^{\infty} \frac{(-1)^j (\Omega^*\mu_2)^{r+j}}{(r + j)j!} - \sum_{j=0}^{\infty} \frac{(-1)^j (\Omega^*/\mu_1)^{r+j}}{(r + j)j!} \right).$$

From the Taylor expansion of $\gamma(r, t)$ (Abramowitz and Stegun (1970), equation (6.5.29)),

$$\gamma(r, t) = \frac{1}{\Gamma(r)} \sum_{j=0}^{\infty} \frac{(-1)^j t^{r+j}}{(r + j)j!},$$

the coefficient $w_0$ is recognized to be $c^*$, reflecting the fact that the difference between the two sides of (5.35) is vanishingly small as $p \to 0$. In addition, the first order coefficient $w_1$ coincides with the derivative of the right-hand side of (5.35) evaluated at $p = 0$. It follows that $w_1 > 0$ is a sufficient condition for the estimator (3.2) to asymptotically guarantee the confidence level $c^*$.

The coefficient $w_1 = v_1 - u_1$ is computed as follows. The term $v_1$ is obtained from (5.37) as

$$v_1 = \frac{1}{(r - 1)!} \sum_{j=0}^{\infty} \frac{(-1)^{j+1} (\Omega^*\mu_2)^{r+j-1}[(d + 1) + \cdots + (d + r + j)]}{(r + j)j!}. \quad (5.39)$$

Substituting $(d+1) + \cdots + (d+r+j) = (r+j)(r+j+1+2d)/2$ into (5.39) and taking into account that $\exp(t) = \sum_{j=0}^{\infty} t^j/j!$, we have

$$
\begin{aligned}
v_1 &= \frac{r+1+2d}{2(r-1)!} \sum_{j=0}^{\infty} \frac{(-1)^{j+1}(\Omega^*\mu_2)^{r+j-1}}{j!} + \frac{1}{2(r-1)!} \sum_{j=0}^{\infty} \frac{(-1)^j(\Omega^*\mu_2)^{r+j}}{j!} \\
&= \frac{(\Omega^*\mu_2)^{r-1} \exp(-\Omega^*\mu_2)}{2(r-1)!} (\Omega^*\mu_2 - r - 1 - 2d).
\end{aligned}
\tag{5.40}
$$

Similarly,

$$
u_1 = \frac{(\Omega^*/\mu_1)^{r-1} \exp(-\Omega^*/\mu_1)}{2(r-1)!} (\Omega^*/\mu_1 - r + 1 - 2d).
\tag{5.41}
$$

From (5.40) and (5.41), and making use of (2.5),

$$
\begin{aligned}
w_1 &= \frac{(\Omega^*/\mu_1)^{r-1}}{2(r-1)!} [M^{r-1} \exp(-\Omega^*\mu_2)(\Omega^*\mu_2 - r - 1 - 2d) \\
&\qquad - \exp(-\Omega^*/\mu_1)(\Omega^*/\mu_1 - r + 1 - 2d)] \\
&= \frac{1}{2(r-1)!} \left( \frac{r \log M}{M-1} \right)^{r-1} M^{-r/(M-1)} [-(r+1+2d)/M + r - 1 + 2d],
\end{aligned}
$$

which is positive if (3.3) holds. Thus, (3.2) asymptotically guarantees the confidence $c^*$ for $d$ as in (3.3). $\qquad\square$

**Lemma 4.** *For $r, n_0 \in \mathbb{N}$ with $r \geq 3$ and $n_0 \leq (r - \sqrt{r})/p$,*

$$
\int_r^{n_0} t^{r-1} \exp(-pt)\, dt \geq \sum_{n=r}^{n_0-1} (n-1)^{(r-1)}(1-p)^{n-r}.
\tag{5.42}
$$

**Proof.** For $n_0$ as given, the sub-integral function in (5.42) is increasing within the integration range. Consequently, for (5.42) to hold, it is sufficient that $n^{r-1} \exp(-np) \geq (n-1)^{(r-1)}(1-p)^{n-r}$ for $n = r, \ldots, \lfloor (r - \sqrt{r})/p \rfloor - 1$. This condition can be shown to be satisfied by means of reasoning analogous to that in the proof of Lemma 1 of Mendo and Hernando (2008a), part (i). $\qquad\square$

**Proof of Proposition 3.** The confidence for the estimator (3.4) is expressed as $c_2(p) - c_1(p)$, $c_1(p) = P[N \leq n_1 - 1]$, $c_2(p) = P[N \leq n_2]$, where $n_1$ and $n_2$ are given by (5.2) with $d = 1$. Let $c_2' = 1 - c_2$. A similar argument as in the proof of Proposition 2, based on the Poisson theorem, shows that $\lim_{p \to 0} c_1(p) = \bar{c}_1$ and $\lim_{p \to 0} c_2'(p) = \bar{c}_2'$, with $\bar{c}_1 = \gamma(r, \Omega/\mu_1)$, $\bar{c}_2' = 1 - \gamma(r, \Omega\mu_2)$. Thus, to establish the desired result, it suffices to prove that $c_1(p) < \bar{c}_1$ and $c_2'(p) < \bar{c}_2'$ for $\mu_1, \mu_2$ as in (3.5).

Regarding $c'_2(p)$, it is shown in Appendix C of Mendo and Hernando (2006) that $P[N \leq \lfloor a/p \rfloor] > \lim_{p \to 0} P[N \leq \lfloor a/p \rfloor]$ for $a > r + \sqrt{r}$. Equivalently, for any $\theta > 1$,

$$P\big[N \geq \lfloor \theta(r + \sqrt{r})/p \rfloor + 1\big] < \lim_{p \to 0} P\big[N \geq \lfloor \theta(r + \sqrt{r})/p \rfloor + 1\big].$$

This can be made to correspond to $c'_2(p) < \bar{c}'_2$ by taking $\lfloor \Omega \mu_2/p - 1 \rfloor = \lfloor \theta(r + \sqrt{r})/p \rfloor$, that is,

$$\mu_2 = \frac{\theta(r + \sqrt{r}) + p}{\Omega}. \tag{5.43}$$

Thus, the inequality $c'_2(p) < \bar{c}'_2$ holds if $\mu_2$ is given by (5.43) for some $\theta > 1$ or, equivalently, if $\mu_2 > (r + \sqrt{r} + p)/\Omega$. Therefore, it holds for arbitrary $p$ if $\mu_2$ satisfies the second inequality in (3.5).

As for $c_1(p)$, from (2.1), it can be expressed as

$$c_1(p) = \sum_{n=r}^{n_1-1} f(n) = \frac{p^r}{(r-1)!} \sum_{n=r}^{n_1-1} (n-1)^{(r-1)}(1-p)^{n-r}. \tag{5.44}$$

According to (5.2) with $d = 1$,

$$n_1 < \frac{\Omega}{p\mu_1} \tag{5.45}$$

and thus

$$\begin{aligned}
\bar{c}_1 = \gamma(r, \Omega/\mu_1) &= \frac{p^r}{(r-1)!} \int_0^{\Omega/(p\mu_1)} t^{r-1} \exp(-pt)\, dt \\
&> \frac{p^r}{(r-1)!} \int_r^{n_1} t^{r-1} \exp(-pt)\, dt.
\end{aligned} \tag{5.46}$$

From (5.44), (5.46) and Lemma 4, it follows that the inequality $c_1(p) < \bar{c}_1$ is satisfied if

$$n_1 \leq \frac{r - \sqrt{r}}{p}. \tag{5.47}$$

Using (5.45), it is seen that (5.47) is fulfilled if $\mu_1$ satisfies the first inequality in (3.5). □

**Proof of Theorem 3.** For $\Omega = \Omega^*$, using (2.4) and (2.5), the conditions in (3.5) are written as (3.7) and (3.8). The left-hand sides of (3.7) and (3.8) are increasing functions of $M$, whereas the right-hand sides decrease with $r$. These inequalities can thus be written as $M \geq h_1(r)$, $M \geq h_2(r)$, where $h_1$, $h_2$ are decreasing functions. This proves that for each $r$, one of the inequalities implies the other. Furthermore, defining $h(r) = \min\{h_1(r), h_2(r)\}$, the allowed range for $M$ is expressed as $M \geq h(r)$ and $h$ is a decreasing function. It only remains to prove that the limiting condition is (3.7) for $r = \{3, 4\}$ and (3.8) for $r \geq 5$.

The left-hand sides of (3.7) and (3.8) are continuous functions of $M > 1$. Considering $r$ as if it were a continuous variable, the right-hand sides are also continuous functions. Assume that (3.7)

implies (3.8) for a given $r_1$, that is, that the latter is satisfied when the former holds with equality. Likewise, assume that (3.8) implies (3.7) for a given $r_2$. The continuity of the involved functions then implies that there exists $t \in [\min\{r_1, r_2\}, \max\{r_1, r_2\}]$ such that both (3.7) and (3.8) hold with equality for $r = t$, that is,

$$\frac{M-1}{\log M} = \frac{t + \sqrt{t}}{t - 1}, \qquad \frac{M \log M}{M - 1} = \frac{t + \sqrt{t} + 1}{t}. \tag{5.48}$$

Multiplying both equalities in (5.48) and substituting into the first yields

$$\log \frac{t + \sqrt{t} + 1}{t - \sqrt{t}} - \frac{2\sqrt{t} + 1}{t} = 0. \tag{5.49}$$

It is easily shown that (5.49) has only one solution, which lies in the interval $(4, 5)$. Thus, one of the two conditions (3.7) and (3.8) is the limiting one for $r \in \{3, 4\}$, whereas the other is for $r = \{5, 6, 7, \ldots\}$. Taking any value from each set, the limiting condition is seen to be (3.7) in the former case and (3.8) in the latter. $\qquad\square$

# Acknowledgements

# References

Abramowitz, M. and Stegun, I.A., eds. (1970). *Handbook of Mathematical Functions*, 9th ed. New York: Dover.

Chen, X. (2007). Inverse sampling for nonasymptotic sequential estimation of bounded variable means. Available at arXiv:0711.2801v2 [math.ST].

Haldane, J.B.S. (1945). On a method of estimating frequencies. *Biometrika* **33** 222–225. MR0019900

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. New York: Springer. MR1639875

Mendo, L. and Hernando, J.M. (2006). A simple sequential stopping rule for Monte Carlo simulation. *IEEE Trans. Commun.* **54** 231–241.

Mendo, L. and Hernando, J.M. (2008a). Improved sequential stopping rule for Monte Carlo simulation. *IEEE Trans. Commun.* **56** 1761–1764.

Mendo, L. and Hernando, J.M. (2008b). Unbiased Monte Carlo estimator with guaranteed confidence. In *Proc. IEEE Int. Workshop on Signal Processing Advances in Wireless Communications* 625–628. New York: IEEE.

Mikulski, P.W. and Smith, P.J. (1976). A variance bound for unbiased estimation in inverse sampling. *Biometrika* **63** 216–217. MR0397965

Papoulis, A. and Pillai, S.U. (2002). *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill.

Petkovšek, M., Wilf, H.S. and Zeilberger, D. (1996). *A = B*. Wellesley: Peters. MR1379802

Prasad, G. and Sahai, A. (1982). Sharper variance upper bound for unbiased estimation in inverse sampling. *Biometrika* **69** 286. MR0655698

Sathe, Y.S. (1977). Sharper variance bounds for unbiased estimation in inverse sampling. *Biometrika* **64** 425–426. MR0494647